People’s Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

University of Constantine 1

Faculty of Letters and Languages

Department of Letters and English Language

N°ordre 43/DS/2020
N° serie 01/Ang/2020

Automaticity and Cognition in Received Pronunciation English Vowel Perception, and Attitudes towards English Language Pronunciation:

The Case of Third-Year Students at the *École Normale Supérieure* Assia DJEBAR, Constantine

Thesis submitted to the Department of Letters and English Language in candidacy for the degree of Doctorat-ès-sciences in *Applied Linguistics and Foreign Language Teaching*

By: **Mr. Ramzi NOUIOUA** Supervisor: **Prof. Hacene SAADI**

Board of Examiners

Chairman:	Prof. Nacif LABED	University of Constantine 1
Supervisor:	Prof. Hacene SAADI	University of Constantine 1
Member:	Prof. Youcef BEGHOUL	University of Constantine 1
Member:	Prof. Riad BELOUAHEM	University of Constantine 1
Member:	Prof. Hacene HAMADA	ENS Constantine
Member:	Prof. Said KESKES	University of Setif 2

02 September 2020

Dedication

I dedicate this work to my late father, who nursed me with affection, love, and determination to pursue success, and to my wife and kids, who have been so inspirational and supportive, as they have been there in every step through the way.

Acknowledgements

My gratitude and thanks go my thesis supervisor, Prof. Hacene SAADI, for his guidance, help and expertise in the realisation of this research work. Without his encouragement and insistence on the completion of the work, it would have never been completed.

I am thankful to members of the board of examiners for their time reviewing the research work and valuable comments on it.

I would like also to thank the participants in this research work, who were cooperative, committed, and patient.

Abstract

The present research investigated perception of RP English prototypic vowels among a group of 53 Algerian 3rd year students of English at the ENS Assia Djebar School in Constantine and their attitudes towards English language pronunciation. The research hypothesised that participants' L1 vowel inventory would not predict perception difficulty and that their attitudes would significantly correlate with their speaking proficiency. The research used a mixed research design to investigate phenomena of interest. Three experimental conditions were manipulated: (a) discrimination of spectral distance among 9 RP English prototypic vowel contrasts, using an A*X* test, (b) identification of RP English prototypic vowels in isolation across durational variation, using an *m*-AFC test, and (c) an *m*-AFC test for identification of RP English prototypic vowels in varying phonetic contexts. The research used a questionnaire to explore Algerian learners' attitudes towards pronunciation. Findings for A*X* discrimination test showed learners' remarkable perceptual sensitivity to spectral distances among 8 vowel contrasts, with *d'* ranging between 1.87 for (/uː/ vs. /ʊ/) and 5.76 for (/ɜː/ vs. /ʊ/), and a null *d'* for (/ɑː/ vs. /ʌ/). The *m*-AFC test for isolated vowels' identification across manipulated durations showed the following identification pattern: /ʊ/ > /ɪ/ > /e/ > /ʌ/ > /æ/ > /ɒ/, with /ɒ/ being significantly the least accurately identified. For long vowels, the pattern was /iː/ > /ɔː/ > /ɑː/ > /ɜː/ > /uː/, with /uː/ being significantly the least accurately identified. In the second *m*-AFC test, identification of in-context vowels followed a varied pattern than in the previous test, with all vowels identified accurately with a hit rate ranging from .74 to .98. The survey demonstrated an overall positive attitude towards English pronunciation learning and instruction and a sensed reluctance in involvement in self-initiated pronunciation learning activities. The research discussed obtained findings within common theories of cross language speech perception, the status of pronunciation among Algerian students of English, drew a brief conclusion and recommended further speech perception research within the Algerian context.

List of Abbreviations

AALP	Attitudes towards Active Learning of Pronunciation
AANLA	Attitudes towards Achievement of Native-Like Accent
ACT	Active Control of Thought
ACT*	Active Control of Thought Star
ACT-R	Active Control of Thought-Rational
AALP	Attitude towards Active Learning of Pronunciation
AIP	Attitudes towards Importance of Pronunciation
AM	Attitudes towards Mispronunciation
ANOVA	Analysis of Variance
APAMP	Attitudes towards Importance of Perceptual Abilities in Pronunciation Mastery
APIF	Attitudes towards Pronunciation Instruction and Feedback
AIPSI	Attitudes towards Importance of Pronunciation in Social Integration
ASL	American Sign Language
AX Test	Same Different Discrimination Test
BA	Bachelor of Arts
BCEF	Bilingual Canadian Speakers of English and French
CPH	Critical Period Hypothesis
d’	Discrimination Index
DRT	Direct Realist Theory
IBM SPSS	International Business Machines Statistical Package for the Social Sciences
IPA	International Phonetic Alphabet
L1	Native Language or Mother Tongue

L2	Second Language
L3	Third Language
M	Mean
MA	Master of Arts
MCE	Monolingual Canadian Speakers of English
MCF	Monolingual Canadian Speakers of French
MD	Mean Difference
MCL	Most Comfortable Listening
m-AFC	Multiple Alternative Forced Choice
ms	Millisecond
MT	Motor Theory
NLM	Native Language Magnet
IPA	International Phonetic Alphabet
P(H)	Hit Rate
P(M)	Miss Rate
PAM	Perceptual Assimilation Model
PDP	Parallel Distributed Processing
PEM	*Professeur de l’Enseignement Moyen*
PES	*Professeur de l’Enseignement Secondaire*
RM-ANOVA	Repeated Measures-Analysis of Variance
RP	Received Pronunciation
SD	Standard Deviation
SLA	Second Language Acquisition
SLM	Speech Learning Model
VOT	Voice Onset Time

List of Figures

List of Tables

Phonetic Symbols

(International Phonetic Association, 1999)

Arabic

Consonants

b	**b**int (daughter)	θ	**θ**aabit (constant)	w	**w**adʒib (duty)
m	**m**in (from)	ð	**ð**iʔb (wolf)	ʔ	**ʔ**ab (father)
f	**f**aata (to pass)	r	**r**ab (god)	x	**x**aaba (to fail)
t	**t**abɣ (tobacco)	l	**l**ij (mine)	tˤ	**tˤ**aaʔir (bird)
d	**d**ijn (religion)	k	**k**ahf (cave)	dˤ	**dˤ**aar (harmful)
n	**n**abaat (plant)	q	**q**aala (to say)	sˤ	**sˤ**aata (to shout)
s	**s**aala (to leak)	x	**x**aal (uncle)	ðˤ	**ðˤ**aalim (unfair)
z	**z**amaan (time)	h	**h**adʒa (to get excited)	lˤ	ðˤi**lˤ** (shadow)
ʃ	ʃaaba (to become white-haird)	ħ	**ħ**adaɵ (event)	ʕ	**ʕ**aada (to come back)
ʤ	ʤaad (serious)	j	**j**amijn (right)		

Vowels

i	ʕ**i**dd (promise!)	u	ʕ**u**dd (come back!)
ij	ʕ**ij**d (feast)	uw	ʕ**uw**d (lute)
a	ʕ**a**dd**a** (to count)	aj	b**aj**t (home)
aa	ʕ**aa**da (to come back)	aw	ʕ**aw**l (support)

French

Consonants

b	**b**ato (ship)	p	**p**ys (flea)	w	l**w**ɛ̃ (far)
m	**m**o (word)	ʁ	**ʁ**ymœ**ʁ** (rumour)	j	m**j**ɛ̃ (mine)
f	**f**u (insane)	l	**l**y (read)	ɥ	s**ɥ**isid (suicide)
t	**t**yb (tube)	k	**k**uto (knife)	ɲ	di**ɲ** (worthy)
d	**d**ɔʁ (sleep!)	v	**v**ilo (bicycle)	ɳ	paʁki**ŋ**
n	**n**ɔ̃ (no)	ʒ	**ʒ**uʁ (day)	z	**z**ɛbʁ (zebra)
s	**s**inɔ̃ (if not)	g	**g**itaʁ (guitar)	ʃ	**ʃ**alœʁ (heat)

Vowels

i	as**i**d (acid)	o	t**o** (early)	ə	s**ə**ʁɛ̃ (calm)
e	s**e**lyl (cell)	ɔ	t**ɔ**ʁ (wrong)	ɑ̃	mam**ɑ̃** (mum)
ɛ	m**ɛ**ʁ (mother)	y	vul**y** (deliberate)	a	b**a**lɔ̃ (ball)
u	l**u** (wolf)	ø	bl**ø** (blue)	œ	k**œ**ʁ (heart)
ɛ̃	sɘd**ɛ̃** (suddenly)	ɔ̃	tɑ̃p**ɔ̃** (stamp)		

RP English

Consonants

b	**b**ɪn	p	**p**ɒt	w	**w**et
m	**m**eɪt	l	**l**aɪf	j	**j**et
f	**f**leɪvə	k	**k**lɪŋ	h	**h**aɪv
t	**t**æb	v	**v**eɪn	z	**z**est
d	**d**aɪs	ʒ	me**ʒ**ə	ʃ	**ʃ**aɪn
n	**n**ɒt	g	**g**et	ð	**ð**æt
s	**s**eɪl	ʤ	**ʤ**iːnz	θ	**θ**ɔːt
ʧ	**ʧ**ɪn	r	**r**eɪn		

Vowels

iː	b**i**ːn	ʊ	f**ʊ**l	ə	**ə**gen
ɪ	b**ɪ**n	ʌ	b**ʌ**t	ɑː	b**ɑ**ːt
e	b**e**d	ɜː	b**ɜ**ːp	eɪ	**eɪ**t
æ	b**æ**d	ɒ	ʃ**ɒ**t	aɪ	**aɪ**s
uː	s**u**ːn	ɔː	ʃ**ɔ**ːt	ɔɪ	**ɔɪ**l
ɪə	t**ɪə**	eə	t**eə**	aʊ	h**aʊ**
əʊ	bl**əʊ**	aɪə	h**aɪə**	aʊə	s**aʊə**

Contents

Chapter 1

Automaticity: Literature Review

Chapter 2

Speech Perception: Literature Review

Chapter 3

Attitude: Literature Review

Chapter 4

Methodology

Chapter 5

Same-different Vowel Discrimination Test

Chapter 6

Test of Vowel Identification in Isolation

Chapter 7

Test of Vowel Identification in Context

Chapter 8

Survey of Attitudes towards English pronunciation

Introduction

Chapter 9

Discussions

Appendices

General Introduction

Described by many as the *Cinderella* or *orphan* of language teaching and learning, pronunciation remains probably the least investigated area in SLA research, coloured by received ideas about the difficulty or even the impossibility of attaining native-like proficiency because of neurophysiological factors beyond human control and the everlasting effect of L1 experience on learning of further languages. These assumptions have guided SLA research and unfortunately backed away optimistic outlooks on effective teaching and learning of L2 pronunciation. Nevertheless, recent findings in L2 phonological acquisition have challenged the long established assumptions of the presumably irreversible effects of age in L2 phonological acquisition and highlighted the need for further research on the role of learning experiences and individual variation in the acquisition of L2 phonology, involving perception and production processes at segmental and supra-segmental levels. Psycholinguistic research literature has constantly emphasised the precedence of speech perceptual processes as a sine qua non of developing proficient L2 pronunciation, and applied social psychological research literature has highlighted the mediating role of psychological factors in shaping learners' L2 pronunciation practices.

The present research aims to contribute to the date substantially limited findings on research examining both Algerian learners' perceptual abilities of segmental elements of RP English vowels and explore learners' attitudinal variation with respect to English pronunciation learning and instruction. Essentially motivated by the lack of speech perception research in the Algerian context, the main objective of the present research is to assess Algerian university learners' perception of RP English vowels. The special status of English as an L2 within the Algerian educational system (being a foreign language introduced after French) makes the present research concern limited to the current learners' perceptual abilities resulting from years of formal instruction. Although the present research may occasionally employ arguments relative to the effects of learners' previous linguistic experiences, it does not pretend to deal firmly with the issue of perception from a developmental perspective, as the substantial lack of already established empirical findings in the field may significantly jeopardise the reliability of any assumptions to advance possibly herein.

The present research attempts to offer comprehensive insight into the issues of automaticity in speech perception and attitude in L2 acquisition in a formal instructional setting. For that purpose, the present research presents a relatively lengthy detailed review of automaticity and speech perception research because of their relative complexity, and a concise review of the attitude concept in SLA research. Chapter 1 introduces the foundations of automaticity through the evolution of psychology as a modern science, bearing on various psychological concepts involved in mainstream research of automaticity in skill acquisition such as computation modelling in cognitive science and attention in models of information processing systems. Rather than presenting strictly language related examples, Chapter 1 employs a variety of examples related to different cognitive domains to illustrate the wide applications of the concept of automaticity.

Chapter 2 provides an extensive review for understanding the nature of the sound from various perspectives and presents major issues in speech perception research, beginning with the inherent philosophical debate over the physical versus the psychological nature of the sound. Chapter 2 carries on with presentation of speech perception issues, namely perceptual units for speech analysis, speech segmentation issue, and provides an extensive review of dominant speech perception theories such as the *Motor Theory*, the *Direct Realist Theory*, and psychoacoustic models. Chapter 2 ends with a review of cross-language speech perception research, the age issue in language learning, and speech perception models in SLA research, including the Native Language Magnet Theory, the Speech Learning Model, and the Perceptual Assimilation Model.

Chapter 3 introduces straightforwardly the concept of attitude from a social psychological perspective, demonstrates the inherent ambiguity of attitude in research, sets a working definition for attitude constructs and ends with a brief review of attitude research in L2 pronunciation learning.

Chapter 4 presents the methodology of the research. It initially states the motivation behind the research and defines its main research questions and hypotheses. It further provides a synopsis of the various methodological steps followed to undertake this research work. This synopsis provides a description of (a) the participants in the research, (b) the research design as including experimental manipulations and a questionnaire, (c) the stimuli materials and the software used for their synthesising, (d) the general procedure to conduct all designed actions, and (e) the computer software used for data processing and analysis.

Chapters 5, 6, 7, and 8 constitute the core of the research fieldwork. Chapter 5 offers a detailed account of the first research experiment, same different vowel discrimination test. It presents an overview of the pilot research and the causes that have led us to the adoption of the current experiment, and provides a full description of the synthetic stimuli materials and procedures followed during experimentation. Then, a detailed presentation of results and their discussion follows. Similarly, Chapter 6 offers a detailed account of the second research experiment, isolated vowel tokens identification test. It proceeds following the logic employed in the previous experiment in Chapter 5 and provides an exhaustive analysis and discussion of obtained findings. Chapter 7 presents the third and the last experiment in the research. It begins with an overview of the experiment and then provides full description of stimuli materials and procedures followed during experimentation. It then proceeds with the analysis and discussion of the obtained findings. Chapter 8 concerns the applied social psychological aspect in the research. The chapter constitutes the exploratory part of the research. The chapter begins with the description of the developed survey and the various elements it includes and then proceeds with the analysis and discussion of findings.

Chapter 9 offers a general discussion of findings, a conclusion, and recommendations for further research.

Finally, the list of references follows along with appendices for further review of detail.

Chapter 1

Automaticity: Literature Review

Introduction

As a theoretical concept adopted in modern psychology, automaticity has always been a centre of interest for researchers investigating observed patterns in human behaviour, their development, and learning effects on these patterns (James, 1890, 1910; Wundt, 1897; Jastrow, 1906; Pavlov, 1927; Watson, 1930). To identify the nature of automaticity and determine its defining attributes, researchers have appealed to numerous psychological concepts such as perception (Ashby, Ennis, & Spiering, 2007; LaBerge, 1975, 1981), attention (Shiffrin & Schneider, 1977; Schneider, Pimm-Smith, & Worden, 1994), memory (Balota, 1983; Mullennix, Sawusch & Garrison, 1992), learning (Anderson, 1976, 1982, 1983, 1992, 1996, 2015), and social cognition (Bargh & Cohen, 1978; Bargh, 1982; Bargh, 1989; Bargh, Schwader, Hailey, Dyer, & Boothby, 2012). Employing diverse terminology such as involuntary acts, reflexes, and acquired habits to refer to automatic behaviours, researchers have also emphasised that automatic behaviours require decreased effort, a minimum of conscious attention and fewer cognitive resources for their processing and performance, as against less automatic actions or voluntary ones that require significant effort, conscious control, and significant memory and attentional resources.

Nonetheless, the probable simplicity with which researchers may conceptualise automaticity has generally proved itself theoretically and empirically illusive. Cumulative literature on the concept and its evolution demonstrate both the complex nature of automaticity and substantial difficulty researchers are likely to encounter in their investigations that may not be without theoretical and empirical costs. Equally, growing literature makes it difficult to provide a broad, clear-cut review of the concept, for most researchers adopt integrative approaches to investigating automaticity without setting an agreed-upon conceptual framework. Hereinafter, within the purview of our research interests, we review what we believe to be the most conspicuous frameworks of automaticity and their underlying foundations in order to provide a clear, comprehensive background for the concept, trace its evolution, and demonstrate its applications and relevance to our research.

1.1 Automaticity in Early Works of Psychology

The concept of automaticity has a long history in psychology and physiological psychology, with forefathers as James (1890, 1910), Wundt (1897), Foster and Sherrington (1897), Jastrow (1906), Pavlov (1927), and Watson (1930), exploring human adaptation to recurring stimulus and patterns of behaviour. Conceptualisations of automaticity have always appealed to terms such as instinct, spontaneous action, impulsive act, reflex, semi-reflex, unconscious response, subconsciously directed behaviour, uncontrolled habit, unlearned behaviour, etc., reflecting agitating, influential theories of behaviour through development of psychology as a modern science.

For Foster and Sherrington (1897),

> We speak of an action of an organ or of a living body as being spontaneous or automatic when it appears to be not immediately due to any changes in the circumstances in which the organ or body is placed, but to be the result of changes arising in the organ or body itself and determined by causes other than the influences of the circumstances of the moment. (p. 992)

This statement highlights the fact that an automatic action is of a *regular* character, as the beat of a heart, a quality that makes it different from relatively a less automatic action, generally called voluntary or volitional, that is *irregular*. This distinction sets a reasonable clear delimitation of automaticity in action that lies in the degree of regularity an action holds to outside of the effect of circumstantial change, thus making sense to claim that irregularity of an action is a function of volition and deliberate control.

Following the same vein in accounting for the mental and motor nature of psychological experiences, Wundt (1897) introduced several orderly notions to elaborate on essential properties of automatic actions, using his proper terms. First, Wundt (1897) employed the notion of simple and complex volitional processes to distinguish voluntary from impulsive acts. As any volitional process, being simple or complex, is incited by an external sense stimulus, Wundt (1897) stated,

> By impulsive act, then, we mean a simple volitional act, that is, one resulting from a single motive, without reference to the position of this motive in the series of affective or ideational processes. Impulsive action, thus defined, must necessarily be the starting point for the development of all volitional acts, even though it may continue to appear along the complex volitional acts. (p. 187)

The single motive from which an impulsive act results, gives voluntary acts a probable defining feature, for voluntary acts' occurrence is a function of *choice* "among a number of simultaneous and antagonistic motives" (Wundt, 1897, p. 188). That is, the fundamental distinction between impulsive and voluntary acts is a function of availability of choices and the need for selection, and, therefore, impulsive acts may be considered intrinsically determined, for they lack the essential characteristic of will or choice. Second, Wundt (1987) made further assumptions on the relationship between motives of impulsive and voluntary acts, and "retrogradation" (p. 192) of voluntary ones. For Wundt (1897), as motives of voluntary and impulsive acts meld together, voluntary acts, with repeated experience, pass to simple or impulsive acts. This generally occurs when conflict of motives becomes less intense and selection among antagonistic motives is repeatedly restricted to the same choice. Third, Wundt (1897) reflected on the mediating role of retrogradation in generating acts that, if performed further in a consistent manner, would render determining motives of voluntary and impulsive acts equally insignificant and transient. Stated differently, initial original stimulation, either affective or ideational, operating as a motive, "causes the discharge of the act before it can be apprehended as an idea" (Wundt, 1897, p. 193), thus generating an automatic act. For Wundt (1897), the gradual suppression of conflicting motives among motives of acts is responsible for generating automatic acts relevant to various aspects of life, including an individual's intellectual, moral, and aesthetic facets of life.

Evoking the possibility of measurement of volitional and impulsive acts, Wundt (1897) further suggested the use of reaction experiments to measure the speed of volitional processes in laboratory investigations, a way to (a) furnish means to understand both mental and motor processes underlying impulsive and voluntary acts, and (b) indirectly investigate the possible relationship between automaticity and time. Reaction time, as a quantitative measurement of deliberate control of behaviour, has proved helpful in psychophysical research investigating the relationship between stimuli and both sensation and perception (Fitts & Posner, 1967, p. 93).

Concurrent to Wundt, Pavlov (1927), in an elaborate investigation of the organism's basic or instinctive acts in a series of original, experimental observations, posited,

> The different kinds of habits [*behaviours*] based on training, education and discipline of any sort are nothing but a long chain of conditioned reflexes. We all know associations, once established and acquired between definite stimuli and our responses, are persistently and, so to speak, automatically reproduced, sometimes even although we fight against them. (p. 395)

Further to relating automaticity to training and education effects, Pavlov (1927) emphasised the necessity for gradual stages in approaching new and difficult tasks,

> extra stimuli inhibit and discoordinate a well-established routine of activity, and how a change in a pre-established dislocates and renders difficult our movements, activities and the whole routine of life. (p. 396)

Pavlov's observations on habit formation have ever since founded the cornerstone of empirically based research enterprise on learning in the history of psychology, *behaviourism*. Within the habit formation[1] tradition and in order to account for differences between

---

[1] It is important to note here that, in psychological literature, the concept of habit formation is different from habituation. Habit formation is synonymous with learning in the behaviouristic tradition, that is, the likelihood of repeating a specific behaviour that, initially preceded by a

deliberately intended and unintended acts, Watson (1930) elaborated Pavlov's ideas in considering automatic acts by offering an alternative classification of acts or responses, that of "learned" and "unlearned" (p. 15) ones, in place of common classification of responses as external (overt) and implicit (covert) responses. For Watson (1930), learned acts include "all of complicated habits and all of conditioned responses" (p. 16), while the unlearned acts include "all of the things that we do in earliest infancy before the processes of conditioning and habit formation get the upper hand" (p. 16). This statement sets an important fact inasmuch as it specifies an age limit for unlearned acts, confined to occur only in an early stage of life, in comparison with learned acts that are presumably acquired after infancy and that may develop into automatic acts themselves. This assumption suggests that automatic acts and formed habits are likely to apply in degrees to all stages of life and may generalise to all aspects of life of an individual. For Watson (1930, p. 238), learned acts in certain situations are almost automatic, suggesting a similar notion of retrogradation of voluntary acts into involuntary or rather automatic ones, as postulated by Wundt (1897).

James (1910) offered an analogous account of the way involuntary responses or "reflex acts" (p. 92) might differ from voluntary ones. In James's (1910) words , a voluntary response is less automatic than an involuntary one that often occurs too quickly to be deliberately intended, "for a man might by conscious effort learn to perform it [*voluntary response*] more skilfully, or even to suppress it altogether" (p. 92). This statement suggests that consciousness in mobilising effort and control of effort suppression are probable delimitation lines between involuntary and voluntary responses, or automatic and less automatic ones. James (1910) further claimed,

---

stimulus, is determined by a positive reinforcer, while habituation is "the progressive reduction of an organism's behavior in response to a repeated stimulus" (Fennel, 2012, p. 4), or simply irresponsiveness.

> It [*voluntary response*] is purely the result of education, and is preceded by a consciousness of the purpose to be attained and a distinct mandate of the will. It is a 'voluntary act.' Thus the animal's [*organism's*] reflex and voluntary performances shade into each other, being connected by acts which may often occur automatically, but may also be modified by conscious intelligence. (p. 92)

According to James (1890), automatic and involuntary acts are equally important and complementary in the life of an individual. In probably one of the most articulately formulated descriptions of the need for automaticity, its development and its multiple ends, James (1890) stated,

> The great thing, then, in all education, is to *make our nervous system our ally instead of our enemy*. It is to fund and capitalize our acquisitions, and live at ease upon the interest of the fund. *For this we must make automatic and habitual, as early as possible, as many useful actions as we can*, and guard against the growing into ways that are likely to be disadvantageous to us, as we should guard against the plague. The more of the details of our daily life we can hand over to the effortless custody of automatism, the more our higher powers of mind will be set free for their own proper work. There is no more miserable human being than one in whom nothing is habitual but indecision …" (p. 80)

James's (1890) statements are clearly consummate insofar as they set ground to distinguish involuntary actions from voluntary ones and the way we should construe the relationship between automatic and less automatic actions within a definite psychological framework, and understand the various ends of the phenomenon.

In a similar effort to explore the subconscious in modern psychological terms in differentiating classes of behaviours, Jastrow (1906) reflected on awareness or consciousness and volition, their mechanisms, respective role in the distribution of attention (as a mental phenomenon) in mental processes, and the subconscious maturation of thought. Jastrow (1906) adopted a psychological scale to distinguish two classes of behaviours in terms of function: (a) a low level class of subconsciously directed behaviours that require "a modest share of awareness" (p. 22), and (b) an intermediate level class of consciously directed behaviours that, usually performed with separated, divided attention, "demand a more moderate range of awareness” (p. 23). For Jastrow (1906), it is likely for subconsciously directed behaviours to require moderate awareness on special psychological occasions’ and for consciously directed behaviours to descend to the level of automatic actions on rare occasions. The functioning of the consciously directed behaviour relates to the subconsciously directed one, as Jastrow (1906) explained,

> Though their [*consciously directed behaviour*] performance involves a variable measure of cooperation of the highest centres, yet their functioning depends specifically upon the integrity of centres between those whose status is in the main physiological and those that demand the most constant directive and conscious control. They fluctuate not only with the interest, ambition, effort, importance that stimulates to their performance, but also with one’s condition and “form””. (pp. 23-24)

This line of argument is remarkably similar to that of James (1890) in that it emphasises the principle of effective utility of assigning different amounts of attention and mental control resources to direct consciously and subconsciously daily life behaviours as an apparent parsimonious and efficacious division of labour (Jastrow, 1906). This division of labour materialises visibly in three key points: (a) the delegation of several repeated activities, routines, or subroutines to semi-automatic mechanisms, freeing commanding attention to deliberation and volition, (b) the suggestion of a range of motives for behaviour, each implying a specific manner and distribution of awareness, and (c) the possibility of scaling the intellectual type of the behaviour.

On the ground of the above, we believe that James (1890, 1910), Foster and Sherrington (1897), Wundt (1897), Jastrow (1906), Pavlov (1927), and Watson (1930) have set some of the earliest essential notions for understanding and investigating automaticity. Though not defining automaticity per se, the postulated notions describe characteristic features of automatic behaviours. These features are necessary to understand basic conceptualisation of automaticity in contemporary psychology and their compilation to form sophisticated definitions for the concept.

1.2 Automaticity in Cognition

With advances of cognitivism in the 1950's, neuroscience, computer science, and novel developments in linguistics, researchers have grown unsure about the resourcefulness of external, behavioural evidence to account for the complexity of human behaviour (Hebb, 1949; Chomsky, 1957; Newell, Shaw, & Simon, 1958; Neisser, 2014/1967; Simon & Newell, 1970; Thagard, 2005; Anderson, 1976, 1982, 1983, 1992, 1996, 2015). Researchers have grown more sceptic about an individual's capacity to access the world without the mediating role of sensory organs and several theoretically elaborated processes of cognition, comprising attention, perception, learning, memory, language, problem solving, reasoning, and thinking (Malim, 1994). Interest in mediating processes responsible for behaviour has caused the appearance of several research perspectives such as computational cognitive science, experimental cognitive psychology, and cognitive neuroscience. These three approaches have formed a multidisciplinary, interconnected area of research difficult to dismantle with clear demarcation lines, and their interrelationships have gradually grown firmer, rendering quite impossible focus on one without considering the others (Eysenck & Keane, 2010).

In the following sections, we first review influential literature on cognition that has set new ground and perspective for the conception of behaviour, notably skill acquisition, learning, and skilled performance, and then highlight the role of cognitive theory in various conceptualisations of automaticity proposed by researchers in the field. As we do not pretend to offer an exhaustive review of the available literature, our review provides few selected historic elements deemed as a sine qua non of cognitive theory. We further deal with some of the cognitive theory elements in detail in an attempt to offer a complete, unambiguous background on the foundational components of the concept of automaticity and shed light on its theory propositions.

1.2.1 Automaticity in computational cognitive science.

Early efforts on configuration of cognitive systems date back to the works of Turing and Von Neumann in the 1940's, whose works on the logical, mathematical, and engineering designs of computers have set the foundation for a theory of information processing and automata (Von Neumann, 1958; Aspray, 1990; Copeland, 2004; Petzold, 2008). In his work, '*The Computer and the Brain*', Von Neumann (1958) proposed an "a priori promising" (p. 1), mathematically guided approach towards the understanding of the nervous system. Being merely a systematised set of formalised speculations on the topic, Von Neumann (1958) offered rationalisations of his proposals and admitted that logics and statistics should primarily, though not exclusively, be the tools for a potential information theory. The ideas of Turing and Von Neumann's on mathematics and logics and the analogy between the computer and the brain, have not set forth the current architecture of computers alone, but have settled for all the quintessential configuration of cognitive systems that comprise (a) an input system, (b) a processing system with memory structures, and (c) an output system. More significantly, their ideas have shaped mainstream cognitive research to account for behaviour and promoted the use of a specific jargon such as input, information processing, encoding speed, direct access to memory, access time to memory, memory registers, serial and parallel processing, and local and distributed representation of knowledge or information (Malim, 1994).

Since then, concepts of information-processing theory have marked cognitive research papers, of which Newell, Shaw, and Simon's (1958) was among the pioneering ones. Their work sets forth a program, a series of instructions to perform, of basic information processes to demonstrate the way an organism generates an observed behaviour. Setting aside the issue of sensory and motor activities alongside the nature of information representation for operational considerations, Newell, Shaw, and Simon (1958) postulated an information-processing theory capable of accounting for the way an individual would perform a series of orderly coordinated operations to solve a problem, a typical behaviour. Their theory postulates,

> (a) A control system consisting of a number of *memories*, which contain symbolized information and are interconnected by various ordering relations…
> (b) A number of *primitive information processes*, which operate on the information in the memories. Each primitive process is a perfectly definite operation for which known physical mechanisms exist.
> (c) A perfectly definite set of rules for combining processes into whole *programs* of processing. From a program it is possible to deduce unequivocally what externally observable behavior will be generated. (Newell, Shaw, & Simon, 1958, p. 151)

As stated, the theory possesses an explanatory and a predictive power in the sense that (a) it explicitly describes the systematic stages of information processing, and (b) it attempts to predict the way the organism executes information processes in varying environmental circumstances. The theory also sets out three essential information storage structures to generate behaviour: (a) a store of memories where interrelated represented information lies in, (b) a number of basic processes, best described in terms of first-order logical connectives[2] operating on mathematical axioms, and (c) a series of definite, clear systematic operations to execute, i.e. a program.

With a similar approach in prioritising mental events as a subject matter for study rather than observed behaviour, Neisser (2014/1967) stated,

> There certainly is a real world of trees and people and cars and even books, and it has a great deal to do with our experiences of these objects. However, we have no direct immediate access to the world, nor to any of its properties. The ancient theory of eidola,

---

[2] First order logic is useful in writing computer programs. Building blocks of first-order logic consist of connectives: $\wedge$ (and), $\vee$ (or), $\neg$ (not), $\rightarrow$ (implies), = (equals); quantifiers $\forall$ (universal quantifier “For all”), $\exists$ (existential quantifier “There exists”); and an infinite number of variables x, x_1, x_2, etc. First-order logic represents symbolised reasoning expressed in sentences; each can be broken down into a predicate P and a subject x, where the predicate defines or modifies the subject. Consider the following example: P (predicate) ‘is a consonant letter’, Q (predicate) ‘combines with a vowel letter’, we can say: $\forall x: P(x) \rightarrow Q(x)$ that translates to ‘For all x, if x is a consonant letter, then it combines with a vowel letter’ or $\exists x: Q(x) \rightarrow P(x)$ that translates to ‘There exists an x, if it combines with a vowel letter, then it is a consonant letter’. For more detail, see Barwise, 1977.

> which supposed that faint copies of objects can enter the mind directly, must be rejected. Whatever we know about the world has been mediated, not only by the organs of sense but by complex systems which interpret and reinterpret sensory information. The activity of the cognitive systems result in—and is integrated with—the activity of muscles and glands that we call "behaviour." It is also partially—very partially—reflected in those private experiences of seeing, hearing, imagining, and thinking to which verbal descriptions never do full justice. (Neisser, 2014/1967, p. 3)

Elaborating on examples from visual and auditory cognition, Neisser (2014/1967) admitted the triviality of information measurement for the cognitive psychologist and emphasised the importance of mental processes, as sensation, perception, imagery, retention, recall, problem-solving, and thinking as mediating stages of cognition. Moreover, Neisser (2014/1967) affirmed the involvement of these aspects of cognition in every psychological phenomenon, and drew an analogy between the work of an individual's mind and the work of a computer program, stressing the necessity not to confuse the program with the computer. For him, the program "is a series of instructions for dealing with symbols…The cognitive psychologist would like to give a similar account of the way information is processed by men" (Neisser, 2014/1967, p. 8). On his turn, Thagard (2005) recognised further usefulness of the computer analogy to involve the stages of discovery, modification, and evaluation of cognitive theories. For instance, computational ideas about various types of programs are likely to suggest new types of mental structures and processes:

> Theory development, model development, and program development often go hand in hand, since writing the program may lead to the invention of new kinds of data structures and algorithms that become part of the model and have analogs in the theory. For example, in writing a computer program to simulate human addition, a programmer might think of a kind of data structure that suggests new ideas about how children represent numbers. Similarly, evaluation of theory, model, and program often involves all three, since our confidence in the theory depends on the model's validity as shown by the program's performance. (Thagard, 2005, p. 14)

For the sake of simulating complex cognitive functions, programmers need not only create computer programs, but also need to generate and evaluate novel data structures on which to execute their programs efficiently. As likely as not, novel data structures will illuminate the multi-faceted issue of information or knowledge representation relevant to mental and motor skills.

1.2.2 Automaticity in computational modelling.

Belief in artificial intelligence and the computer analogy has caused the emergence of sophisticated computational models to describe internal systems responsible for human behaviour. These internal systems represent discrete structures responsible for various processing of information from the moment environmental stimuli are encoded to the moment a response is produced. These internal systems involve encoding modalities, temporal and memory-based structures, and an output system (Baddeley, 1999). Several features of the functioning of these internal systems are requisite for understanding the concept of automaticity. Among the suggested computational models, *production systems* (Simon & Newell, 1970; Anderson, 1983; 2015; Chomsky & Halle, 1968; Chomsky, 2002) and *parallel-distributed processing models* (McClelland, Rumelhart, & Hinton, 1986; Smolensky, 1986; Zisper & Rabin, 1986; McClleland & Elman, 1986; Rogers & McClleland, 2004) are the predominant computational paradigms. These models have proved interesting findings in demonstrating the way initially limited, partial capabilities in a developing system grow almost continuously into ones that are more powerful (Simon & Newell, 1970). These computational models offer hypothetical accounts for acquisition of various skills and their development over the life span, and equally hold divergent assumptions on issues relating to nature of knowledge formation, representation, and localisation. Plain presentation of basic premises of these computational models is helpful in providing a prerequisite background for the understanding the foundation of automaticity in cognitive sciences.

1.2.2.1 Production systems.

Research accounting for processes underlying cognitive and motor functions originates in production systems aiming primarily at mathematicising (formalising) the logical proof

(McClelland, Rumelhart, & Hinton, 1986; Pullum, 2011) and ascertaining the highly systematised work of the mind that characterises most automatic behaviours and explaining their developmental process. Archetypal researches in the field include: (a) Simon and Newell's (1958, 1970) work on a theory of problem solving; (b) Chomsky's work (1957) on natural language grammar in syntax, commonly known as transformational generative grammar (TGG); and (c) Anderson's work (1976, 1982, 1983, 1996, 2015) on higher cognitive functions, termed Adaptive Control of Thought, or simply ACT. However, their proposals diverge on a number of issues relating to the uniqueness of cognitive functions and the unitary or faculty nature of their underlying systems.

1.2.2.1.1 Work of Simon and Newell.

For Simon and Newell (1970), a production system is a bipartite form, a condition-action pair, analogous to an 'If…, then…' rule[3]. The condition defines a test or a set of tests to execute, and performance depends on satisfaction of condition. If the latter is not satisfied, there will be no production, and the control is transferred to another production or condition-action pair, forming a cyclic operational mode where there is no limit for number of condition-action pairs. To illustrate, we consider the following straightforward example to describe the behaviour of a learner in forming a two-letter English word:

Rule 1: *If* letter 1 is a consonant, *then* add a vowel.

Rule 2: *If* letter 1 is a vowel, *then* add a consonant.

When Rule 1 condition is true, Rule 1 is fired and vowel insertion is carried out. When Rule 1 condition is not true, control is transferred to Rule 2. When Rule 2 condition is true, Rule 2 is fired and consonant insertion is carried out. This simple production system can further be

---

[3] First-order logic sentences and production systems structures are expressed alike; however, it is important to note that: (a) first-order logic sentences define or modify the subject, representing a static system, and (b) rules in production systems explain how an observed behaviour is generated, representing a dynamic system.

extended to take on a structurally more complex system, such as that one whose goal is to form an English meaningful word out of the letters ‘t’, ‘e’, and ‘a’:

Rule 1: *If* [letter 1 is ‘t’], *then* [*If* letter 2 is ‘e’, *then* add ‘a’]

Rule 2: *If* [letter 1 is ‘e’], *then* [*If* letter 2 is ‘a’, *then* add ‘t’]

Rule 3: *If* [letter 1 is ‘a’], *then* [*If* letter 2 is ‘t’, *then* add ‘e’]

Thus, we can generate more elaborate production systems, applying the basic ‘If…, then…’ rule, by modifying the structure of both condition and action, each to include even sub-production rules themselves with precise specifications. The practice of such production rules creates firmly established algorithms, notably in case of processing of repeated similar information, where algorithmic calculations become faster and more efficient resultant of accumulating traces or instances of calculations (procedures) in carrying out similar tasks.

1.2.2.1.2 Work of Chomsky.

As part of a whole theory on grammar of natural language and with focus on language competence, being the worth-investigating, idealised, internalised system of language, Chomsky (1957) argued for a faculty and modular approach to the study of the mind (Fodor, 1983) and derived several of his ideas of syntactic structures from production systems (Pullum, 2011). In describing the internalised system accountable for generating structurally and not necessarily semantically grammatical sentences, Chomsky (1957) stated,

> Suppose that we have a machine that can be in any one of a finite number of different internal states, and suppose that this machine switches from one state to another by producing a certain symbol (let us say, an English word). One of these states is an initial state; another is a final state. Suppose that the machine begins in the initial state, runs through a sequence of states (producing a word with each transition), and ends in the final state. Then we call the sequence of words that has been produced a “sentence”. Each such machine thus defines a certain language; namely, the set of sentences that can be produced in this way. (p. 19)

The operational mode of the suggested system requires an initial, an intermediate, and a final state. Grounding the form of grammar associated with the theory of linguistic structure upon constituent analysis, Chomsky (1957) conceived of grammar as:

> A finite set Σ of initial strings and a finite set F of 'instruction formulas' of the form $X \rightarrow Y$ interpreted : "rewrite X as Y." Though X need not be a single symbol, only a single symbol of X can be rewritten in forming Y…Given the grammar [Σ, F], we define a derivation as a finite sequence of strings, beginning with an initial string of Σ, and with each string in the sequence being derived from the preceding string by application of one of the instruction formulas of F. (p. 29)

As such, grammar of natural language consists of a set of finite rules operating on an immediate, initial, available linguistic information at different representation levels, including phonemes, morphemes, and phrases. Every rule orderly operates on a given linguistic piece of information (input) and transforms it to a new one with a new, derived structure (output). Following Chomsky's line of argument (1957), the application of the simple opening rule in the grammar system, S → NP + VP that translates to a sentence rewrites a noun phrase plus a verb phrase, is capable of generating a set of infinite sentences via further, indefinite application of recursive, enumerable transformation rules. Application of these rules in generating sentences over time equally suggests the likelihood of creation of automatic procedures, not only in processing linguistic units but also in linguistic performance as a function of repeated practice.

1.2.2.1.3 Work of Anderson.

Founding his arguments on the computer analogy in defence of unitary approach to the study of cognition, Anderson (1983) emphasised,

> There are three lines of evidence for the unitary approach. One is the short evolutionary history of many of the higher human intellectual functions, such as those concerned with mathematical problem solving. The second is that humans display great plasticity in acquiring functions for which there was no possibility of evolutionary anticipation. The third is that the various cognitive activities have many features in common...I claim that a single set of principles underlies all of cognition and that there are no principled differences or separations of faculties. (p. 5)

According to Anderson (1976, 1982, 1983, 1992, 1996, 2015), one simple theory would suffice to account for the human complex cognition that involves entire higher-order functions, including language, mathematics, reasoning, memory, and problem solving. Nevertheless, Anderson's theory of architecture of human cognition underwent significant changes since its initial formulation and maintained production systems and memory structures as underlying components of human cognition. ACT, ACT*, and ACT-Rational versions have been continuously devised to account for more complex cognitive functions, using adapted and changing terminology as a function of application area such as solving mathematical problems, language learning, memory experimentation, etc. For the sake of illustration, we believe that reviewing of ACT* will satisfactorily provide an exhaustive framework of ACT that has a broad generalisability. Considering cognitive tasks as goal-directed, problem-solving cases, the ACT* theory has origins in memory experiments, specifically theory of human associative memory (HAM) and production systems (Newell, Shaw, & Simon, 1958, Simon & Newell, 1970; Anderson & Bower, 1974). The crucial assumption of the ACT* theory is the distinction between the "*knowing that*" and "*knowing how*" knowledge (Ryle, 2009/1949, p. 14), or simply between declarative and procedural knowledge, sometimes referred to as explicit and implicit memory, respectively.

Fundamentally, the ACT* is a learning by doing system, production system based, instantiating the interlinking processes of three hypothetical structures of memory: (a) working memory, (b) declarative memory (explicit memory), and (c) production memory (procedural memory). The central part of the ACT* theory is the working memory that processes information the system can access. Accessible information originates in three different sources: (a) locally active, available information in the permanent declarative memory; (b) information deposited by encoding mechanisms from outside world; and (c) information resulting from action of productions.

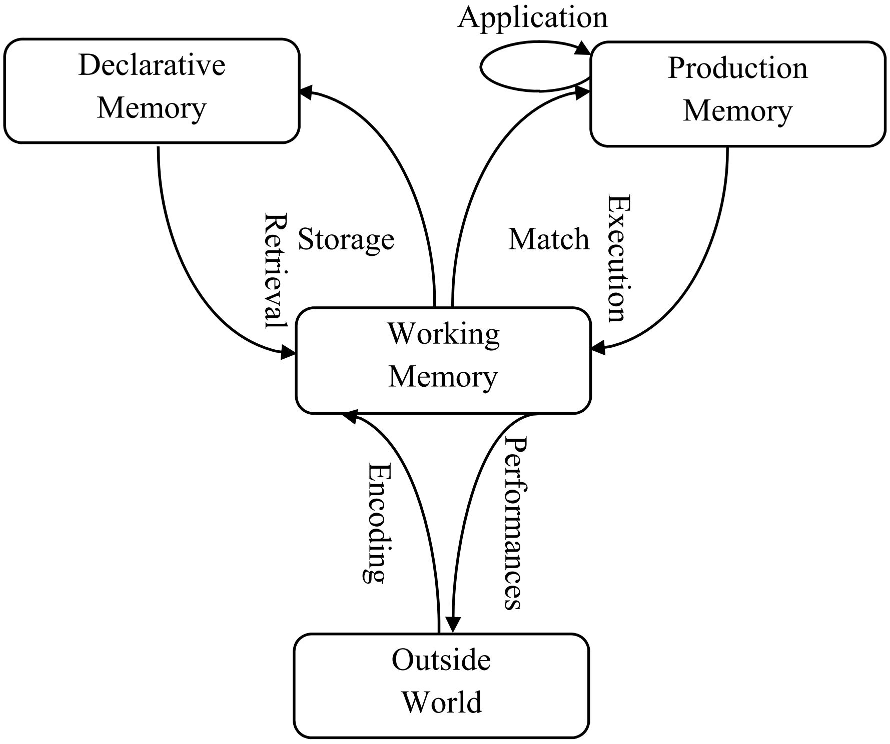


Figure 1. A general framework for the ACT* production system, illustrating the three memory structures and their communicating processes. From "*The Architecture of Cognition*" by J. R. Anderson, 1983, p. 19. Copyright 1983 by Harvard University Press. Reprinted with permission

The ACT* system embraces six information processes: (a) the *encoding process* deposits information obtained from outside world into the working memory; (b) the *performance process* converts commands in the working memory into behaviour; (c) the *storage process* creates permanent records of the contents of the working memory in declarative memory, (d) the *retrieval process* retrieves information from declarative memory; (e) the *match process* synchronises information in the working memory with the conditions of productions; and (f) the *execution process* deposits matched production actions into the working to await for performance. Even though the ACT* theory does not refer to automaticity as a concept per se, it offers a tenable explication for the phenomenon of automaticity through the principles of learning and strengthening phenomena as a function of practice. On this ground, there is much reason to claim that automaticity may build up as a power function of practice of production rules, decreasing both processing and processing time, likely to lead to a more consistent, fast, and smooth processing of information even in the presence of interfering phenomena or concurrent processes. It is important to note that considerable overlap in production rules between current information-processing task and concurrent tasks is likely to affect processing capacity and impose limitations on carrying them out flexibly and successfully, a common case reported in automaticity research under the label of multi-task performance (Hirst, Neisser, & Spelke, 1978, p. 54; Anderson, 1992, pp. 177-178).

Nonetheless, it is imperative to distinguish accrual of automaticity as a power function of practice and repetition from improvement and proficiency of a performance, for automaticity may not necessarily imply skilled performance or reduced error rate. Improvement in performance and emergence of a skilled performance and decrease of processing time in carrying out a cognitive function, are very likely the subject matter of the *power law of learning or practice*, a power function of the relationship between performance measurements, number of trials of practice and error rates (Newell & Rosenbloom, 1981; Anderson, 1982) and practice conditions (Snoddy, 1926). Research on the effects of practice and practice conditions on improving skills and performances involves all types of motor and intellectual skills, like walking, speaking, listening and writing, reading and thinking, etc. (Bryan & Harter, 1899; Fitts & Posner, 1967; Hirst, Neisser, & Spelke, 1978).

1.2.2.2 Parallel distributed processing models.

Recognising the appropriateness of artificial intelligence and the ingenuity of computer programs in simulating and capturing the fluidity and adaptability of human information processing, several cognitive researchers believe that the program or software is not the whole story behind human cognition. They believe that people are smarter than computers and that learning is more than explicit rule formulation as outlined in conventional production systems. McClelland, Rumelhart, and Hinton (1986) highlighted,

> People are far better at perceiving objects in natural scenes and noting their relations, at understanding language and retrieving contextually appropriate information from memory, at making plans and carrying out contextually appropriate actions, and at a wide range of other natural tasks. People are also far better at learning to do things more accurately and fluently through processing experience…people are smarter than today's computers because the brain employs a basic computational architecture that is more suited to deal with a central aspect of the natural information processing tasks that people are so good at. (p. 3)

Supported by physiological plausibility and inspired by functioning of cerebral mechanisms, some cognitive researchers have argued for the inadequacy of the localisationist view to account for the function of the mind, and have alternatively proposed information-processing models based on the assumption of the unitary function of the mind and the dynamic nature of knowledge structures. These models are called neural networks or parallel distributed processing (PDP) models, advocating novel conceptualisations of information-processing, representation, and most importantly the phenomenon of learning. The latter is conceptualised as a spontaneous by-product of processing activity. PDP models consider the interplay of multiple sources of knowledge in understanding human cognition because of complexity and demanding requirements of everyday situations, which necessitate involvement of various knowledge structures. Equally, PDP models have numerous applications to issues like motor control, perception, memory, problem solving, and language learning.

The general framework for PDP models rests upon neurone-like units or simply nodes connected together and affecting each other either by excitation or by inhibition. McClelland, Rumelhart, and Hinton (1986) specified the major characteristics of PDP models that would include:

(a) A *set of processing units*
(b) A *state of activation*
(c) An *output function* for each unit
(d) A *pattern of connectivity* among units
(e) A *propagation rule* for propagating patterns of activities through the network of connectivities

(f) An *activation rule* for combining the inputs impinging on a unit with the current state of that unit to produce a new level of activation for the unit
(g) A *learning rule* whereby patterns of connectivity are modified by experience
(h) An *environment* within which the system must operate.

(McClleland, Rumelhart, & Hinton, 1986, p. 46)

The figure below illustrates the basic framework of PDP models. As neurophysiological findings about the work of the brain inspire and drive PDP basic systems, connectionist networks mimic the activity of the brain units, replacing the ‘computer metaphor’ by the ‘brain metaphor’. As illustrated in *Figure 2* below, a PDP model comprises three layers of units: (a) a set of input units receiving signals from sensory sources or neighbouring internal units; (b) a set of output units sending signals to motoric systems or other external systems unconcerned with modelling; and (c) a set of hidden units, invisible to external systems, sending signals to systems concerned with modelling.

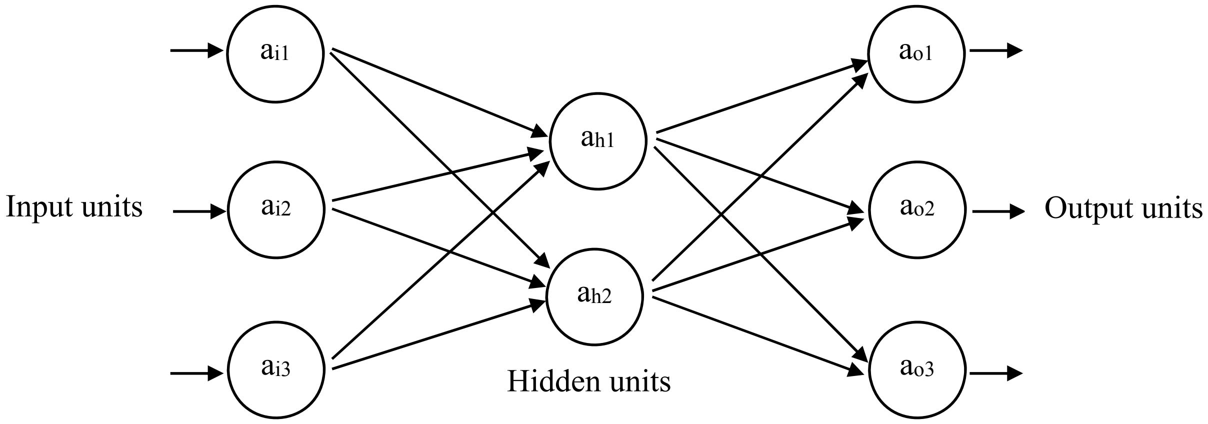


Figure 2. A multi-layered PDP model, with input units, hidden units, and output units. From “A General Framework for, Parallel Distributed Systems”, by Rumelhart, Hinton, and McClelland, in D. E. Rumelhart, J. L. McClelland, and the PDP Research Group (Eds.), “*Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations*”, pp. 45-76. Copyright 1986 by the MIT Press. Adapted with permission.

It is important to state that every unit stands for a particular knowledge representation, likely to exist in the form of a feature, a letter, a word, a concept, or any abstract element. When taken as a whole, the latter makes a meaningful perceptual analysis. The functioning of a PDP model builds upon Hebb's principle (1949) that assumes a probabilistic and statistical nature of communication between neurons, manifesting in excitation and inhibition phenomena, largely formulated by cognitive researchers in 'the neurons that fire together wire together'.

Though it is very conventional among cognitive researchers to model PDP systems in formalised entities and mathematical equations, we limit ourselves, in the present context, to a brief explanation of the way PDP models operate, appealing to essential premises. The work of a PDP model is a linear function of activation and output values of units over time. Understanding this linear function necessitates knowledge of the following basic assumptions using our example illustrated above in *Figure 2* (for detailed specifications, see Rumelhart, Hinton, & McClelland, 1986, pp. 45-76):

(a) At time (t), each input unit has an activation value like $a_{i1}(t)$, $a_{i2}(t)$, $a_{i3}(t)$ that, if it exceeds the *threshold value*, the input unit produces a single output to another unit.

(b) Unidirectional arrows represent links between units. Each unidirectional arrow is associated with a real number called strength or weight of connection, as $w_{i1h1}(t)$, w_{i2h2}, $w_{i2h1}(t)$, $w_{i2h2}(t)$, etc.

(c) At time (t), each hidden or output unit takes an activation value that is the weighted sum of all of the connected units, calculated using the main formula:

$a_{h1} = \Sigma a_i w_{ih}(t)$ (main formula)

$a_{h1} = [(a_{i1} w_{ih1}) + (a_{i2} w_{ih1}) + (a_{i3} w_{ih1})](t)$, and

$a_{h2} = [(a_{i1} w_{ih2}) + (a_{i2} w_{ih2}) + (a_{i3} w_{ih2})](t)$.

As illustrated above, it seems clear that PDP models have different implications for both processing and learning. Contrary to production systems, PDP models model cognitive functions without explicit rules as those found in production systems, assume neither the presence of static memory structures nor a central processing unit, and generalise the following claims:

(a) Knowledge representation and processing exist in a distributed form throughout activation patterns of layers of different processing units.

(b) Learning is a function of finding the right connection strengths and the right circumstances.

(c) The role of experience lies in modification of processing and knowledge structure, involving the modification of patterns of interconnectivity. Additionally, the role of experience can result in development of new connections, loss of existing ones, and modification of the weight of existing connections (Rumelhart, Hinton, & McClelland, 1986, p. 52).

1.3 Attention in Automaticity Research

Since the early days of cognitive theory, researchers have debated the close relation between automaticity and attention, claiming that automatic behaviours can be performed with little attention, if any, whereas less automatic or voluntary ones are very sensitive to attentional resources (Schneider, Pimm-Smith, & Worden, 1994; Pashler, 1998; Logan, Taylor, & Etherton, 1999; Styles, 2006). However, given the complex nature of attention as a key principle in cognitive psychology and its centrality in automaticity research, we provide an essential review of researchers' conceptualisations of attention as a preliminary step to understand the implication of attention in the development of automatic behaviour and skilled performance. It is to note that the review is not inclusive of all raised issues in attention research literature, and that cited examples are pertinent to our research interest.

1.3.1 Groundwork on attention.

James (1890) offered one of the earliest descriptions of attention and its association with consciousness, span of consciousness (the number of things an individual can attend to), and control of the mind.

> It [*attention*] is the taking possession by the mind, in clear and vivid form, of one of what seem several simultaneous possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatter-brained state… (p. 261).

Lying in the heart of an individual's experience, attention describes the focus of perceptive consciousness and awareness on a set of stimuli among a number of immediately available stimuli. James (1890, pp. 269-272) suggested: (a) source, (b) interest, and (c) choice or will as essential elements making the core concept of attention. For source, attention is either sensorial or intellectual: it is sensorial when a sense drives it and intellectual when an idea other than a sense object does. For interest, attention is either immediate or derived (apperceptive): it is immediate when the idea or a sense object is "interesting in itself without relation to anything else" (p. 269) and derived or apperceptive when the idea or a sense object owes "its interest to association with some other immediately interesting thing" (p. 269). For choice or will, attention is either involuntary (effortless and passive) or volitional (effortful and active). In information-processing theory terms, volitional attention occurs in the case of top down processing, while involuntary attention occurs in bottom up processing. According to James (1890, pp. 269-272), the three principles suggest that: (a) volitional attention is always apperceptive, for it is improbable for an individual to mobilise an effort to attend to an object in the absence of a remote interest; (b) sensorial and intellectual attention are either involuntary or volitional; and (c) involuntary intellectual attention is either immediate or apperceptive. More interestingly, relating immediate sensorial attention to age, James (1890) emphasised,

> Sensitiveness to immediately exciting sensorial stimuli characterizes the attention of childhood and youth. In mature age we have generally selected those stimuli which are connected with one or more so-called permanent interests, and our attention has grown irresponsive to the rest. But childhood is characterized by great active energy, and has few organized interests by which to meet new impressions and decide whether they are worthy of notice or not, and the consequence is that extreme mobility of the attention with which we are all familiar in children, and which makes their first lessons rough affairs. (p. 270)

The view of attention as subject to ageing influence[4], probably as a function of selection of an individual's interests, resulting in probable diminution of sensitivity to sensorial stimuli in performance tests, has been predominantly entertained by research on visual attention using the spotlight metaphor[5] in various tasks to assess the spatiotemporal dynamics of attention (for detailed frameworks on visual attention, see D'Aloisio & Klein, 1990, pp. 447-466).

1.3.2 Cognitive conceptualisations of attention.

Though behaviourist psychologists have avoided the implication of attention in their research due to the highly introspective nature of the phenomenon (Eysenck, 1982, p. 7; Lovie, 1983), contemporary accounts of attention have consistently employed many of James's (1890) ideas to define attention using new terminology, reflecting change in mainstream psychological theories, particularly cognitivism. Cognitive researchers have brought interest to attention as a major factor in human cognition, investigating: (a) nature of attention (Kahneman, 1973; Pashler, 1998; Styles, 2006), (b) the development of attention through the life span (Bornstein, 1990; Trehub & Trainor, 1990; Tipper & McLaren, 1990; Chapman, 1990), (c) attention deficits (Burke, 1990; Graf, Tuokko, & Gallie, 1990), (d) the relation of attention with perceptual processing (Ackerman, 1990; Guttentag & Ornstein, 1990; Shiffrin & Schneider,

---

[4] In fact, qualitative and quantitative changes in cognitive performance due to ageing are common issues of debate in cognitive ageing research that involves memory, attention, cognitive-motor skills, language, problem solving, and personality (Lovelace, 1990, p. 407).

[5] Visual attention is likened to a zoom lens that enhances processing of an object that falls within its focus. This visual attention decreases with increasing distance from the attended location.

1977; Goldstone, 1998), and (e) the role of attention in the development of automaticity (Schneider & Kintz, 1967; Mullennix, Sawusch, & Garrison, 1990; Johnson & Ralston, 1994; Ruthruff, Allen, Lien, & Grabbe, 2008; Winfred, 2011).

Basic to the understanding of attention in ordinary language are the issues of: (a) the nature of attention as a causal mechanism or a natural consequence of other processes (James, 1890, p. 291; Styles, 2006, p. 9), (b) mental effort and resources allocation in information processing (Posner & Boies, 1971; Kahneman, 1973; Shiffrin & Schneider, 1977), and (c) stages of information processing and the putative placement of the attention within the information-processing model (Johnston & Dark, 1986, p. 47; Moors & De Houwer, 2006, p. 297).

1.3.2.1 The emergent nature of attention.

James (1890) advanced the view of attention as a resultant and stated,

> Everywhere attention is voluntary, it is possible to conceive of it [attention] as an effect, and not as a cause, a product and not an agent. The things we attend to *come to us* by their own laws. Attention *creates* no idea; an idea must already be there before we can attend to it. Attention only fixes and retains what the ordinary laws of association bring "before the footlights" of consciousness… Effort is felt only where there is a conflict of interests in the mind. (p. 293)

Some cognitive researchers have clearly articulated the emergent nature of attention in terms of priming effects[6], where prior experience with a stimulus affects uncontrollably its subsequent processing and recognition. Therefore, an existing piece of information in the long-

---

[6] Priming describes the process by which an implicit memory trace, resulting from prior exposure to a stimulus, affects its subsequent processing. This can occur at both lower-order, sensorial analysis and higher-order, semantic analysis (Rabitt & Vyas, 1979). For instance, in an auditory priming experiment, the prime stimulus is a sound, with specific frequencies (F_1 500Hz, F_2 600Hz), may subsequently facilitate identification of similar instances of the sound, with slightly varying frequencies (F_1 520Hz, F_2 630Hz; F_1 550Hz, F_2 650Hz; F_1 600Hz, F_2 700Hz, etc.). In this conceptualisation, attention is considered a passive by-product of priming effects.

term memory influences unconsciously and uncontrollably a similar encoded sensorial stimulus. In cognitive theory, this view represents a typical example of parallel processing of information, where, a piece of information may undergo two stages of processing at the same time (Malim, 1994, p. 9). Assumptions of parallel-distributed processing models proponents provide a sophisticated theoretical account of such a phenomenon, as displayed in *Figure 3* below, where perceptual processing and long-term memory activation operate simultaneously.

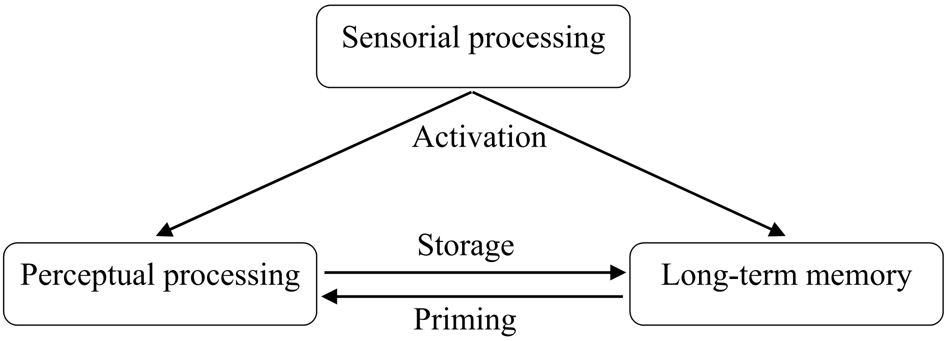


Figure 3. A simplified model of parallel information processing system.

For instance, Jacoby and Dallas (1981) conducted a research on the relationship between presentation of a word and the effect it had on its subsequent perceptual recognition in recognition memory performance. Their research demonstrated the effects of memory in an individual's performance of a perceptual task, as Jacoby and Dallas (1981) stated,

> Prior experience with material can make that material more easily identified or comprehended in perceptually difficult situations. Unlike with standard tests, effects or prior experience on a perceptual task do not logically require that a person be aware that he or she is remembering. Indeed, amnesic patients purportedly show effects of practice that they do not remember having engaged in that prior experience. (p. 306)

In a research on semantic priming, Conrad (1974) explored lexical disambiguation in the language in case of distinct meanings for words, and found that activation of words' meanings in memory was evoked simultaneously when the words were heard in sentences. She concluded, "context which is effective in disambiguating lexical ambiguities in the language has its effects only at a relatively late stage in the cognitive processing in language comprehension" (Conrad, 1974, p. 130).

1.3.2.2 The causal nature of attention.

To emphasise the causative role of attention in information processing, cognitive researchers have differentiated between qualitatively different domains of stimulus processing described in dichotomous labels such as: (a) unconscious vs. conscious, (b) automatic vs. controlled, (c) pre-attentive vs. attentive, and (d) passive vs. active (Kahneman, 1973; Shiffrin & Schneider, 1977; Deutsch & Deutsch, 1964; Johnston & Dark, 1986; Treisman, Vieira, & Hayes, 1992; Moors & De Houwer, 2006), as displayed in *Figure* 4 below. The two qualitatively distinguished domains of information processing are based on two premises. The first premise suggests that: (a) processing of information is orientated by either temporal data originating in environmental stimuli, internally stored information, or both (Treisman & Gelade, 1980), and (b) attention is likely a pre- or post-perceptual mechanism. The second premise advocates the basic conceptualising distinction between attention- and non-attention-demanding processes (Eysenck, 1982, p. 27).

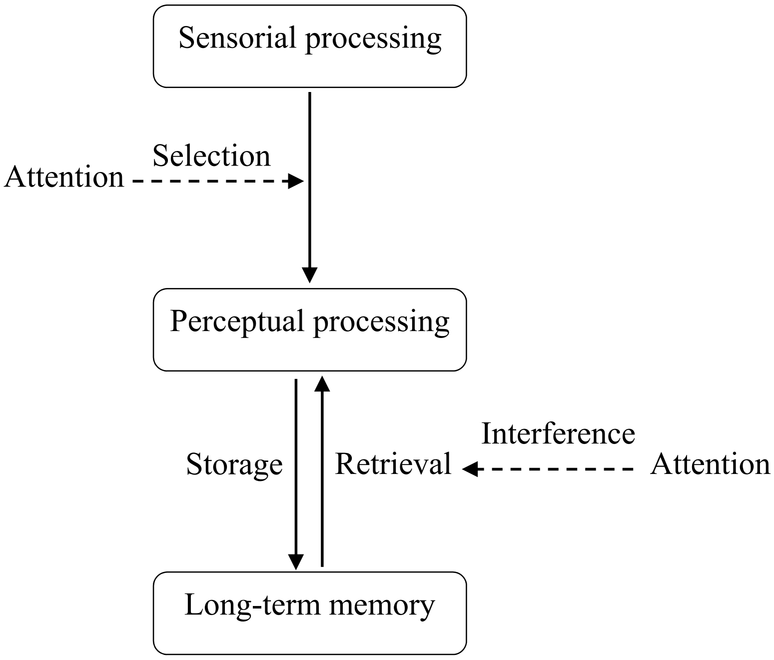


Figure 4. A simplified model of information processing system with pre- and post-perceptual processing locations for controlled attention.

As for the first premise, it is to note that perceptual organisation explanations similar to bottom up processing of information have appeared originally in the works of the Gestalt psychology. Drawing examples from visual perception and focusing on environmental stimuli, the Gestalt psychologists have emphasised the role of innate laws of organisation in arranging patterns and shaping perceptual forms (Wertheimer, 1938). Equally, some cognitive researchers, suggesting the importance of bottom up processing of information, have emphasised the role of analysis of environmental stimuli and extraction of their physical features in pattern recognition (Cole & Scott, 1974; Cole, Rudnicky, Zue, & Reddy, 1980; Scott, 1980). In an experiment to understand the process of spectrogram reading by a spectrogram[7] expert reader, Cole, Rudnicky, Zue, and Reddy (1980) reported that an expert

---

[7] A spectrogram is a graphic display of spectral features (spectrum of frequencies) of sounds as a function of time.

reader could identify 85% of all phonetic segments from spectrograms of fluent speech. They suggested the implication that sounds were accompanied by specifiable acoustic features and could be perceived *directly* from the information in the acoustic signal (p. 43). Similarly, borrowing much from the Gestalt psychologists in a series of experiments stressing the importance of the relations among acoustic components of the speech signal as a source for information for perception, Scott (1980) claimed,

> Acoustic signals, as they exist in our physical environment, are frequency-integrated, complex waveforms. The auditory system, like all perceptual systems, strives to maintain as accurate a representation of the physical environment as possible. (p. 70)

On the other hand, other cognitive researchers, suggesting top down approaches of information processing, have emphasised the role of an individual's already formed memories and expectations in directing their perceptions, i.e. an individual may easily recognise perceptual forms and patterns because an individual expects them to occur in certain locations (Bond & Garnes, 1980). In an attempt to explicate the functions of fluent speech perception system by examining the nature of its failures, Bond and Garnes (1980) revealed interesting findings about the involvement of the phonological, lexical, and semantic aspects in understanding. They emphasised,

> Listeners actively employ grammatical information in speech perception—on phonological, lexical, and sentence levels. Listeners are aware of "fast speech" rules and compensate for either real or supposed phonological reductions which have affected lexical representation. Listeners are also aware of the segments and sequences possible in language, and "hear" only those that are possible. (Bond & Garnes, 1980, p. 128)

A third group of cognitive researchers has argued for the bi-directionality of speech processing, suggesting the latter to be the result of an interaction between lower- and higher-order elements. Considering the insufficiency of sounds in their physical form for perception of speech, cognitive researchers have focused on the parallel processing of different linguistic information, comprising simultaneous sensorial and semantic processing. Sensorial processing involves extraction of relevant physical features of stimuli or their acoustic features, while semantic processing involves syntactic, lexical, and morphological processing. Coordination in processing of structurally and functionally appropriate information is essential to perception, and is subject to processing constraints. In this respect, Cole and Jakimik (1980) proposed an attractive speech perception model specifying orderly stages capable of simultaneously integrating lower- and higher-order elements of speech.

> Words are recognized through the interaction of sound and knowledge. The words in an utterance are recognized one after the other. A word's recognition provides syntactic and semantic constraints that are used to recognize the following word. A word's recognition locates the sounds which begin the following word. The sounds in a word are processed sequentially. A word is consciously recognized when the sequential analysis of sound eliminates all word candidates but one. (Cole & Jakimik, 1980, p. 161)

1.3.3 Theories of attention.

Most cognitive researchers have considered different theories of attention emphasising the importance of the cocktail party[8] phenomenon in implying processing capacity limitations, and have equally proposed theoretical models to account for attention, with varying claims in terms of serial or parallel information processing and fate of unattended stimuli. Attention theories can be classified into two classes along a practical, chronological dimension: classical theories and contemporary ones.

1.3.3.1 Classical theories of attention.

Classical theories of attentions have been proposed by Broadbent (1958), Deutsch and Deutsch (1963), and Treisman (1964). Broadbent hypothesised about the presence of a bottleneck in one of the stages of information processing, the role of which is to preclude

---

[8] The cocktail party phenomenon is inherent in attention theories. In the latter, psychologists use the expression to refer to the phenomenon that the environment bombards individuals with a large amount of stimuli that, given the substantial empirical evidence reported by individuals and research, is impossible to attend to all at once.

completion of successful processing of information. We think that a straightforward presentation of Broadbent's (1958) theory is a prerequisite to the understanding of the rationale of the limited capacity view of attention. Working quintessentially on data obtained from recall tasks on dichotic listening experiments[9], Broadbent (1958) developed a "Filter Theory" (p. 42) to account for a limited capacity of attention that is responsible for selection of information to be processed. Broadbent's (1958) theory supposes the existence of a sensorial buffer at the entrance to the nervous system responsible for selection of a type of stimuli to pass for processing. In accounting for attention, the *Filter Theory* demonstrates a specific flow of information and allocation of processing resources, implying the short-term memory structure (Broadbent, 1958, p. 216). Equally, the Filter Theory makes the following assumptions: (a) simultaneous stimuli encoded by different modalities access the sensory buffer in parallel; (b) based on the stimuli with the most prominent physical characteristics, the filter allows of only one type of stimuli to pass for processing, while other stimuli are retained in the buffer for future processing; and (c) due to the limited capacity of the short-term memory, the filter carefully processes information to prevent information overload in the short-term memory.

Several researchers have reviewed Broadbent's (1958) model of selective attention on the ground of the latter's shortcomings in accounting for the existence of more than one level of selection (Malim, 1994, p. 16). They have recognised the plausibility of the selective filter to account for selective attention, but have re-examined the validity of the physical characteristics as a basis for selection of information. In a recall task, Gray and Wedderburn

---

[9] In a dichotic, listening experiment, the experimenter presents subjects with two materially different sets of auditory stimuli, with one set to hear at the right ear and the other set to hear at the left ear. Broadbent (1958) used this type of experiment in presenting two sets of three digits each in a recall task and found that subjects recalled the digits ear by ear and not in the order in which he presented them. According to Broadbent (1958), the order of recall suggested the use of a physical feature, i.e. ear of arrival, as a basis for selection of information by the filter.

(1960) employed a dichotic, listening experiment in which they presented subjects with interspersed messages presented to whichever ear. Their experiments material included lists of words and digits arranged as in the following example “Mice 5 cheese” presented to the left ear and “3 eat 4” presented to the right one (Gray & Wedderburn, 1960, p. 182). The key finding of the experiment was that recall of messages was in the order “Mice eat cheese” and “3 5 4”, suggesting selective attention to occur also after semantic processing of information rather than before as claimed by Broadbent (1958). Gray and Wedderburn (1960) concluded that, “the ear of arrival is only one possible cue for grouping…Subjects are simply using whatever cues are available to interpret sensory events” (p. 184).

Some researchers have argued that the selective filter may not comprehensively explain the cocktail party phenomenon, where an individual with focused attention on one conversation can still pick relevant information from an unattended conversation. For instance, in a series of dichotic, listening experiments designed to investigate the storage of irrelevant message during selective attention to one of two dichotic messages, Treisman (1964) obtained findings suggesting that processing of both attended and unattended messages require identification of words and their meaning rather than simple identification of sounds. Treisman’s (1964) findings supported her statement that “the filter acts by attenuating rather than blocking irrelevant signals” (p. 459), suggesting a model of selective attention called the *Attenuation Model*.

Deutsch and Deutsch (1963) stated that the concept of the filter, as proposed by Broadbent (1958), was too simplistic to account for selection of wanted from unwanted information in received messages in speech. They suggested a model of selective attention deriving some of its essential elements form the theory of learning and motivation, claiming that some “degree of general arousal is thus necessary for attention to operate” (Deutsch & Deutsch, 1963, p. 84). One of the salient characteristics of their suggested model is that all sensory messages impinging upon the organism are processed at both sensorial and semantic levels, and that the limits of attention apply only to consciousness, memory, and responses. A second key characteristic of the proposed model is that selective attention is a two-dimensional function of message importance or pertinence and the organism's level of arousal, which fluctuates between asleep, drowsy, and alert positions.

1.3.3.2 Contemporary theories of attention.

Recent theories of attention have focused on the conceptual distinction between two automatic and controlled cognitive processes involved in different types of tasks (Treisman, 1964; Schneider & Kintz, 1967; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977; Shiffrin & Schneider, 1984; Treisman & Gelade, 1980; Chow, 1986; Logan, 1988; Logan, 1992; Logan, Taylor, & Etherton, 1999; Schneider & Chein, 2003). Highly frequent tasks require automatic, parallel processing of encoded input, and infrequent, difficult tasks necessitate controlled, serial processing. As against automatic processing, controlled processing requires significant consciousness.

In two auditory shadowing experiments[10] investigating the role of attention in efficiency of performance in a task requiring verbal responses to one or two simultaneous auditory messages, Treisman (1964) identified three functional stages for input processing: (a) a compulsory stage of processing for all inputs based on general physical characteristics of the sounds, (b) a selective discarding or attenuation of inputs by a filter in case of concurrent inputs or overloaded messages, and (c) the identification of words and meaning carried out only for selected messages (p. 459). According to Treisman's (1964), the findings suggest the existence of two qualitatively different information processing mechanisms, one of which is compulsory

---

[10] In an auditory shadowing experiment in attention research, auditory messages are presented to subjects, either monochotically or dichotically, and are asked to attend to and repeat back continuously only a relevant message.

and unconscious, required for the processing of all inputs, and another conscious mechanism, responsible for processing of selected ones. The differences between the suggested input processing mechanisms are concomitant with top down and bottom up processing of information, driven by semantic interpretation and physical characteristics of inputs, respectively.

Combining ideas from both analytic and synthetic theories of perception[11], Treisman and Gelade (1980) suggested an account of attention involving focused and divided attention called "feature-integration theory" (p. 98). Their model assumes a two-mechanism model of attention responsible for early, pre-attentive (automatic), and in parallel processing of encoded information, and a late, consciously focused attention stage of information processing in charge of objects' recognition and identification. Treisman and Gelade (1980) stated,

> the immediacy and directedness of an impression are no guarantee that it reflects an early stage of information processing in the nervous system. It is logically possible that we become aware only of the final outcome of a complicated sequence of prior operations. "Top down" processing may describe what we consciously experience; as a theory about perceptual coding it needs more objective support. (p. 98)

In a similar vein, Schneider and Shiffrin (1977), and Shiffrin and Schneider (1977) offered classic articles on the distinction between automatic and controlled information processing mechanisms that require different levels of attention, and the role of consistent practice in reducing attention necessary for performing tasks. Schneider and Shiffrin (1977) and Shiffrin and Schneider (1977) suggested a two-process theory of human information

---

[11] Analytic and synthetic theories of perception represent the controversy between the Gestalt psychologists and associationists' view of experience or perception of complex wholes. For the Gestalt psychologists, the perception of the whole precedes its parts, and that, only when needed, individuals proceed in analysing objects into their constituent elements or features (best formulated in the dictum, *all is more than sum of its parts*). However, for associationists, the claim is that individuals experience the whole by combining its constituent elements (formulated in the dictum, *all is the sum of its parts*).

processing applied to attention phenomenon, controlled and automatic processes. Drawing the basic distinctions between the suggested processes, Schneider and Shiffrin (1977) stated,

> Automatic processing is activation of a learned sequence of elements in long-term memory that is initiated by appropriate inputs and then proceeds automatically—without subject control, without stressing the capacity limitations of the system, and without necessarily demanding attention. Controlled processing is a temporary activation of a sequence of elements that can be set up quickly and easily but requires attention, is capacity-limited (usual serial in nature), and is controlled by the subject. (p. 1)

In a series of experiments, Schneider and Shiffrin (1977) employed a same-different recall task in two experimental conditions, called consistent and varied mapping, to investigate controlled, visual search and automatic, visual detection tasks, i.e. *attention demanding vs. no attention demanding tasks*. In consistent mapping condition, qualitatively similar items (digits and consonants) were utilised either as a memory set or as distractors. In other words, if a memory set contained digits, then distractors would contain only consonants, and vice versa. In varied mapping condition, qualitatively similar items were utilised simultaneously as a memory set and as distractors, i.e. if a memory set contained digits, then distractors would contain digits too, and vice versa. Frames were presented to subjects who had to decide whether any of the presented items was identical to any of the presented items in the memory set, while encouraged to maintain high accuracy expressed in *hits* and fast responses expressed in *reaction time*. Manipulating memory set size and frame size on several series of trials, Schneider and Shiffrin's (1977) findings demonstrated a significant effect of memory load (memory set size and frame size) on accuracy level of subjects' responses, and found the performance in all consistent mapping conditions were much better than in the easiest varied mapping ones (Schneider & Shiffrin, 1977, p. 14). Accuracy level findings led the authors to suggest two probable conclusions: (a) the use of automatic detection and controlled search in consistent and varied mapping conditions, respectively, and (b) the ability of using automatic detection may develop as a function of training procedure in which stimuli can be consistently mapped to responses. Further, the latter may occur with increasing practice that takes a great number of trials. However, reaction time findings did not provide enough evidence for the authors to assume uniformity of attention mechanisms across experiments. The authors suggested that reaction time might depend on the relative importance given to responding quickly or accurately (Schneider & Shiffrin, 1977, p. 32). Schneider and Shiffrin (1977) concluded,

> Automatic processing is learned in long-term store, is triggered by appropriate inputs, and then operates independently of the subject's control. An automatic sequence can contain components that control information flow, attract attention, or govern overt responses. Automatic sequences do not require attention, though they may attract it if training is appropriate, and they do not use up short-term capacity. They are learned following the earlier use of controlled processing… automatic detection develops when stimuli are consistently mapped to responses; then the targets develop the ability to attract attention and initiate responses automatically, immediately, and regardless of other inputs or memory load. (p. 51)

Carrying on their laborious work, Shiffrin and Schneider (1977) extended their initial experimental effort to: (a) demonstrate the qualitative differences between the suggested modes of information processing, (b) investigate the probable effect of perceptual learning in the development of automatic processes, and (c) understand the role of categorical perception in the controlled attention and in the development of automaticity. Comparing their suggested, two-process theory of information processing with existing theories of attention, Shiffrin and Schneider's (1977) emphasised that: (a) encoding of information and feature abstraction is subject to minimal control; and (b) conscious filtering, blocking, or attenuating occurs after perceptual processing; and (c) training conditions are vital to the development of automatic processing and detection. Some of their research findings suggest that: (a) an automatic response can develop with practice in the absence of previous information categorisation; (b) an automatic response is resistant to change; (c) an automatic response can be unlearned and a new automatic response be learned, but only after a highly significant retraining; (d) information categorisation improves controlled processing; and (e) information categorisation improves controlled processing and speeds up the acquisition of automatic processing. In a recent work on the theory of automatic and controlled processing of information, Schneider and Chein (2003) suggested a distributed, computational model for processing information that involves modules responsible for categorisation, buffering, association, and encoding prioritisation of information, recognising the fact that: (a) "controlled and automatic processing are complementary modes of behavior supported by different processing architectures" (p. 555), and (b) that there is "a developing synergism between the behavioural, computational, and biological interpretations of dual processing theory" (p. 555).

1.3.3.3 Instance theory of automatisation.

Though widely admitted to have outlined the main features of automaticity, Schneider and Shiffrin (1977) and Shiffrin and Schneider's (1977) findings have been occasionally reviewed by several researchers on the ground that (a) their theory provides mere description of automatic and controlled processes, but does not explain the way a controlled process becomes an automatic one as a result of great practice in a consistent condition; and (b) that their theory fails to explain the interference effect reported by Stroop (1935) in his experiments[12] (Logan, 1988; Logan, 1990; Logan, 1992; Logan, Taylor, & Etherton, 1999). Logan (1988) proposed a theory of attention aimed to account for the nature of automatisation and the way it develops as a function of consistent practice and memory. Logan (1988) stated,

> automatization is construed as the acquisition of a domain-specific knowledge base, formed of separate representations, *instances*, of each exposure to the task. Processing is considered automatic if it relies on retrieval of stored instances, which will occur only after practice in a consistent environment. Practice is important because it increases the amount retrieved and the speed of retrieval; consistency is important because it ensures that the retrieved instances will be useful… The focus on learning avoids many problems with the modal view that stem from its focus on resource limitations. (p. 492)

---

[12] Stroop (1935) investigated the effect of an interfering colour stimulus (except for black) on reading the name of a colour other than the one the name denotes, e.g. the word *blue* printed in *red*. Stroop (1935) found that the colour in which the word was printed slowed down the reading of the word. Such an effect bears the name of the investigator, the *Stroop Effect*.

Three main assumptions establish the instance theory of automatisation: (a) mandatory encoding, (b) mandatory retrieval, and (c) instance representation. For encoding, the theory assumes that every attended stimulus is encoded mandatorily in the long-term memory, and the quality of encoding depends on the stimulus conditions and attention constraints. For retrieval, the theory assumes that retrieval of information from long-term memory is an unavoidable, mandatory consequence of attention, and an attended stimulus is enough to retrieve any information associated with it in the past. The theory assumes that a stimulus is encoded, stored, and retrieved separately every time it is received, regardless of the number of times it has been encountered (Logan, 1988; Logan, Taylor, & Etherton, 1999). As suggested by its assumptions, the instance theory explains automaticity development as a function of task practice in a consistent environment, and that repeated exposure to consistent task environment creates memory traces, forming a task-specific knowledge base. As such, automaticity is not only reflected in information processing occurring during learning and information encoding, but is equally a pure memory phenomenon. Automaticity is abandonment of a step-by-step information processing in favour of a single-step retrieval of information from memory, causing a significant reduction in demanding cognitive resources, specifically attention and processing time (Logan, 1992, p. 909). Simply put, automatic performance is better viewed as a function of learned information and retrieval.

Conclusion

Regardless of the existence of various conceptualisations of automaticity, there seems to be a basic agreement among psychologists to involve certain rudimentary features of it, involving (a) regularity, (b) speed, (c) attention, and (d) memory. The concept denotes regular, fast processes of encoding information and forming of long-term memory representations of encoded input. With early practice within a consistent environment, mapping of input onto representations becomes an involuntary process that is less demanding for awareness and attentional resources, resistant to change, and that predominantly develops in an early stage of life. Therefore, automaticity is a process that manifests itself in information-processing routines or behavioural responses that depend more on long-term memory representations of encoded input than on its received characteristics, i.e. physical properties. However, the extent of unlearning these routines for processing information and the establishment of new ones with training and experience is also possible, as research findings and firm scientific evidence prove inconclusive either way.

For the purpose of our current research, we conceive of automaticity as a process of appraisal of information that varies as a function of experience. Whether automaticity involves no attention or selective attention, we argue that the involvement of the latter in appraisal of information does not necessarily reflect choice or volitional control as much as it reflects a product of long-term, consistent experience. For instance, it is uncommon to believe that infants exercise some sort of volitional control over their attentional resources to perceive the sounds of their native language. However, with long-term language experience, infants come (a) to establish highly overlearned speech perception routines that are capable of detecting meaning in their native language speech, and (b) to use these speech perception routines often unconsciously. Conscious use of these speech perception mechanisms occurs in case of failure of extraction of the linguistic message. We further operationalise automaticity in terms of behavioural response patterns, involving basic discrimination and identification perceptual tasks, as a promising revealer of the nature of perception mechanisms.

As the focus of the first part present research relates quintessentially to the application of automaticity in the field of speech perception, the following chapter will introduce and review the various concepts employed in speech perception research, cross-language speech perception theories, and the controversial issues raised thereof.

Chapter 2

Speech Perception: Literature Review

Introduction

Speech has been a centre of research interest among psychologists and linguists for numerous reasons, the most important of which is the extraordinary ability individuals possess in associating meaning with a stream of sounds, the nature of which is still debated. Speech researchers have formulated theories on the nature of speech, its acquisition mechanisms, and its development over an individual's life span. Theoretical assumptions and empirical considerations, in both linguistics and psychology, have yielded inconsistent answers to account for speech phenomenon, employing varying hypotheses, frameworks, and terminologies. This intricate research condition has broadened speech investigation with joint multidisciplinary efforts to provide answers to fundamental issues pertaining to larger questions about language learning. Hereinafter, we offer a background for our research questions and hypotheses, introduce basic common concepts utilised in speech research by both psychology and linguistics in relation to the ontology of the sound and speech underlying mechanisms. Then, we review some of the most conspicuous suggested explications and models accounting for cross-language speech perception, with exclusive focus on segmental aspects of the language.

2.1 Nature of the Sound

Research on speech perception and production and their elements has raised several issues, initially involving philosophical debates over mind, matter, and perception (Smith, 2009). As O'Callaghan and Nudds (2009) put it,

> Debates about the nature of sounds have focused upon such questions as whether sounds are mind-dependent or mind-independent, whether they are individuals or properties, and whether they are object-like or event-like. Also, there has been considerable debate about just where sounds are located. (p. 4)

The initial issue in investigating speech concerns the nature of its constituent elements, or what theorists conventionally consider as sounds, being the immediate and proper objects of audition, for it is uncertain that an individual "could hear something without hearing a sound"[13] (O'Callaghan & Nudds, 2009, p. 5). To explicate the nature of the sound phenomenon, philosophers have debated three views about the sound: (a) the "*Property View*", (b) the "*Wave View*", and (c) the "*Event View*" (O'Callaghan, 2009, pp. 27-36).

2.1.1 The Property View.

Within the '*Property View'*, the sound is an attribute of the object that produces it. This stands in a similar situation where colour and figure are attributes of objects an individual perceives through their eye. Therefore, the sound is an attribute to the object or body that produces it when the object or body vibrates at a certain frequency and amplitude. In the case of speech, the sound is an attribute of the speech organs that is located at their origin "rather than as filling the air" (Pasnau, 1999, p. 313). Liberman and Mattingly (1985, 1989), Liberman (1996), and Fowler (1986, 1989) repeatedly engaged the '*Property View*' of the sound to investigate the proper objects of speech perception. Nonetheless, the main argument held against the '*Property View*' of the sound is that it does not explicitly assume the necessity for a transmitting medium for the sound. This argument is known as the '*vacuum argument*'.

2.1.2 The Wave View.

The '*Wave View*' posits the very standard version or the received view on the nature of the sound. The latter is viewed as train of waves generated by air disturbance and that moves

---

[13] It is to note that some philosophers disagree with the view that the sound is the object of auditory perception. Some philosophers have argued that it is possible to hear silence that does not necessarily involve the sound, illustrating their case with few, relevant examples of deaf and hearing individuals as well. Exemplary cases, involving multimodal perception, help understand this view, where a complete account of an individual's perception in a single modality may equally involve other modalities (for more detail on this unconventional idea, see Sorensen, 2009).

through the surrounding environment as a longitudinal compression wave (O’Callaghan, 2009). That is, an individual immediately hears the sound as neither an object or a body nor a feature of any of them; the individual rather hears the sound as a pattern of pressure variation, which constitutes the wave disturbance in the surrounding environment (Searle, Jacobson, & Kimberley, 1980). Within the ‘*Wave View*’, just as the sound is a wave caused by an object that originates in a relatively *distal*, *determinate* location, an individual does not perceive it as travelling through the air as waves conventionally do. Stated differently, an individual locates distally both the sound as a wave and its source, and hears the sound where it occurs rather than where it travels.

The main issue with this view is that it avoids dealing with sound locatedness problems such as echoes, sound interference, and sound transmission. If the ‘*Wave View*’ does not explicitly suppose the existence of laws for possible perceptual errors in locating sounds in different hearing conditions, it will pose serious issues of auditory illusion and veridical recovery[14]. That is, the sound may never be precisely located and accurately perceived, for it is a movement of air itself (O’Callaghan, 2009, p. 28). The movement of the air and change in pressure (albeit neglected) across time (albeit very short) offer good reason to think that initial attributes of the sound, as a distal stimulus (near its emitting source), varies from its attributes, as a proximal stimulus (near its receiving organism).

2.1.3 The Event View.

The ‘*Event View*’ of the sound thoroughly considers the issues of auditory illusion and veridical recovery and offers plausible answers by conceptualising the sound as an event that has spatial, spectral, and temporal patterns. This view postulates that the sound is a relatively

---

[14] Veridical recovery is a long-held debate in psychology of perception. It is about whether it is possible for an individual’s senses to recover perfectly (without distortions) all information available in stimuli in the surrounding environment in service of their organism. It is also termed “the inverse problem” (Kluender & Kiefte, 2006, p. 155).

stable, temporal event as against its attributes that persist less for change over time. O’Callaghan (2009) put it,

> sounds are particular events of a certain kind. They are events in which a moving object disturbs a surrounding medium and sets its moving. The strikings and crashings are not the sounds, but are the causes of sounds. The waves in the medium are not the sounds themselves, but are the effects of sounds. Sounds so conceived possess the properties we hear sounds as possessing: pitch, timbre, loudness, duration, and … spatial location. When all goes well in ordinary auditory perception, we hear sounds much as they are. (p. 28)

Within this view, the sound is “distally located and stationary relative” (O’Callaghan, 2009, p. 49) to its source and is not merely a property of a material object or a body. Given its consideration of the vacuum argument and the locatedness issue, the ‘*Event View*’ seems the most advantageous as against the two previous views. Accordingly, the sound is conceived of as a product of a reconstruction process of an event originating in body organs, mediated by physical medium and having an impact on the auditory system.

2.2 Important Issues in Speech Perception

Speech perception has significantly marked research on language, accounting for the frequency of speech as probably the most frequent mode of communication. Psychologists, linguists, and computer scientists have adopted different research paradigms, invoking issues proper to their own fields, to explicate the complexity of speech perception and the remarkable, probably inherent advantages speech has over writing (Ferreira & Anes, 1994; Remez, 1994; 2005; Liberman, 1998). Though the ultimate goal in speech perception is to investigate processes that listeners use to derive meaning form percepts and represent speech out of a signal, speech perception researchers have proved remarkable disunity in defining the phenomenon and its underlying mechanisms. Numerous accounts of encountered problems in speech research have been offered to explain the constituent elements listeners use in processing the speech signal in order to reconstruct the intended message of the speaker (Nygaard & Pisoni; 1995; Frazier, 1995; Rosenblum, 2005; Bernstein, 2005). These accounts have raised several difficulties in dealing with the complex, dualistic[15] nature of speech, relating to psychology and linguistics. The issues involve: (a) the ontology of the sound and psycholinguistic representation; (b) linearity, segmentation, and lack of invariance; (c) variability and perceptual constancy (speaker variability, variability in speaking rate); (d) perceptual organisation of speech; (e) linguistic experience; (f) biological constraints; and (g) relation between speech perception and production (Clark & Clark, 1977; Nygaard & Pisoni, 1995). Despite the relevance of these issues in speech research, most researchers have predominantly focused their interest on issues of psycholinguistic representation, perceptual organisation of speech, linguistic experience, and biological constraints. Hereinafter, we briefly describe some of the fundamental speech perception problems in relation to psychology and linguistics, with focus on issues that are closely relevant to our research interests. Additionally, we note that, as cognitive traditions have always inspired speech perception research, we employ common terminology used in the field, as being inherent in the topic, with review only when we deem it necessary.

2.2.1 Physical vs. psychological reality of the sound.

Originating in inherently different conceptualisations of ontology of language, natural empiricist and pure mentalistic frameworks have continually guided research on language and shaped controversies between different views on language, its learning, and its relation to mind. With respect to speech as probably the most utilised form of communication, a common view assumes that speech is a pure, physical phenomenon. The latter materialises in observable

---

[15] The dualistic nature of speech refers to the realities of the speech phenomenon: (a) the sound as the physical medium of language produced via mobilisation of specialised biological organs and received by ears, and (b) information as the semantic component in an individual's mind that is not a haphazard combination of sounds. With this respect, Jakobson (1978) stated, "we have known for a long time that a word, like any verbal sign, is a unity of two components. The sign has two sides: the sound, or the material side on the one hand, and meaning, or the intelligible side on the other… But while the fact that there is such a combination is perfectly clear, its structure has remained very little understood" (pp. 2-3).

acoustic, motor, and auditory events that are respectively: (a) the intended motor gestures that form a physiological prerequisite for speech; (b) the acoustic aspects that are temporary, deliberate, modulated air pressure disturbances available for transmission; and (c) the auditory aspect that forms the impressions transmitted, modulated air disturbances impinging on the hearing sense (Wilder, 1975). The three aspects, individually or in concert, are sufficient to explain the communicative act through speech (Lisker & Abramson, 1964; Lieberman & Blumstein, 1988; Diehl & Kluender, 1989; Fowler, 1989; Ladefoged, 1996; Marchal, 2009; Lieberman, 1970; Liberman & Mattingly, 1985; Fry, 1979; Raphael, Bordon, & Harris, 2011; Liberman, 1996; Liberman, 1998; Scruton, 2009). This pure empiricist paradigm has dominated early speech research and focused mostly on the natural and physical laws of speech (Chomsky, 1986, p. 2). Despite providing a substratum of evidence about the physical nature of speech, the natural empiricist paradigm has neglected to explicate the functions of these sounds within a speech system. Jakobson (1978) claimed,

> The fact that linguistic sounds are signifiers was deliberately put aside, for these linguists were not at all concerned with the linguistic functions of sounds, but only with sounds as such, with their 'flesh and blood' aspect, without regard for the role they play in language. (p. 5)

On this ground, speech is not the raw sounds as produced by an individual using speech organs, but is rather a product of a mental activity following the grammar of language, as part of the work of the mind, in service of efficient communication (Jespersen, 1924; Jakobson & Halle, 1956; Chomsky, 1957; Chomsky, 1986; Chomsky & Halle, 1968; Stevens, 2006). In this view, speech is an abstract phenomenon for which sounds, in their concrete base, are simply its accessible vehicles, and should be investigated in relation to its underlying form.

This divergence of views gives no chance to carry research on speech without making prior assumptions on the nature of speech, its components, and their representations as attested by Trubetzkoy (1969), "the various aspects of the speech process are so disparate that their study must be divided into several subsciences" (p. 4). Phonetics and phonology stand as obvious examples of these disciplines. Invoking notions of surface and underlying representations, many speech researchers have substantiated the disagreement between phoneticians and phonologists on how to represent speech (Jakobson and Halle, 1956; Chomsky & Halle, 1968; Trubetzkoy, 1969; Frazier, 1995; Pierrehumbert, 1990; Remez, 1994; Archibald, 1993; Gelfand, 2005; Clark & Clark, 1977). In differentiating the subject matter of in phonetics and phonology, Jakobson and Halle (1956) stated,

> While phonetics seek to collect the most exhaustive information on gross sound matter, in its physiological and physical properties, phonemics, and phonology in general, intervenes to apply strictly linguistic criteria to the sorting and classification of the material registered by phonetics. (p. 7)

Accordingly, in view of phonetics, speech researchers should investigate speech sounds for their concrete properties that are articulatory, acoustic, and auditory that correspond analogously to the three parts making up the elements of a communicative act. Articulatory phoneticians concern themselves with the physiological mechanisms and biological organs responsible for production of speech segments. Therefore, articulatory phoneticians describe speech segments as a train of motor gestures, involving airstream mechanism and mobilisation of various parts of the vocal apparatus (lungs, trachea, larynx, mouth, teeth, lips, etc.) in the modulation of air stream to produce sounds with certain qualities aimed at producing certain effects (Bickford & Floyd, 2006).

Acoustic phoneticians describe sounds in terms of their acoustic properties just the same way they study notes and tones produced by music instruments, and make extensive use of spectrograms that represent frequencies of sound elements of utterances as a function of time. The guiding principle of acoustics is that cavities (pharyngeal, oral, bilabial, and nasal) within the vocal tract are acoustic resonators, i.e. air containers, which vibrate at specific frequencies. To have a simple idea on the sound frequency, it is necessary to involve a simple, straightforward example. Similar to the way a guitar string vibrates and makes a note, the vocal cords vibrate to produce voice. The vocal cords vibrate at a basic frequency differently among man, women, and children, resulting in the fundamental frequency of the voice that is represented as F_0 (Clark & Clarck, 1977; Raphael, Borden, & Harris, 2011). The voice simultaneously produces multiples of this fundamental frequency called harmonics, which is why acoustics call fundamental frequency as the repetition rate of a complex periodic sound. For instance, if the F_0 of a voice equals 250 Hz, then the voice produces harmonics at 500 Hz, 750 Hz, and 1000 Hz, etc. The passage of the vibrations through the pharyngeal and oral cavities enhances some of these vibrations and diminishes others, leading, therefore, to the creation of sound patterns. Spectrogram analysis shows the bands of enhanced frequencies, called formants.

Finally, auditory phonetics concerns itself with the anatomy and function of the peripheral, auditory system (the one separate from the cerebral area responsible for audition) accounting for the way aural stimuli impinge upon the auditory modality as they pass from the outer ear to the middle and inner ears to the auditory nerve, that is, *transduction*.

For phonologists, the material aspect of speech sounds is only the apparent side of the story, and the intended message of an individual should be the real subject of research. Phonologists' interests are linguistically driven and centre upon two general themes: (a) the representational system of the speech raw material that "captures the true 'essence' of the sounds in words and sentences" (Clark & Clark, 1977, p. 188); and (b) the grammar that governs speech underlying forms in relation to its phonetics forms (Chomsky & Halle, 1968). For Jakobson and Halle (1956),

> Linguistic analysis gradually breaks down complex speech units into morphemes as the ultimate constituents endowed with proper meaning and dissolve these minutest semantic vehicles into their ultimate components, capable of differentiating morphemes from each other. These components are termed distinctive feature. Correspondingly, two levels of language and linguistic analysis are to be kept apart: on the one hand, the

> semantic level involving both simple and complex meaningful units from the morpheme to the utterance and discourse and, on the other hand, the feature level concerned with simple and complex units which serve merely to differentiate, cement and partition or bring into relief the manifold meaningful units. (pp. 3-4)

Accordingly, feature analysis of speech in its phonetic realisation in terms of a system of contrast is the most appropriate premise for a functional, semantically driven theory. The Prague School and American Structuralism linguists used feature analysis to describe and classify speech sounds as *phonemes*[16], characterised in a system of oppositions capable of defining lexical contrast (Jones, 1967; Cole & Hualde, 2011). A prominent example of application of feature analysis in linguistic research is the notion of the '*minimal set*' that stands for the slightest acoustic and articulatory difference capable of distinguishing meaning among lexicons.

---

[16] Speech segments are considered units, identifiable and distinguishable out of their atomic elements (articulatory, acoustic, and auditory features, on the one hand, and their functions in the speech system as being semantically relevant or irrelevant, on the other). However, phoneticians and phonologists disagree over the existence of any unitary parallelism between these atomic and semantically relevant, minimal units. Use of terminology such as phone, phoneme, and allophone illustrates the divergence. The term 'phone' is kept for use in phonetic research to "refer to the smallest perceptible DISCRETE SEGMENT of sound in a stream of speech" (Crystal, 2008, p. 361) or to "designate the sum of all distinctive and non-distinctive properties occurring at a specific point in the sound flow" (Trubetzkoy, 1969, p. 37). The term 'phoneme' is rather defined as "a family of sounds in a given language which are related in character and are used in such a way that no one member ever occurs in a word in the same phonetic context as any other member" (Jones, 1976, p. 10). Alternatively, as put by Trubetzkoy (1969), "the phoneme is the smallest distinctive unit of a given language" that "cannot be analysed into still smaller successive distinctive units" (p. 35). Given that phones empirically outnumber phonemes in a language sound system, the term 'allophone', represented graphically like a phone, ensures complete association of phones to phonemes. That is, phones whose differences are irrelevant for linguistic information in speech of a language community are assigned to the same phoneme, accounting, therefore, for the discrepancy between phonetic and phonological representations. Phonetic and phonemic notations substantiate the difference in use of previous terminology in speech studies, by representing phones and phonemes in square brackets [] and slashes //, respectively. In our research, we use the expression 'speech sounds' as a general expression to refer to both phones and phonemes (for a detailed review on the use of term 'phoneme', see Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967, p. 431; Dresher, 2011).

Favouring the underlying structure of language, other linguists have focused their attention on the development and causation of speech as a cognitive ability rather than the final, refined product, seeking to capture the regularities of speech within a general framework that applies to all languages (Chomsky, 1959; Chomsky & Halle, 1968). For example, Chomsky and Halle (1968) attempted to conceive a "universal phonetics" (p. 4) that would describe the universal articulatory features that are available to use for all languages, and the way these features are set to combine and distribute to form the phonetic representation of the lexicon in a specific language. In Chomsky and Halle's (1968) view, the underlying form, i.e. the lexicon or formative, relates to the universally phonetic form (the lexicon pronunciation or surface form), the latter resulting from single- or multi-step mapping. A single-step mapping does not necessitate intermediate stages between underlying and phonetic forms, whereas a multi-step mapping requires a series of intermediate forms, which are set by the language specific, combinatorial and distributional patters, known as phonotactics possibilities.

With focus upon phonetics or phonology, there is compelling evidence not to undermine the usefulness of either in understanding speech. Taking for granted the need for phonetic research, Clark and Clark (1977) argued for an equal need for phonological research, being psychological in character, because of the regularities that characterise speech in a language and the limits or constraints that the latter exhibits in word formation and pronunciation. Expressing a similar view, Pierrehumbert (1990) stated,

> Phonological representation is responsible for describing the qualitative contrasts in sound which can be used to convey qualitatively different meanings in any given language, or in all languages. The entities it posits are attributed to the speaker/listener, since this is where the association between sound and meaning takes place. Phonetic representation is responsible for describing speech as a physical phenomenon. That is, it covers measurable properties of articulation, acoustics and audition. (p. 375)

Moreover, Pierrehumbert (1990) plainly articulated the complementarity between phonetic and phonological representations to allow of a comprehensive understanding of speech, as he stated, "[A] theory encompassing phonology, phonetics and their relation to each other is needed as a foundation for a theory of language processing and language acquisition" (p. 375).

2.2.2 Perceptual units for speech analysis.

The standard view of speech considers the latter a chain of recognisable, discrete, continuous units called phones or phonemes. Straightforward and precise this statement may seem, it is, in fact, deceptively simple. The relative easiness of relevant, linguistic segmentation of informational elements in the written language seems to have no analogous, clear counterpart for speech, for, at least, they relate to two different modalities. The issue of perceptual units for analysis concerns speech research inspired by cognitive tradition as the issue raises the question of the primacy of the perceptual unit for both perception and production of speech, be it the distinctive feature[17], the phone, the phoneme, the syllable, or the lexicon, etc. Speech cognitive mechanisms concern lower- and higher-order levels of speech analysis, therefore, bearing equally upon broad phonetic and specific phonological principles (Foss, Harwood, & Blank, 1980).

As regards phonological principles, for example, Cutler, Mehler, Norris, and Segui (1986) stated,

> Speech segmentation procedures may differ in speakers of different languages. Earlier work based on French speakers listening to French words suggested that the syllable functions as a segmentation unit in speech processing. However, while French has relatively regular and clearly bounded syllables, other languages, such as English, do not. No trace of syllabifying segmentation was found in English listeners listening to

---

[17] In common phonological terms, distinctive features are articulatory features that characterise the minimal speech unit or segment termed the 'phoneme'. The latter is identified as a unitary entity that is the product of a series of articulatory gestures, starting from passage of air in the vocal tract to its release outside of it. Canonical distinctive features include: (a) air stream mechanism (egressive vs. ingressive sounds), (b) air obstruction in the vocal tract (vowels vs. consonants), (c) voicing (voiced vs. voiceless sounds), (d) place of articulation, indicating the physical contact between articulators (bilabial, alveolar, inter-dental, labio-dental, etc.), (and e) manner of articulation, indicating air obstruction of constriction in the vocal tract (stops, fricatives, nasals, laterals, affricate, approximant, etc.).

> English words, French words, or nonsense words. French listeners, however, showed evidence of syllabification even when they were listening to English words. (p. 385)

With respect to syllable stress at the word level, Cutler and Norris (1988) argued, "segmentation at strong syllables in continuous speech recognition serves the purpose of detecting the most efficient locations at which to initiate lexical access" (p. 113). Nonetheless, other speech researchers argue that processing of phonemes and syllables are perceptually contingent on larger linguistic units such as words or phrases (Bever, Lackner, & Kirk, 1969; Rubin, Turvey, & van Gelder, 1976). In a 2-experiment research investigating the effect of lexical information on detection of initial phonemes measured in reaction time, Rubin, Turvey and van Gelder (1976) reported that their subjects detected initial phonemes that appear in meaningful words faster than in nonsense words, suggesting access to lexical information to occur prior to phonological information. Accordingly, given conflicting empirical evidence on the use of various speech units in perception, there is compelling evidence to stand in favour of an interdependence of representations, arguing for no strict hierarchical view of speech perception in which lower-order units are initially processed to provide higher-order units for processing at the next level. It is safer, therefore, to argue that any particular unit of speech depends on attentional requirements of the task, available information, and processing contingencies, as Goodman, Lee, and DeGroot (1994) suggested, "it appears that many levels of linguistic structure provide informative constraints on one another" (p. 3).

2.2.3 Segmentation, linearity and lack of invariance.

As listeners have the impression to perceive speech as a series of discrete, linguistic segments such as phonemes, syllables, words, and so on, their capacity of carving up the continuous speech signal into appropriate units out of which they recover the intended message of the speaker is far from straightforward. As Cole and Jakimik (1980) stated, "[T]he impression that words in fluent speech are separated in time is an illusion. Words exist in the mind of the perceiver—not in the stimulus" (p. 138), it is empirically unclear whether speech signal units are linearly arranged one after another for immediate perception. The presumable linearity of immediately available, speech segments is not found in the physical signal due to speech sounds change resulting from coarticulation and assimilation effects[18], where boundaries between perceived phonemes are illusive. Briefly, coarticulation represents the temporal overlap of acoustic and articulatory properties of adjacent vowels and consonants in natural speech (Delattre, Liberman, & Cooper, 1955), a fact that makes it hard to understand the way listeners systematically segment the speech signal into invariable segments. From the articulatory viewpoint, coarticulation occurs when two articulators are simultaneously moving to produce two different speech sounds. For words such as /pʊt/ and /spuːn/ with initial /CV/ or /CCV/ syllable, where /V/ is a rounded vowel, lip rounding for /ʊ/ or/uː/ can begin initially at the very start of syllable of the word, that is, during the articulation of the word-initial /p/ and /s/, respectively, in case none of the other sounds necessitates an antagonistic articulatory movement (Raphael, Borden, & Harris, 2011, p. 140).

Another phenomenon that argues against the presumable linearity of speech segments involves the modification or alteration in the movement of an articulator, a case known as assimilation. In the usual case, a speech sound follows a planned sequence of articulatory gesture to produce a speech sound with a certain audible and perceptual effect. For example, the production of the consonant [t] involves the following sequential steps: (a) an upward movement of air in the vocal tract through the trachea to the larynx (egressive consonant); (b) no vibration of the vocal folds (voiceless); (c) a blockage of the nasal cavity (oral); (d) raising of the tongue tip to articulate with the alveolar ridge and completely block the air (alveolar); (e) build-up of a short-lasting pressure in the oral cavity (stop); and (f) quick burst of air

---

[18] It is to note that coarticulation effect describes the temporal overlap between articulatory gestures for consonants and vowels, and is not equivalent to assimilation that describes the effect the articulatory gestures of a consonant or a vowel has on the articulatory gestures of another (Raphael, Borden, & Harris, 2011, p. 139; Crystal, 2008, pp. 39-40).

(plosive)[19]. During assimilation, the movement of the articulator takes a shortcut and modifies one or all of the articulatory steps to match that or those of a preceding or following, neighbouring sound, resulting in partial or complete assimilation and anticipatory or carryover assimilation. The assimilation may bear upon the place of articulation as in '*insert this*', where the articulation of the sound /t/ is modified by lowering the tip of the tongue to articulate against the upper incisors instead of the alveolar ridge, allowing a more efficient move to the place of articulation of the following /ð/ that is a dental sound. The slight modification of the place of articulation results in the production of a dentalised /t/, constituting a partial assimilation that is a phonetic (or allophonic) change, semantically irrelevant (Raphael, Borden, & Harris, 2011, p. 137). Another frequent example of modification of place of articulation concerns the tongue-palate, physical contact in the production of consonant /t/ in the word 'talk' as against its production in the word 'tick', being backward and forward, respectively.

2.3 Theories of Speech Perception

It is practical to consider rudimentary assumptions of speech perception theories to describe and review them. The available literature on speech perception shows that proposed theories classify into conventional and less conventional ones. As the less conventional theories posit that humans perceive speech as a series of articulatory gestures, the conventional theories posit that speech perception is part of a general auditory and multimodal learning mechanism (Liberman, 1992, 1998; Diehl, Lotto, & Holt, 2004; Galantucci, Fowler, & Turvey, 2006; Raphael, Borden, & Harris, 2011). Hereinafter, we briefly introduce the preliminary,

---

[19] It is usually sufficient in identification of speech sounds to describe either the manner the air builds pressure inside the vocal tract or the way it bursts out, but not both. While the former is an acoustic index of energy (or air pressure) forming inside the vocal tract below and above the concerned articulators, the latter is an articulatory description of the physical contact between articulators that causes either a complete or incomplete obstruction of air in the vocal tract. The common terminology used is that of Jakobson and Halle (1956), for whom speech sounds are identified as either discontinuous or continuant (p. 30). Common recent vocabulary includes terms such as plosives, non-continuants, and stops (Chomsky & Halle, 1968).

psychological theories that set the theoretical background for cross-language speech perception research.

2.3.1 The Motor Theories.

2.3.1.1 The Motor Theory for Alvin Liberman.

Liberman at the Haskins Laboratories contributed the most to the development of the original motor theory of speech perception that, since then, several speech researchers have continually reviewed and modified in light of supporting empirical evidence. Speech perception researchers refer to the original work of Liberman as the *Motor Theory* (MT) of speech perception, and refer to further versions as the Direct Realist Theory (DRT) of speech perception (Diehl, Lotto, & Holt, 2004). The Motor Theory of speech perception views the speech sound as a property of its emanating body, suggesting an innate, modular[20], species-specific mechanism to be responsible for speech. This innate speech module functions as a speech decoder (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, 1970, 1992, 1998; Liberman & Mattingly, 1985). The essential claim of the Motor Theory of speech perception is that speech is a special code whose elements are not the commutable phones or phonemes but articulatory gestures. The fusing nature of articulatory gestures and variability of acoustic features make it improbable to think of a one-to-one correspondence between speech entities and the supposedly, linguistically perceived phonemes, rendering the belief in their physical and temporal discreteness much of an illusion than of a tangible fact. Thinking of speech as a highly efficient code for communication rather than an alphabet or a cipher[21], Liberman (1998) posited,

---

[20] In the Motor Theory of speech perception, the specialised module is simply one among many special modules responsible for other perception systems that are biological endowment of humans. The speech module enjoys no special status compared to others.

[21] In its basic form, a cipher is a system for symbolic representation of a message in a straightforward, invariant, one-to-one relation between the perceived unit and its symbolic representation. For example, if letters of the alphabet (the plaintext units) are represented in

> the ultimate constituents of speech are not sounds, but articulatory gestures. Having evolved exclusively in service of language, they form a natural class, a phonetic modality. Being phonetic to begin with, they do not require to be made so by cognitive translation. And that, very simply, is the advantage of speech over writing/reading. Speech has the corollary advantage that it is managed by a module biologically adapted to circumvent limitations of tongue and ear by automatically coarticulating the constituent gestures and coping with the complex acoustic consequences. But a result is that awareness of phonetic structure is not normally a product of having learned to speak: the module "spells"—that is, sequences phonetic segments—for the speaker and recovers the segments for the listener, leaving both in the dark about the way it is done; the gestural representations are immediately phonetic in nature, precluding the cognitive translation. (p. 111)

Fitting the concept of modularity as developed by Fodor (1983), the Motor Theory of speech perception assumes the existence of a special speech closed module responsible for encoding and decoding speech sounds in a different manner from perceiving non-speech sounds that are processed in a rather general, auditory open module (Liberman & Mattingly, 1989). The Motor theorists argue that humans are naturally equipped with a tuned, auditory system to perceive speech sounds exactly the way humans produce them, having dedicated speech perception-production mechanisms for recognising intended, articulatory gestures and recovering acoustic

---

serial numbers (the cipher-text units) in the form (a, 1), (b, 2), (c, 3), (d, 4), (e, 5), etc., each word will have a unique representation in the form (cab, 312), (bed, 254), (bad, 214), etc. The cipher-text unit will not change with the change of the neighbouring units of the plaintext unit (Raphael, Borden, & Harris, 2011). That is, whether 'b' is followed by 'e', 'a', or 't', its representational cipher-text unit remains the same. However, a code is a more complex system for symbolic representation that uses, for example, cipher units along with further detail on their combinatorial features. Using the same example mentioned above, it is visibly clear that, in a code, the letter 'b' in 'bed' would not be only represented by the number '2' for it needs to take another form to embody its combinatorial feature when it is followed by 'e' in 'bet' which is perceptually presumably different from it when the letter 'b' is combined with 'a' as in 'bat'. Regarding whether phonemes are enciphered alphabetically, Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) claimed, "[T]here are reasons for supposing that phonemes could not be efficiently communicated by a sound alphabet—that is, by sounds that stand in one-to-one correspondence with the phonemes. Such reasons provide only indirect support that speech is a code rather than an alphabet. They are important, however, because they indicate that the encoded nature of speech may be a condition of its effectiveness in communication" (p. 432).

patterns. First, for the Motor Theory proponents, the speech-closed module is a specialised neural structure that compares parametric descriptions of the acoustic signal with analogous descriptions of production processes that incorporate “complete information about the anatomical and physiological characteristics of the vocal tract” (Liberman & Mattingly, 1985, p. 26). Second, acoustic patterns materialise in redundant elements, making speech perception efficient under the most difficult conditions that involve significant inter-speaker and intra-speaker variability. As humans hardly ever produce speech sounds in isolation, speech sounds overlap and significantly influence each other in a way that causes the lack of invariance, improbable segmentability of the acoustic signal, and high transmission efficiency. On these grounds, some speech researchers believe that humans encode speech sounds rather than encipher them in acoustic cues (Delattre, Liberman, & Cooper, 1955; Liberman, Harris, Hoffman, & Griffith, 1957; Raphael, Borden, & Harris, 2011, p. 244).

For example, Liberman et al. (1967) illustrated the complex mapping between phonemes and their acoustic realisations as perceived by listeners in synthetic syllables as a function of varying formant frequencies at the syllable onset. Invoking the example of the phoneme[22] /d/, being one of the first phoneme-like segments that appears in the vocalisations of a child, Liberman et al. (1967) accounted for the role of formant transitions and context effects as cues for speech sounds. As for formant transitions, Liberman et al. (1967) stated that first formant transition “is a cue for the perception of manner and voicing” (p. 435); however, “transitions of the second formant carry important information about the place of production of most consonants” (p. 435). As for context effect, Liberman et al. (1967) demonstrated the way neighbouring speech sounds would cause variation in the acoustic properties of each other without affecting their perception as individual phonemes. *Figure 5* shows spectrographic

---

[22] We do not use the term ‘phoneme’ here to express our standpoint on the appropriateness of phonological representation of the speech sound, we simply use it as originally employed in the authors’ work in review.

patterns of the phoneme /d/ as a short burst of about 50 ms, with relatively a fixed, first formant transition and a variable, second formant transition as a function of the following vowel. Liberman et al. (1967) argued for a very complex relation between a perceived consonant and its acoustic cues that considerably vary as a function of both the consonant pairing with different vowels and its position with respect to the same vowel as well. For example, the acoustic cues of the perceived phoneme /d/ in /dɪ/ are different from those of /d/ in /ɪd/, whose acoustic cues are also different from those of /d/ in /ædɪ/.

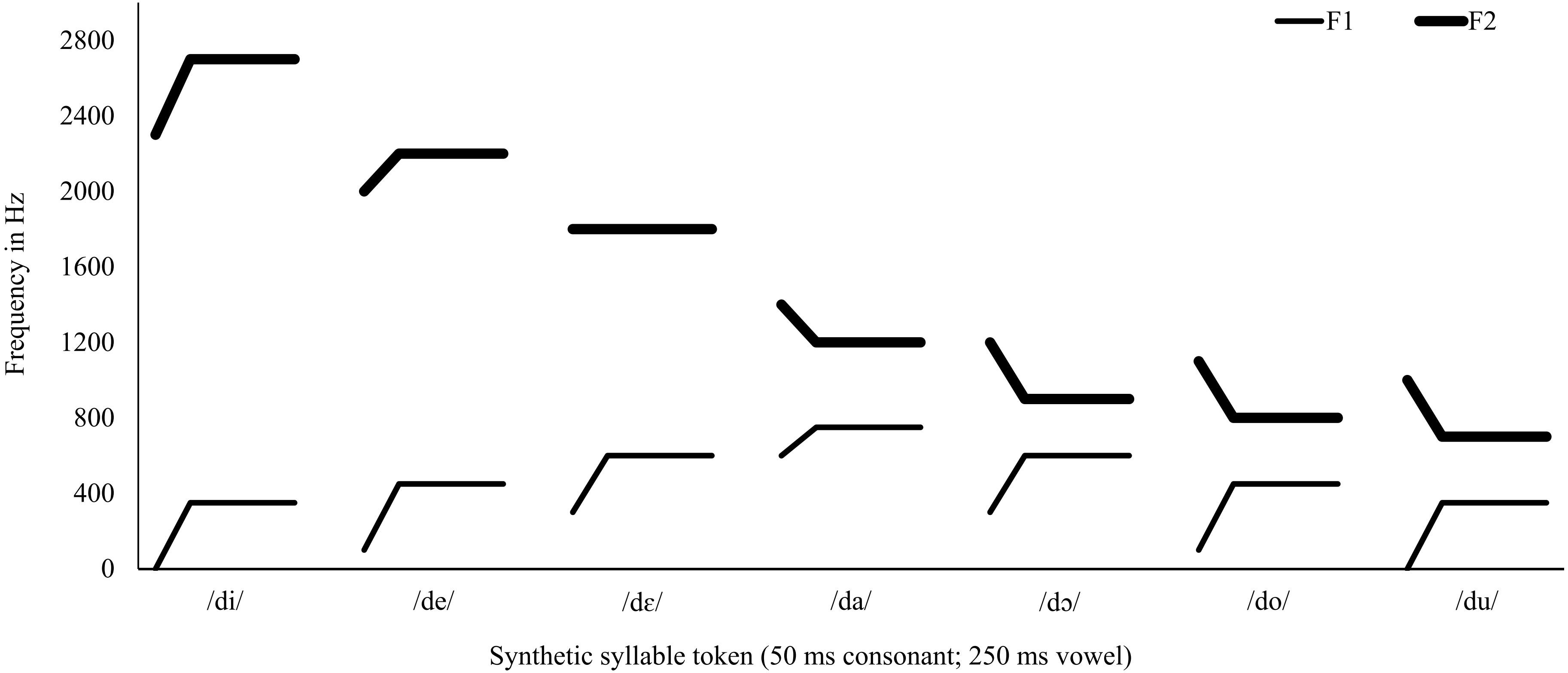


Figure 5. Two-formant patterns for simplified, synthetic syllables with an initial /d/ at the syllable onset. The steady state, formant transitions that take relatively a longer time are vowels. Notice the rise and fall in the transition of F_2 along the syllable tokens. For example, in /di/, the second formant transition rises from 2200 Hz to 2600 Hz, and, in /du/, the second formant transition falls from 1200 Hz to 700 Hz. From "*Perception of the Speech Code*" by A. M. Liberman, F. S. Cooper, D. P. Shankweiler, & M. Studdert-Kennedy, 1967, p. 436. Copyright 1967 by the American Psychological Association. Adapted with permission.

2.3.1.2 The Direct Realist Theory.

Though perhaps the most cited theory in speech research, the *Motor Theory* of speech perception has received as many critical commentaries in the field of speech perception as positive recognition outside of the field. Working at the Haskins Laboratories, Fowler (1986, 1989, 1991, 1996) provided an alternative theory to the Motor Theory of speech perception, confirming some of its claims and rejecting others[23]. In order to account for the phenomenon of speech perception, the Direct Realist Theory (henceforward DRT) drives many of its principles from the concepts of affordance and information as proposed in the ecological approach of perceptual development (Gibson & Pick, 2000). Clear explanation of these concepts' implication in the DRT of speech perception is necessary to understand equally its tenets and designation. First, as for the affordance, being the objective properties of the environment and events, the DRT asserts that articulatory gestures are the actual objects of perception (i.e. distal stimulus) and not the neuro-motor commands or the intended motor gestures. This reciprocal relation describes the mutual adaptability between speech perception and production, stipulating that perception guide production, which, in turn, yields information for further guidance on perception, resulting in a continuous perception-action cycle (Gibson & Pick, 2000). Second, as for information being "the structured distribution of energy in an ambient array that specifies events or aspects of events in the environment" (Gibson & Pick, 2000, p. 18), the DRT affirms the fact that acoustic and auditory events serve as sources of speech information. That is, the acoustic or auditory event, created by a disturbing organism

---

[23] Fowler (1996) argued for embedding the theory of speech perception in the context of a larger theory of perception in which perceptual systems are considered to have a universal function. In a comprehensive review article of the Motor Theory of speech perception as originally proposed by Liberman and his colleagues, Galantucci, Fowler, and Turvey (2006) argued against the uniqueness of speech perception and maintained similar claims with respect to articulatory gestures as the object of perception and the recruitment of motor system in speech perception.

(the vocal tract in our case), does not resemble the organism; rather, the acoustic or auditory event specifies the organism and its path of locomotion in relation to the listener. This relation of specification is of paramount importance, for rich information capable of thoroughly specifying its source in the environment, does not necessitate any intermediate representation or hypothesising processes, allowing, therefore, of direct perception of the environment.

In her account of the foundations of the DRT of speech perception, Fowler (1989) stated that her proposal concerning speech perception derived from Gibson's direct, realist perspective, and argued,

> The real-world events of speech that structure a medium are, most locally, activities of the vocal tract…The informational medium causally structured by those activities is the acoustic signal. In the theory [the DRT], then, activities taking place in the vocal tract are objects of perception. Properties of acoustic signals are not perceptual objects, and "auditory" properties cannot be perceptual objects… (pp. 149-150)

Therefore, from the DRT perspective, patterns in informational media are caused by gestures of the vocal tract which perception recovers directly from patterns in physical energy, either acoustic, visual, or haptic (Fowler, 1986). Equally, as perception directly recovers gestures of the vocal tract, it is likely that properties of gestures of the vocal tract are therefore available to speech perceiver, who is capable of producing or imitating them. Grounding its arguments on properties of produced speech, the DRT offers an account of inherent issues in speech production such as coarticulation and mapping of articulatory gestures into phonetic segments. As coarticulation effects are largely assimilative (Fowler & Saltzman, 1993), the DRT introduces the concepts of "*coordinative structure* or (*synergy*)" and "*phonetic gesture*" (Fowler, 1986, p. 150) to explain the link between articulatory coordination and the presumable influence coarticulated, phonetic segments have on each other.

In view of Fowler (1989), coordinative structures are a temporary organisation of phonetic gestures for the realisation of a phonetic segment. Simply put, a coordinative structure is a stage of articulation that defines or realises a component of a phonetic component, either a consonant or vowel. For example, the coordinative structures necessary for the articulation of an egressive, pulmonic, stop consonant consist of three main stages, (a) the closing phase, (b) the closure phase, and (c) the release phase (Raphael, Borden, & Harris, 2011, p. 129). The closing phase is the transient cessation of the emitted airstream instigated by a complete obstruction of the vocal tract. The closure phase is the compression stage during which the occlusion obstructs airflow from coming out of the vocal tract, causing an increased, intraoral pressure. The last stage causes the held airflow behind the occlusion articulators to resume its flow out of the vocal tract. The three phases of articulation combined constitute phonetic segments that are generally the object of linguistic analysis in linear phonologies such as that of Chomsky and Halle (1968). On this ground, Fowler (1989) considered articulatory gestures probable "minimal linguistic components of an utterance" and "products of the minimal organizational structures of speech production" (p. 151).

The DRT demonstrates the role of coordination to explain articulatory constraints that implement phonetic gestures to create transient dependencies among articulators in the service of automatic and rapid distribution of activity of the entire articulatory system. The DRT refutes the context sensitive nature of coarticulation in speech in its linguistic essence, and claims that phonetic segments in natural speech do not affect each other in their linguistic identities. In this perspective, Fowler and Saltzman (1993) stated,

> coarticulatory effects in a given utterance result primarily from context-specific interactions among invariant gestural units during periods of coproduction, rather than from context-specific alterations in the intrinsic linguistic identities of these units…context-sensitivity in acoustic or articulatory flows arises from two sources: the time courses of the activation waves for the gestures in the utterance (the "speech plan"), and the manner in which the coordinative constraints of temporally overlapping (coproduced) gestures blend or interact with one another" (p. 190).

That is, coarticulatory effects in the production of consonants and vowels does not result in either their acoustic merging or articulatory, gestural assimilation, but in their coproduction. Therefore, coproduced, articulatory gestures are organised in the acoustic signal in parallel and independent canals just as two music tones, originating in two different music instruments, temporally overlap and accompany each other. As assumed by the DRT in relation the close relation between speech perception and production, the listener, having a prior knowledge of her articulatory system, arrives effortlessly at recovering the phonetic gestures along with their temporal sequencing and overlapping.

2.3.1.3 VOT and categorical perception of consonants.

The most frequently reported acoustic cue for consonant perception in speech research based originally on the MT and DRT is the notion of voice onset time, or conventionally VOT (Lisker & Abramson, 1964, 1970) and its role in the development of categorical perception (Studdert-Kennedy, 1976). According to Lisker and Abramson (1964), "the time interval between the burst[24] that marks release and the onset of periodicity that reflects laryngeal vibration" (p. 422) is a highly systematic, effective cue serving the separation of phonemic categories (cognate pairs) in several languages. A central finding in Lisker and Abramson's (1964) research is the empirical fact that languages choose to implement and realise their voice contrasts differently from each other via negative and positive VOT's. For example, the Puerto Rican Spanish uses long negative VOT's at -108 ms, -110 ms, and -138 ms to realise the stop consonants /g/, /d/, and /b/, respectively, and short positive VOT's to realise their voiceless counterparts /k/, /t/, and /p/ at +29 ms, +9 ms, and +4 ms, respectively as well. However, English uses relatively shorter negative VOT's at -88 ms, -102, and -101 ms to realise the stop consonants /g/, /d/, and /b/, respectively, and longer positive VOT's to realise their voiceless counterparts /k/, /t/, and /p/ at +80 ms, +70 ms, and +58 ms, respectively as well. Lisker and Abramson (1964) suggested further that phonetic aspects such as voicing, aspiration, and force

---

[24] A speech burst is the acoustic effect produced by the release of a closure in the vocal tract. It is generally associated with the noise accompanying the production of plosive consonants (Studdert-Kennedy, 1976, p. 173).

of articulation would be predictable consequences of the VOT serving as acoustic differentiators.

Speech researchers have extensively used VOT as a supporting evidence for the phenomenon of perceptual organisation of speech sounds called categorical perception. Probably the most studied phenomenon in perception given its importance concerning the remarkable stability and constancy in perception, categorical perception is a phenomenon understood as the ubiquitous use of a remarkable ability to categorise environmental stimuli (Repp, 1984; Ashby, Ennis, & Spiering, 2007). However, the notable abilities of environmental stimuli categorisation by an individual are a subject of disagreement among researchers as to whether perceptual categorisation abilities are instant, effortless, and attention-free (i.e. automatic) or whether they involve awareness and further cognitive representations (Shiffrin & Schneider, 1977; Schneider & Shiffrin, 1977; Repp, 1984). First findings on categorical perception were obtained from original experiments held by Liberman and his colleagues at the Haskins Laboratories. Undertaken experiments were a variety of specifically designed tests of discrimination and identification of synthetic speech sounds, the most common of which is the classical AB*X* discrimination test. For example, Lisker and Abramson (1970) carried out identification tests using computer-generated sound stimuli varying in temporal delay between the burst of air and the onset of voicing from -150 ms (voicing prior to air release) to +150 ms (voicing post to air release). Asking the participants to identify the heard stimuli as instances of /p/ or /b/, Lisker and Abramson (1970) found that participants almost unanimously identified the heard stimuli as /b/ from -150 ms up until +10 ms VOT (with 100% identification rate) and /p/ from +50 ms to +150 ms VOT (with more than 90% identification rate). The abrupt decline in the identification the phonemes /b/ and /p/ as a function of VOT is nowadays a conventional acoustic through which listeners are said to discriminate categorically voiced and voiceless stops in a predictable way.

2.3.2 Psychoacoustic theories.

Auditory and multimodal theories of speech perception are less articulate theories as against motor theories and constitute actually an amalgam of approaches from information theory, developmental psychology, and learning theories. Auditory or psychoacoustic approaches of speech perception emphasise the role of general, auditory mechanisms the listener uses to filter auditory information and map it onto phonetic representations, while multimodal approaches highlight the involvement of other modalities as an intermediate cognitive stage of speech representation, lying behind its perception and production. Hereinafter, we only provide key elements that make up the core of auditory and multimodal speech perception models. These key elements, we believe, are sufficient for the current section. However, we note that we will explain some specific, cross-language speech perception models that are auditory, multimodal-based and concomitant with our research interests.

2.3.2.1 Guiding framework.

The guiding framework for auditory theories of speech perception is a compilation of basic assumptions about: (a) the nature of speech sounds, (b) the improbable, modular nature of speech perception, and (c) the underlying auditory and phonetic processes governing speech perception (Lane, 1962; Massaro, 1975; Pisoni, 1973, 1975, 1977; Schouten, 1987; Repp, 1987; Studdert-Kennedy, 1976). The first assumption about the nature of speech sounds rests upon the *'Event View'* of the sound that assumes speech sounds to be psychoacoustic entities caused by the speaker and received by the listener. In this view, speech sounds are merely signals whose relation with the speaker's message is as arbitrary as letters of the alphabet relate to the writer's message. Relative to this arbitrary relation, Repp (1987) claimed, "except in very special circumstances, the sounds of speech as such do not play an important role in speech communication … it is the more abstract, articulatory information that is used by listeners to decode the linguistic message" (p. 7). According to Repp (1987), it is more plausible to view speech perception as a relational process that involves input-driven and knowledge- or memory-based processes related to understanding of language. Additionally, Repp (1987) assumed,

> it is not the stimulus as such (or its auditory transform) that is perceived, but rather its relationship to the phonetic knowledge base; perception thus is a relational process, a two-valued function. Its output is also two-valued: The relation of the input to the pre-existing internal structures yields (potential) awareness of the structure that provides the best fit which may be experienced as degree of confidence and uncertainty. (p. 13)

The second assumption about the nature of speech is that the latter, as a higher-order, cognitive function, has no presumable, modular advantage, and that speech sounds' perception and production are multi-modal, involving multiple sources of information and their complex integration (Massaro, 1987; Repp, 1987; Rosenblum, 2005). In an original, empirical research on the pairing of incongruent auditory and visual syllables in speech perception tests, McGurk and MacDonald's (1976) demonstrated the way repeated utterances of some syllables, dubbed on to lip movements for other syllables, were reported by listeners as none of the syllables or a combination of them. For example, listeners, whose age ranged from 3 to 40, reported the auditory component [ba-ba] dubbed on to articulatory gestures for [ga-ga] as fused [da], which constitutes a compromised perceived segment originating in neither components. When the auditory and visual components were reversed, listeners reported varying responses from fused perceived segment [da-da] to a combination of segments ([gabga], [bagba], [baga], [gaba]). Similar findings were obtained when listeners were presented with the auditory component [pa-pa] dubbed in on articulatory gestures for [ka-ka]. McGurk and MacDonald (1976) confirmed that, even when they included themselves in the study, they could not familiarise with the perceptual illusion of hearing fused or combined phonetic segments over hundreds of trials: a finding that researchers have reported as evidence for automatic integration of visual and auditory modalities in speech perception. This *McGurk effect* in audio-visual speech integration has ever since demonstrated the oscillation between visual and auditory sources of information in speech perception. Although several researchers have called for careful re-examination of

McGurk effect using more parametric, statistical analyses of individual participants in replication experiments, researches have provided quite similar findings (Mallick, Magnotti, & Beauchamp, 2015).

The third assumption about underlying speech processing mechanisms posits the existence of interdependent sound processing mechanisms (Cutting, Rosner, & Foard, 1976; Cutting, 1978; Macmillan, Braida, & Goldberg, 1987): (a) a general auditory mechanism responsible for processing of auditory signals impinging upon the auditory sense; and (b) a secondary, language-dependent, phonetic mechanism responsible for detecting acoustic patterns constantly present in speech. In a summary of empirical findings on cognitive processing mechanisms responsible for infants' speech perception, Kuhl (1979) concluded,

> the infant's perception of speech sounds indicate that the human infant makes discriminations that depend upon fine temporal and frequency changes in a complex auditory array. In addition, infants demonstrate perceptual constraints when listening to speech that can be characterized as adult like and appear to be predisposed to perceive certain speech-sound categories such as vowel and fricative categories… At this time, we cannot separate mechanisms in the human infant that are specifically linguistic from those that are generally psychoacoustic in nature. (p. 41)

Moreover, in a 2-experiment investigation on perceptual organisation of 18 synthetic speech syllables and 18 synthetic, musical stimuli in two experimental conditions (using loudspeakers and earphones), Cutting and Rosner (1974) found that their adult participants' identification and discrimination of synthetic speech syllables and musical stimuli were quite categorical in a parallel way equally constant in both experimental conditions. In spite of the varying points of peaks in discrimination and identification between synthetic musical stimuli and synthetic speech syllables, Cutting and Rosner (1974) stated, "this does not impair the overall similarity of the results" (p. 567) and that similarity of the results "is considerably greater than expected" (p. 567). On the ground of the obtained findings, Cutting and Rosner (1974) concluded,

> Taken together, these results suggest that certain aspects of speech perception are intimately related to processes and mechanisms exploited in other domains. The many categories in speech may be based on categories that occur elsewhere in auditory perception" (p. 564).

This capacity of perceptual organisation of both speech and non-speech sounds has been reported to be characteristic of adult listeners as well as infants (Jusczyk, Rosner, Cutting, Foard, & Smith, 1977; Cutting, Rosner, & Foard, 1976). In a similar study to that of Cutting and Rosner (1974), Jusczyk, Rosner, Cutting, Foard, and Smith (1977) used the same study design to explore 2-month old infants' perception of 4 synthetic non-speech sounds (varying in maximum amplitude occurrence after onset) using high-amplitude sucking technique[25]. Jusczyk et al. (1977) found that infants could perceive non-speech sounds in a categorical way, and suggested that the perceptual organisation of speech sounds in categories "is a consequence of general properties of the auditory system rather than of a special system devoted entirely to the perception of speech" (1977, p. 53).

2.3.2.2 Continuous perception of vowels.

The comparable lack of regularity in the acoustic features of consonants and vowels makes the principle of general auditory processing of speech sounds more plausible and affects the validity of the argument regarding categorical perception of all speech phonemes. Given the fact that vowels and fricative consonants have relatively steady-state formant transitions (F_1 and F_2) and to be less dependent on context, invoking a probable acoustic invariance and a

---

[25] High-amplitude sucking technique, or simply HAS technique, is a habituation technique used to assess an infant's perceptual abilities and sensitivity to stimuli differences. The rationale of HAS technique is that the sucking rate is easily conditioned by change of stimulus and level of arousal. For example, in an experiment using this procedure, an experimenter fosters infants' sucking response that has an amplitude exceeding a specified criterion with the presentation of the same sound stimulus until the amplitude drops to a decrement criterion, i.e. until habituation is established. Then, infants are put in a control and experimental group, the experimenter presents a new sound stimulus to experimental infants and the first stimulus to control infants to observe potential changes in sucking rate (for detailed description, see Kuhl, 1984).

different mode of perception that requires no special decoding, regardless of reported variability in vowel formant transition in normal speech. The reported variability is essentially due to the speaking rate (i.e. vowel reduction), phonetic context (i.e. surrounding sounds), gender, and age differences (Lindblom, 1963; Lindblom & Studdert-Kennedy, 1967; Liberman et al., 1967; Raphael, Borden, & Harris, 2011, p. 200). Pisoni (1973, 1975) and Pisoni and Lazarus (1974) argued for a dual cognitive process for consonants and vowels on the ground of the distinction between categorical perception of consonants and continuous perception[26] of steady-state vowels that motor theorists have repeatedly reported (Fry, Abramson, Eimas, and Liberman, 1962; Stevens, Liberman, Studdert-Kennedy, & Öhman, 1969). According to Pisoni (1973), the different abilities listeners have in discriminating synthetic, consonantal and vowel stimuli drawn from different phonetic categories are due to differential cognitive processing at the level of the short-term memory. Pisoni (1973) postulated,

> the differences between consonant and vowel discrimination are primarily due to the differential availability of auditory short-term memory for the acoustic cues which distinguish these two classes of speech sounds… As a result, we can conclude that auditory short-term memory for the acoustic properties of vowels is better than auditory short-term memory for the acoustic properties of consonants. (p. 259)

---

[26] Categorical and continuous perception of speech sounds are observed phenomena in speech discrimination and identification experiments. In categorical perception, listeners prove able to discriminate consonantal, synthetic stimuli with varying acoustic properties as long as they belong to different phonetic categories. This discrimination is claimed to be guided essentially by linguistic knowledge. The sharp and abrupt change in discriminatory power of synthetic consonantal stimuli that marks phonemic category boundary is not observed in synthetic vowel stimuli, where reported findings demonstrate better intra-phonemic or within-category distinctions and no clearly demarcated category boundary. Such a better discriminatory power for vowels is referred to as continuous perception.

2.3.3 Early stages of speech perception.

A common assumption in auditory and multimodal theories of speech perception is that infants are genetically predisposed to be sensitive to speech temporal and frequency patterns, suggesting that speech perception is a process of feature detection, realised through a process called *template matching*. Comparing infants to certain songbirds, Marler (1970) argued that template matching would be a general, genetic learning process of perceptual categorisation of acoustic patterns of speech, materialising in discriminatory and identification abilities of functional, input language sounds. That is, when the environment provides sufficient auditory stimuli, intrinsic, auditory mechanisms are activated to process abstract patterns of speech and store them in templates in service of vocal learning. In view of Marler (1970), speech perception is fundamentally a selective, vocal learning from adult speech in which auditory mechanisms allow individuals to hear and remember sounds from the surrounding environment, and to match the produced sounds to those sounds already stored in memory as an intrinsic reinforcement feedback. The use of the latter is primordial for adaptation and fine-tuning of auditory mechanisms in infants to perceive linguistically relevant patterns and become less sensitive for irrelevant, redundant features.

Several researchers have further defined cognitive and biological mechanisms responsible for speech perception and reported interesting findings about the way neonates, infants, and children develop specialised auditory systems for speech as against non-speech sounds in the first years of life (Werker & Pegg, 1992; Werker & Tees, 1992; Vouloumanos, Hauser, Werker, & Martin, 2010; Kuhl, Willimas, Lacerda, Stevens, & Lindblom, 1992; Maye, Weiss, & Aslin, 2008). In a review of their empirical work on infants’ speech perception and its development, Werker and Pegg (1992) highlighted the great capacity infants have to discriminate nearly all phonetic contrasts, including those that do not occur in their language-learning environment, and that this discriminatory capacity remarkably decreases within the first year of life. In view of Werker and Peggy (1992), this change in perceptual capacity is more of a perceptual reorganisation in shaping infants phonological learning and reorganise their perceptual sensitivities for optimal use as a function of native language phonemic categories. In the same vein, Kuhl, Willimas, Lacerda, Stevens, and Lindblom (1992) reported similar results and equally confirmed the effects of infants' early, linguistic experience on the formation of "phonetic prototypes" (p. 225) or categories that function as ideal exemplars of infants' native language (henceforward, L1) phonemes, therefore, easing the perception of speech. This phenomenon of perceptual tuning and observed decrease in abilities of perception and production of non-native phonemes is a highly controversial debate in cross-language speech perception.

2.3.4 Cross-language speech perception.

Several researchers have proposed several models to account for the perceptual organisation and production of speech pertaining to children's probable innate ability to learn effortlessly their L1 speech (Werker & Pegg, 1992; Werker & Tees, 1992; 1999; Kuhl, 1984; 1993, 1994; Marean, Werner, & Kuhl, 1992) and adults' probable altered ability to learn effortfully a second or a foreign language (Penfield, 1959; LaBrière, 1966; Lenneberg, 1967; Singleton, 1998; Amaro, Flynn, & Rothman, 2012). Linguistic research efforts in the field of second or foreign language learning (henceforth SLA[27]) have proliferated divergent hypotheses among researchers accounting for individual variation in learning speech of an L2. SLA researchers have highlighted the role of several factors affecting L2 phonological acquisition,

---

[27] We use the term SLA as a generic term to refer to all processes of learning a language other than the L1, regardless of divergent, theoretical considerations in linguistic research in differentiating between the status of language, whether it is second, foreign or interlanguage, and the psychological mechanism responsible for it, whether it is a built-in, unconscious process (acquisition) or an experience-nurtured, conscious one (learning). For a detailed account of the use of the terms SLA vs. FLA (foreign language learning), see Ellis (1999, pp. 11-15), Johnson and Johnson (1999); the use of the term interlanguage, see Selinker (1972); for the use of the terms acquisition vs. learning, see Krashen (1981, p. 8).

involving primarily maturational constraints (age) and cognitive factors, and secondarily educational and motivational factors (Singleton, 2003). In the following section, we limit our review to the role of age and cognitive factors as essential factors serving our major research purposes. We note that account of age effects is disproportionate to that of cognitive factor effects, for the latter are discussed in detail in Chapter 1 and in reviewed models of phonological acquisition (Sections 2.5.1, 2.5.2, and 2.5.3).

2.3.4.1 Age.

For several researchers, age is a significant factor for acquiring a language, be it an L1 or an L2, a concept termed the *Critical Period Hypothesis* (CPH). Penfield (1959) hypothesised the presence of a biological timetable for language acquisition that involves several brain mechanisms and specific, linguistic stages. According to Penfield (1959), the brain speech mechanisms consist of mapping of auditory impulses in neuronal connections in both brain hemispheres, involving cerebral areas specialised for both language and motor functions, with a left-hemispherical dominance even among right-handed people. Accordingly, auditory mapping leave behind neuronal basis for memory that corresponds to individual speech segments or sound units. Penfield and Roberts (1959) assumed the existence of an automatic relationship between the sound unit as an acoustic stimulus and the conceptual unit as an auditory image that is formed in the brain (p. 246). From learning individual sound conceptual units to word units, the brain image of the articulatory movement required to produce the learned words and the execution of the movement by employment of special motor areas in the cortex "becomes a skill that is eventually automatic but can be controlled voluntarily" (Penfield & Roberts, 1959, p. 248). This skill is firmly connected with age as the authors (1959) claimed,

> There seems to be little if any relationship between general intellectual capacity and the ability of a child to imitate an accent. Pronunciation is essentially an imitative process. Capacity for imitation is maximum between 4 and 8. It steadily decreases throughout later childhood. (p. 243)

This age limitation for successful imitation of speech forms an impediment to acquisition of further languages that would depend on the employment of L1 sound units inventory as a basis for L2 speech acquisition. The employment of L1 speech units' inventory to approach L2 speech units would be theoretically responsible for accented speech, an observed phenomenon in L2 learners' speech.

Lenneberg (1967) popularised the CPH view and stressed maturational constraints effects on language acquisition, resulting in lateralisation and termination of organisational plasticity, or simply the loss of neural plasticity that is a neural substrate for language acquisition. Arguing for specific onset (around 3 months of age) and offset for language acquisition (around puberty), Lenneberg (1967) noted,

> there are many skills and tasks that are much better learned during the late teens than in early childhood and a great deal of general learning has no age limitation whatever… our ability to learn foreign languages tends to confuse the picture. Most individuals of average intelligence are able to learn a second language after the beginning of their second decade, although the incidence of "language-learning-blocks" rapidly increases after puberty. Also automatic acquisition from mere exposure to a given language seems to disappear after this age, and foreign languages have to be taught and learned through a conscious and labored effort. Foreign accents cannot be overcome easily after puberty (p. 176).

In spite of the purely clinical aspect in Lenneberg's (1967) research work, Lenneberg's (1967) generalised his comments to include normal individuals, and assumed that accentedness in L2 learners would be systematic rather than pathological.

Ever since the publication of the work of Lenneberg in 1967, the notion of the CPH has stimulated an everlasting, inconclusive debate about the potential relationship between age of L1 acquisition and further linguistic experiences. Jonhson and Newport (1989) amended the initial version of the CPH and suggested two versions instead: (a) the "*exercise hypothesis*" and (b) the "*maturational state hypothesis*" (p. 64). The exercise hypothesis posits that infants and children have an inherent, superior capacity to acquire languages that, if exercised during the critical period, would maintain capacities intact for further linguistic experiences at any age, and vice versa. However, the maturational state hypothesis posits that L2 acquisition capacity decreases as a function of age irrespective of the onset of L1 experience. Operationalising age of arrival to the US as age of acquisition of English as an L2 in their study, Johnson and Newport (1989) found that performance of 46 native speakers of Korean on grammaticality judgement task was a linear function of their age of arrival up to puberty. That is, early Korean arrivals between the age of 3 and 7 and native speakers of English performed almost equally on all grammaticality judgement tasks[28] of English sentences, whereas late Korean arrivals between the age of 8 and 39 performed significantly poorly than early Korean arrivals, providing support for the maturational state hypothesis. The strong correlation between age of arrival and native-like performance indicates a gradual decline in linguistic performance starting at age 8. Johnson and Newport (1989) concluded, "human beings appear to have a special capacity for acquiring language in childhood, regardless of whether the language is their first or second" (p. 95). Several studies have sustained the maturational state hypothesis and constantly argued that differences in L2 acquisition, especially phonological learning, would be due to the nature of the human brain, not its nurture (Scovel, 1969). Scovel (1988) singled out phonological acquisition, specifically pronunciation, as the only linguistic aspect that was subject to the CPH. Hinging his arguments on the assumption that maturational effects are irreversible, Scovel (1988) argued that L2 pronunciation would be the only aspect of language to have a neuromuscular basis, a neuromotor involvement, and a physical reality that would make of it an impossible aspect to acquire past age 12 (p. 101).

Several researchers have amended early formulation of the CPH proposals accounting specifically for L2 phonological acquisition. Long (1990) argued for a softer version of the

---

[28] It is to note here that the notion of CPH in SLA research applies not only to foreign accentedness but extends to grammatical aspects of the L2, including syntax, morphology, tense use, etc. that are assumed a part of a general metalinguistic knowledge and awareness (Long, 1990; Bialystok, 1991).

CPH, claiming for rather relative, sensitive periods than an absolute critical period for acquiring different aspects of L1 and L2. As regards L1 acquisition, Long (1990), in a review of a series of L1 and L2 acquisition studies, admitted the relevance of maturational constraints to account for some significant differences in L1 acquisition among feral children, on the one hand, and deaf children learning ASL (American Sign Language), on the other. Long (1990) further stated that language acquisition would show irregularities if begun late around the age of 6 and 8, and that even late L1 children learners would exhibit an accelerated rate of development as against younger starters. As regards L2 acquisition, Long (1990) did not only refute the existence of more than a single critical period, but suggested the existence of several sensitive periods. Long (1990) stated,

> There are sensitive periods governing the ultimate level of first or second language attainment possible in different linguistic domains, not just phonology, with cumulative declines in learning capacity, not a catastrophic one-time loss, and beginning as early as age 6 in many individuals, not at puberty, as is often claimed. (p. 255)

These sensitive periods explain: (a) the different rates and paces of L2 development among young children, older children, and adults, and (b) the L2 proficiency level. In view of Long's (1990) proposals, older children or adults' would prove several advantages in rate and pace of L2 acquisition of some linguistic aspects as against young children. These advantages range from a shorter advantage in L2 phonological acquisition to a relatively longer advantage in acquisition of L2 syntax and morphology. However, these advantages would be short-lived, and young children would show significantly more L2 proficiency on the long run.

In spite of frequently reported evidence in favour of the CPH accounting for incomplete L2 acquisition, some researchers have claimed that there is no age limit for acquiring any language as L1 throughout the life span. Wode (1994) suggested, "human beings can learn more than one language as their L1s, that they can add additional ones throughout their entire lifespan, i.e. as L2s, L3s, etc., and that they can forget languages and relearn them" (p. 145). However, other researchers have only claimed some CPH claims do not provide the best explanations for "the earlier the better" (Flege, 1999, p. 101) phenomenon in L2 phonological acquisition. In a study of performance on grammaticality judgement task of French sentences by a group of 20 native speakers of English who began learning French as adults, Birdsong (1999) obtained unexpected findings. Birdsong (1999) found that: (a) 15 out of 20 participants scored almost as equal as native speakers of French did; and (b) age of arrival predicted the level of performance, though the participants had moved to France as adults. Equally, Bongaerts (1999) reported interesting findings of ultimate attainment in L2 phonological acquisition involving a group of highly successful, late L2 Dutch learners of French and English, who learned the L2 within an instructional context. For Bongaerts (1999), judging his participants' speech as native-like or authentic by native speakers of the language provided quite an empirical counterevidence to the assumptions of the CPH, and demonstrated that "claims concerning an absolute biological barrier to the attainment of a native-like accent in a foreign language are too strong" (p. 154).

2.3.4.2 Transfer of knowledge representation.

Several researchers have attempted to account for the differences in L2 learning away from neurophysiological and maturational constraints to place them within a general, cognitive framework of skill acquisition (Anderson, 1982). These researchers have emphasised both the non-specific nature of language as a skill and the role of general cognitive mechanisms responsible for the acquisition of any language (L1 or L2), some of which are equally shared in acquisition of other cognitive and psychomotor skills. The rationale for approaching language acquisition as a skill is the assumption that

> the learning of a wide variety skills shows a remarkable similarity in development from initial representation of knowledge through initial changes in behaviour to eventual fluent, spontaneous, largely effortless, and highly skilled behaviour, and that this set of phenomena can be accounted for by a set of basic principles common to acquisition of all skills. (DeKeyser, 2007, p. 97)

It is to note that the concept of memory is primary to all skills acquisition and their mechanisms, involving: (a) explicit and implicit rule learning, (b) priming, (c) automaticity, and (d) practice that increases performance and reduces error rates, following the power law of learning. First, it is commonplace to think of L1 acquisition as an innate process activated by interaction with a linguistic environment. A child acquires her L1 speech effortlessly and without any formal instruction, involving implicit acquisition of rules of grammar, syntax, morphology, and pronunciation. However, except in cases of bilingual situations where the learner acquires both L1 and L2 simultaneously outside of educational context, regardless of whether she proves a "balanced bilingual" (Crystal, 2008, p. 53) or not, an L2 is generally acquired in an educational context through explicit acquisition of rules of various language aspects. Second, priming stands for the effects prior knowledge representation may have on further experiences. In SLA research, priming has been equated with the phenomenon of transfer in which already acquired, linguistic patterns are transferred to analogous situations, as Lado (1957) suggested, "individuals tend to transfer the forms and meanings, and the distribution of forms and meanings of their native language and culture to the foreign language and culture" (p. 2). There is probably no better example than L2 phonological acquisition in which transfer and interference have been extensively used as an explanation for production and perception of L2 sounds, respectively. Contrastive linguists have widely utilised the concept of language transfer in assuming the fact that L2 learners are very prone to using, consciously or not, L1 features in their initial acquisition steps of L2 perception and production. Third, automaticity denotes the transition from attentive processing of linguistic information to a non-attentive mode of processing that materialises in effortless and fast use of already existing linguistic knowledge in an individual's memory. A good example to illustrate formation of automaticity in skill acquisition is the assumption that speech perception is accomplished by development of automatic perceptual mechanisms at the phonetic level that manifest themselves in categorical perception of sounds (Johnson & Ralston, 1994). The fourth cognitive mechanism underlying L2 acquisition is the importance of frequency of correct practice that leads to equally improved L2 acquisition at both perception and production.

2.4 Theories of Second Language Speech Perception

SLA research concerned with phonological acquisition and pedagogical methods for teaching pronunciation has long addressed the relationship between L2 sounds' perception and production. Researchers have claimed for both a linear relationship between L2 speech perception and production and the precedence of L2 speech perception over production (Trubetzkoy, 1969; Kuhl, 1991; Flege, 1995; Best, 1991; 1995). However, some researchers have suggested an asymmetrical relation between L2 sounds' perception and production and the probable precedence of production over perception (Brière, 1966; Sheldon & Strange, 1982; Neufeld, 1984; Borrell, 1990). It is not our intention in the following section to review empirical evidence supporting either claim, but to provide three widely cited models in speech perception research that focus on segmentals.

2.4.1 The Native Language Magnet Theory (NLM).

Based on the assumption that humans come into the world with important built-in auditory abilities, the Native Language Magnet (henceforth NLM) accounts for infants' speech perception and the way language experience with the native language modifies early-life discrimination abilities as a function of learning (Kuhl, 1991; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Kuhl, 1993). The NLM suggests that a developmental process of internal organisation of phonetic stimuli guides speech perception at an early age in the infants' life, regardless of the language to which infants are exposed. In the NLM terms, infants first demonstrate an important ability to: (a) perceive speech in an auditory mode chiefly directed by general auditory mechanisms, and (b) "recognize the coherence between speech presented visually and auditorially" (Kuhl, 1993, p. 248). Kuhl et al. (1992) stated,

> At the beginning of life, human infants exhibit a similar pattern of phonetic perception regardless of the language environment in which they are born. They discern differences between the phonetic units of many different languages they have never heard, indicating that the perception of human speech is strongly influenced by innate factors. (p. 606)

The NLM theory suggests that by 6 months of age, language-specific mechanisms substitute for early built-in general auditory mechanisms as infants experience further exposure to their L1, months before infants begin to acquire word meaning. Building on confirmed findings from visual perception research, the NLM proposes that speech stimuli are graded quantitatively and qualitatively just as other complex stimuli from other domains (Kuhl, 1991), i.e. speech stimuli are rated in terms of effectiveness and representativeness. The theory claims that infants' innate ability to discriminate phonetic contrasts of all languages start to diminish by 6 months of age as infants' speech perceptual system adapts and is attuned to their L1 at both phonemic and phonotactics levels. Using statistical properties of language input as part of an internal organisational system to discover structure (Saffran, 2003), the speech perceptual system of 6-month old infants is capable of grouping L1 speech stimuli into categories and further preferring some stimuli to others. Stated differently, infants' initial perceptual abilities provide a division of speech stimuli into categories that language input reinforces with experience. According to the NLM theory, speech statistical patterns and constraints of the L1 phonological system cause "the formation of stored representations of native-language phonetic categories" (Kuhl, 1993, p. 249), i.e. the formation of L1 phoneme categories. Therefore, L1 phoneme category representations would form exemplars or prototypes that adult speakers of the native language identify as ideal representatives of specific phonemic categories. These prototypes act as "perceptual magnets" and "perceptually attract nearby members of the category, and this provides internal structure to the category" (Kuhl, 1993, p. 249).

As it builds upon both infants' innate abilities and sensitivity to language experience, the NLM theory suggests that infants: (a) initially partition acoustic stimuli into categories as part of general auditory perceptual processing mechanisms, whereby they exhibit an enhanced discriminability near the boundaries that make the phonetic categories, and (b) subsequently exhibit phonetic prototype effect that reflects the learned categories (Kuhl, 1993). For instance, phonetic research has demonstrated that vowels are largely characterised by their relative first and second formants (F_1 and F_2) and that various languages have different sets of vowel sounds that carve up the F_1/F_2 space differently from each other. As first and second formant frequencies characterise prototypic vowels in a given language, the magnet effect may cause non-prototypic vowels (nearby vowel sounds with different first and second formant frequencies to the prototypic one) to be perceived "as more similar to the category prototype than to each other" (Kuhl et al., 1992, p. 607). The magnet effect causes: (a) the non-prototypic vowels to be perceptually drawn towards the prototype, and (b) the equal physical (here the acoustic difference) differences between non-prototypic vowels to shrink.

Using the head-turn conditioning technique[29], Grieser and Kuhl (1989) tested sixteen 6-month old infants' categorisation of 64 novel computer-synthesised speech stimuli as variants of two vowel centres identified by adults as prototypes of two vowel categories (/i/ and /ɛ/), with 32 stimuli from each category. The variants were different from the two prototypic vowel centres in either F_1, F_2, or both of them. Obtained results showed that infants categorised accurately the synthetic stimuli over 90% of the time, confirming the hypothesis that young infants can categorise vowel stimuli into two categories when the synthetic stimuli vary in category goodness. Grieser and Kuhl (1989) stated that their results

---

[29] The head-turn conditioning technique or the HT paradigm is a frequent technique in infants' speech perception research, whereby infants are trained to turn their heads away from the experimenter towards a loud speaker when there is a change in the speech sound category (as identified by adults) within a time interval of less than 5 seconds. A correct head-turn is rewarded with a visual reinforcer, while an incorrect head-turn is not reinforced. In control trials, the experimenter plays the same speech stimulus and monitors the head-turn responses to assess the head-turning chance probability.

> support the conclusion that infants categorise novel /i/ and /ɛ/ stimuli after training on a single good exemplar of each category, even then the novel exemplars are no longer all "good" or prototypical exemplars of their respective categories. (p. 582)

Obtained results of a similar research with 32 American and 32 Swedish 6-month-old infants showed a magnet effect around prototypic vowels in infants' L1, /i/ for American infants and /y/ for Swedish infants (English does not have /y/ and Swedish does not have /i/). Using the HT paradigm to assess infants' categorical perception of computer-synthesised vowel stimuli varying in spectral distance from prototypic and non-prototypic vowel centres, infants showed magnet effect for their native-language prototypic vowel only. These results were consistent with previous infants' speech perception researches that would contribute to the understanding of the fast and efficient processing of speech information at the first year of life (Werker & Tees, 1984, 1999).

By drawing similarities between the NLM research findings and Liberman's findings in adults' categorical perception tests at the Haskins Laboratories, Kuhl (1993) provided further support to the NLM general claim that the evolution of infants' innate discrimination abilities are experientially used in the selection of L1 contrastive sounds. Therefore, enhanced discrimination abilities for infants, become structured through ambient language experience and maintain their significant sufficiency for the acquisition of the L1. In a revised version of the NLM theory, termed the *NLM-expanded* (NLM-e), Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola, and Nelson (2008) reiterated the three phases in the development of infants' speech perceptual system: (a) the initial phase at which infants exhibit universal-language discrimination abilities that derive from general auditory mechanisms that are not species-specific; (b) the second phase at which infants' sensitivity to L1 produces phonetic representations (prototypes) that start acting as perceptual magnets; and (c) the third phase at which the perceptual magnet effect causes perception distortion and produces facilitation in native and a reduction in non-native language phonetic abilities. Furthermore, Kuhl et al. (2008)

argued that the dual change in infants’ speech perception skills towards the end of the first year of life, predicts infants’ later language abilities. According to Kuhl et al. (2008), better L1 phonetic perception at 7.5 months of age predicts faster L1 advancement, whereas better non-native phonetic perception predicts slower L1 advancement. While faster L1 advancement indicates neural commitment, acute non-native language discrimination abilities indicate uncommitted neural circuitry (Kuhl et al., 2008, p. 979).

Though primarily aimed at accounting for the evolution of infants’ speech perceptual abilities of their L1, Kuhl (1993) stated that the NLM theory might help explain the reported findings in research on adult’s acquisition of a non-native language. The NLM theory appeals to the *proximity principle* to demonstrate that not all non-native language sound contrasts are hard to discriminate: “the nearer a new sound is to a native-language magnet, the more it will be assimilated by it, making the new sound indistinguishable from the native-language sound” (Kuhl, 1993, p. 259). However, the application of the NLM theory in the field of SLA research should include more than auditory mechanisms and acoustic characteristics of speech sounds alone in speech perception, as SLA experience occurs post to the time language learners have learned to utter words and understand them.

2.4.2 The Speech Learning Model (SLM).

The Speech Learning Model (henceforth SLM) has been developed to address L2 phonological acquisition and account for individual variation in learning a second language, mainly adult-child differences (Flege, 1981, 1987a, 1987b, 1991, 1992, 1995, 1997, 1999, 2003a, 2003b, 2008). The SLM builds essentially on several assumptions regarding (a) L2 speech learning mechanisms, (b) age constraints, (c) the role of L2 input, and (d) the symmetrical relation between L2 perception and production. The learning mechanisms of L2 speech and maturational constraints make the core elements of the SLM, as the latter argues against the claims of the CPH (neurological maturation and reduced neural plasticity) and its inconclusive, though observed, findings. Flege (1995) argued that L1 acquisition mechanisms would not cease to function due to maturational constraints and would always be active for the learning of an L2, and that adult learners' accented L2 speech and perceptual difficulties would result from the reorganisation of their phonetic system. The latter, according to Flege (1995), may undergo changes as a function of L2 experience and, therefore, improve significantly adult learners' pronunciation of L2. The SLM holds both symmetrical relationship between L2 speech perception and production and sufficient L2 input are a prerequisite for unaccented speech. Flege (1995) put it,

> The aim of our research is to understand how speech learning changes over the life span and explain why "earlier is better" as far as learning to pronounce a second language (L2) is concerned. An assumption we make is that the phonetic systems used in the production and perception of vowels and consonants *remain adaptive over the life span*, and that phonetic systems reorganize in response to sounds encountered in an L2 through the addition of phonetic categories, or through the modification of old ones. (p. 233)

The refined version of the SLM builds on two main developed principles: (a) the phonological translation hypothesis that concerns L2 speech production (Flege, 1981), and (b) the classification equivalence principle that concerns L2 speech perception (Flege, 1987a, 1987b). In arguing for the role of L1 in L2 acquisition rather than age and neurological maturation, Flege (1981) advanced the "phonological translation hypothesis" (p. 448) to demonstrate the general tendency of adult speakers to produce L2 sounds with spectral and temporal characteristics typical of their L1 sounds. Flege's (1981) phonological translation hypothesis corroborates previous research findings of Caramazza, Yeni-Komshian, Zurif, & Carbone (1973) and Munro (1993). Caramazza et al. (1973) demonstrated that, regardless of similar articulatory features in English and French voiceless plosives /p, t, k/, the latter are phonetically realised with (a) shorter mean VOT values (18, 23, and 32 ms, respectively) in French words by monolingual Canadian speakers of French (MCF), and (b) longer mean VOT values (62, 70, and 90 ms, respectively) in English words by monolingual Canadian speakers of English (MCE). Caramazza et al. (1973) further demonstrated that (a) bilingual Canadians (BCEF) realised /p, t, k/ with longer mean VOT values (20, 28, and 35 ms, respectively) in French words as against MF participants, and (b) shorter mean VOT values (39, 48, and 67 ms, respectively) in English words as against MCE participants. Caramazza et al.'s (1973) analysis of variance in the distribution of mean VOT values in the phonetic realisation of /p, t, k/ in both French and English words between groups, revealed (a) no significant differences in the phonetic realisation of /p, t, k/ in French words between MCF and BCEF participants, and (b) significant differences in the phonetic realisation of /p, t, k/ in English words by MCE and BCEF participants. These findings led Caramazza et al. (1973) to suggest that BCEF participants did not have a complete bilingual phonological system free of interlanguage interference, which revealed some unidirectional interference from French (probably stronger language) to English (probably weaker language).

Munro's (1993) research findings on the articulation of English vowels by native speakers of Arabic demonstrated a similar pervasiveness of L1 vowels' acoustic properties in the realisation of English tense-lax vowel pairs. According to Munro (1993), the temporal and spectral properties of English vowels as articulated by native speakers of Arabic were significantly different from those of English vowels as articulated by native speakers of English. Some of the observed differences are (a) exaggerated lengthening of tense vowels /iː, uː/ and (b) tendency to articulate back vowels with a low F2, suggesting that,

> Because Arabic long-short pairs exhibit a greater duration difference than English tense-lax pairs, they [*participant native speakers of Arabic*] may have produced these vowels with the exaggerated duration effects. (Munro, 1993, p. 51)

Regarding speech perception, Flege (1987a, 1987b, 1991) introduced the '*equivalence classification principle*' to account for reported perceptual difficulties encountered by adult language learners in the formation of new L2 sound categories. This fact is considered the cause of L2 articulatory distortions, given that speech perception precedes its production and that speech perceptual representations formed in the long-term memory guide L2 sound production. That is, clear perception of both L1 and L2 sound contrasts as well as firm formation of their respective phonetic categories are a sine qua non of unaccented L2 pronunciation[30]. Unlike the contrastive analysis hypothesis[31] that suggests the facilitating role of similarity between L1 and L2 sound inventories in learning L2 pronunciation, the SLM hypothesises that the greater the perceived similarity of an L2 sound from the closest sound of the L1, the more likely a new category will not be formed for the L2 sound. As regards age and the notable effect of early exposure to L2, the SLM hypothesises that, as L1 sound inventory is firmly established and acoustic templates are stabilised through childhood, neighbouring L2 sound categories may not form. Flege (1987c) argued that children might pronounce L2 sounds better than adults "because they are more likely to develop new phonetic categories as a result of exposure to sounds in L2 which are acoustically non-identical to sounds in L1" (p. 172). Flege (1987c) specified further that probable incomplete formation of an L2 sound category, which is similar to an L1 equivalent, would make both L1 and L2 sound categories assimilate, and lead to a merged L1-L2 sound category or a composite.

To investigate the issue of L1 and L2 sounds' perceived similarity and dissimilarity, Flege (1991, 1992) discussed several methods to categorise L2 sounds into new or similar ones

---

[30] It is to note that equivalence classification principle is similar to Trubetzkoy's (1969) notion of the sieve or filter, suggesting that pronunciation difficulties arise from perceptual problems and false evaluation of L2 phonemes. Trubetzkoy (1969) stated, "the phonological system of a language is like a sieve through which everything that is said passes. Only those phonic marks that are relevant for the identity of the phoneme remain in it. The rest falls down into another sieve in which the phonic marks, relevant for the function of appeal, are retained… Starting from childhood, each person becomes accustomed to analysing what is said in this fashion. This analysis is carried out quite automatically and unconsciously" (pp. 51-52).

[31] The contrastive analysis hypothesis considers transfer of L1 forms and meanings at both perception and production levels as essential in learning an L2. As Lado (1957) put it, "we assume that the student who comes in contact with a foreign language will find some features of it quite easy and others extremely difficult. Those elements that are similar to his native language will be simple for him, and those elements that are different will be difficult" (p. 2).

based on (a) phonetic notation of L1 and L2 sounds, (b) spectral features of vowels as plotted on the F_1/F_2 perceptual map (the 2-dimensional high-low vs. front-back vowel space), and (c) phonetic distinctness tests, including discrimination and classification tests. Thus, it is necessary that (a) different IPA symbols be used to represent auditorially distinguished L1 and L2 sounds; (b) spectral distance between an L2 vowel and its closest L1 vowel be determined to decide whether the L2 vowel can be considered new; and (c) category goodness judgement (rating) tests be used to assess L2 perceived similarity or dissimilarity with the closest L1 sound. As regards L2 vowel categorisation along the F_1/F_2 perceptual space, the SLM assumes that it is likely that an L2 vowel, occupying an unexploited spectral space by the L1, would make of the former a new, easy-to-learn vowel, and, when the L2 vowel is created, perceptual dissimilarity is maximised to avoid perceptual confusion in the L1-L2 vowel space. Yet, Flege (2005, April) admitted that, the treatment of an L2 vowel occupying the unexploited spectral space in the L1 would take time to emerge as a result of both developmental and learning processes. Flege (2005, April) provided predictions about the final learning state of similar and dissimilar L2 vowels, suggesting L2 dissimilar vowels to be produced poorly at an early stage then would be produced more accurately with predominant L2 use. The SLM suggests the use of perceptual experiments to assess L1 and L2 perceived dissimilarity involving, for instance, digitally-synthesised sound tokens along continua delimited by L1 and L2 prototypic sounds (for a detailed account of speech perception tests, see McGuire, 2010).

Findings from Caramazza et al. (1973) corroborate some of the SLM claims, namely the composite L1-L2 sound. The latter revealed that bilingual Canadians’ control for production when switching between English and French did not match bilingual Canadians’ phonological discrimination perceptual ability. Bilingual Canadians could produce the voiceless stops /p, t, k/ with mean VOT values in intermediate positions between those typical of French and English. However, they appeared to have used the same criteria in their perceptual decision to discriminate synthetic 350-ms syllable continua (voiceless stop followed by a vowel), with VOT values ranging from -150 ms to +150 ms following the release burst. This may indicate that, regardless of their relative success in phonetic realisation of /p, t, k/ in a typical manner in both English and French, bilingual Canadians failed at maintaining separate representations for them at the perceptual level.

2.4.3 The Perceptual Assimilation Model (PAM).

Building on the concept of perceptual learning and development that emphasise the dynamic nature of perception and action mechanisms (Gibson & Pick, 2000), Best (1991, 1995) developed the Perceptual Assimilation Model[32] to account for (a) the nature of information in speech perception; (b) the relation between speech perception and production; and (c) the impact speech organisation in L1 may have on further language learning experience. The PAM address the cognitive mechanisms underlying L1 speech perception and language-specific influences on speech perception at the segmental level. Specifically, the PAM addresses several issues related to the development of the L2 learner's speech perception mechanisms in discerning L2 sound contrasts and the varying degrees of difficulty in doing so by L2 learners. Drawing essentially from the direct realist view of speech perception (Chapter 2, Section 2.4.1.2), the PAM specifies the primitives for speech perception and their cognitive structure as either abstract stationary representations or dynamic concrete patterns.

Bearing much in common with motor theories of speech perception, the central premise of the PAM is that speech segments are the property of articulatory organs, and that "the perceptual primitives [*of speech perception*] are articulatory gestures" (Best, 1995, p. 176). The PAM assumes that the mechanisms underlying L1 perceptual primitives are life-long processes

---

[32] It is to note in here that perceptual process of assimilation in the PAM is completely different from the notion of assimilation in speech production. Perceptual assimilation describes a process of phonetic analogy, while assimilation in speech production refers to the influence one sound has upon the articulation of another.

restrained neither by age nor by maturational factors, as Best (1995) put it, “perceptual learning continues into adulthood” (p. 198). As regards speech perception in infancy, the PAM makes similar claims to the NLM in that discovery of L1 critically distinctive features guide infants’ speech perceptual mechanisms. According to Best (1995), perceptual learning is orientated towards the formation of higher-order invariants, as vocabulary items and syntactic structures, out of lower-order ones, i.e. speech segments. The organisation and detection of higher-order invariants by infants will not be possible, unless infants learn to pick the lower-order critically distinctive features. Best (1995) put it,

> Educated speech perception should, therefore, actively seek and extract the invariants of language-specific articulatory gestures and constellations [*clusters or patterns*] of intergestural phasing at all levels from segments to syllables, words, and so forth. Language specific gestural constellations are complex articulatory events, which are specified by higher-order invariants in the signal that automatically account for contextual variations such as speaking rate, speaker differences, and allophonic variation, so that perceivers “hear through” irrelevant lower-order variations. The converse of the efficiency of extracting higher-order invariants in native speech may be an increased difficulty in picking up lower-order gestural invariants of unfamiliar non-native categories that are irrelevant to critical native distinctions. (p. 185)

Therefore, perceptual attunement to L1 speech would have considerable implications for perception of L2 phonetic patterns.

As the PAM considers articulatory gestures atoms of the phonological structure, it indirectly assumes that, articulatory gestures infants can produce form a universal phonetic domain, of which the L1 phonological space purposefully selects those gestures that serve phonologically contrastive functions. A corollary of this attunement to the L1 (or the native phonological space) is that infants may discard those lower-order gestural invariants that have no phonological function in the L1 and which they could notice at an early stage in life. Much like the NLM, the PAM assumes that both attunement to an L1and further experience with it are very likely to determine the perception of L2 gestural constellations, given the notable, shared gestural constellations among language. As L2 segments “are those whose gestural elements or intergestural phasing do not match any native constellations" (Best, 1995, p. 193), the PAM suggests, the degree of similarity to or discrepancy from L1 sounds will guide the perception of L2 sounds, following the latter's proximity to L1 phonological space.

Regarding L2 speech perception, the PAM claims that the universal phonetic domain and the dynamic articulatory gestures provide some dimensions for assessing the relative similarity or discrepancy between L2 and L1 sounds, given that the "phonological patterning in languages obeys the constraints provided by the physical structure of the vocal tract and the movement that its biomechanical components afford" (Best, 1995, p. 187). In PAM terms, L1 sound inventory guides early L2 speech learning by virtue of L2 learners' detection of gestural constellations, as provided by spatial proximity in constriction degree and gestural schedule. That is, by virtue of experience with L1, the L2 learner is very likely to detect the active organs involved in the articulation of an L2 sound and its articulation stages. Consider the case of the voiceless velar fricative /x/, as it exists in the Arabic language in words like /xilaaf/ and /xabarun/ (meaning dispute and news, respectively). In PAM terms, a learner of Arabic as an L2 (whose L1's phonological space does not include this sound) is expected to detect (a) the active articulators employed in the articulation of these L2 and L1 sounds, and (b) the similarity to, or discrepancy from the gestural constellations employed in L1, in case the discrepancy is important. The first step is likely to allow the learner recognise the involved articulators (say, the velum or the uvular) and the gestural phasing (say, plosive or fricative). Detection of similarity or discrepancy in the gestural constellations between an L1 sound and the L2 sound (/x/ in our example) will cause the learner to assimilate perceptually the L2 sound.

Considering three possible patterns of perceptual assimilation, the PAM assumes that an L2 sound segment is perceptually assimilated to (a) a native category if the learner detects important similarity in gestural constellations; (b) non-native category if the learners detects considerable discrepancy; and (c) non-speech category. Perceptual assimilation of L2 to a native category occurs when the learner hears the L2 sound as (a) a good exemplar of the L1 category, (b) an acceptable though not ideal exemplar of the L1 category, and (c) a notably deviant exemplar of the L1 category. Perceptual assimilation to a non-native category occurs when the learner recognise the L2 sound as a speech-like segment that does not resemble any exemplar of L1 sounds. The perceptual assimilation of an L2 sound to a non-speech category happens when the learner fails at establishing neither similarity to, nor difference from any existing L1 sound (Best, 1995).

In a similar way to predict patterns of perceptual assimilation of L2 sounds to L1 sounds, the PAM further predicts the degree of differentiation or discriminability among L2 sound contrasts. For this, the PAM offers (a) six pairwise perceptual assimilation patterns, and (b) the predicted discriminability level for each pairwise comparison, as put below.

> (a) *Two-Category Assimilation (TC Type)* Each non-native segment is assimilated to a different native category, and discrimination is expected to be excellent.
> (b) *Category-Goodness Difference (CG Type)* Both non-native sounds are assimilated to the same native category, but they differ in discrepancy from native "ideal" (e.g., one is acceptable, the other is deviant). Discrimination is expected to be moderate to very good, depending on the magnitude of difference in category goodness for each of the non-native sounds.
> (c) *Single-Category Assimilation (SC Type)* Both non-native sounds are assimilated to the same native category, but are equally discrepant from the native "idea"; that is, both are equally acceptable or both are equally deviant. Discrimination is expected to be poor (although it may be somewhat above chance level).
> (d) *Both Uncategorizable (UU Type)* Both non-native sounds fall within phonetic space but outside of any particular native category, and can vary in their discriminability as uncategorizable speech sounds. Discrimination is expected to range from poor to very good, depending upon their proximity to each other and to native categories within native phonological space.
> (e) *Uncategorized versus Categorised (UC Type)* One non-native sound assimilated to a native category, the other falls in phonetic space, outside native categories. Discrimination is expected to be very good.
> (f) *Nonassimilable (NA Type)* both non-native categories fall outside of speech domain being heard as nonspeech sounds, and the pair vary in the discriminability as nonspeech sounds; discrimination is expected to be good to very good.
>
> (Best, 1995, p. 195)

Predictions of the PAM have been measured and assessed in several researches investigating L2 sound contrasts firstly in infants, and then they have been generalised to L2 sound contrasts in adults, employing categorisation and identification tests. In her review of developmental cross-language research, Best (1995) affirmed infants' (a) remarkable decline in ability to perceive L2 sound contrasts between 6 and 12 months of age, and (b) developmental reorganisation in their speech perception even before emergence of their L1 first words. The latter is probably the cause for the unavoidable effect of L1 on acquiring or learning further languages in life.

Best and strange (1995) investigated adult native speakers of American English and Japanese on their discrimination and identification of approximant consonants /w, j, r, l/ in initial position in synthetic syllables, given their various phonological status and phonetic realisations in American English and Japanese[33]. The synthetic stimuli were series of /rak, lak/, /wak, jak/, and /wak, rak/ continua, with every pair containing 10 equal-step generated stimuli in which simultaneous changes in F_2 and F_3 were made to move from the first consonant to the other in the contrast. Using the A*X*B procedure, the authors ran 10 repetitions of the two A*X*B orders for the seven stimuli pairings that differed by a 3-equal step along the continuum (i.e. 1-4, 2-5, 3-6, 4-7, 5-8, 6-9, 7-10). For consonant identification in the stimuli, Best and strange (1995) used a 2-alternative forced choice identification test (2-AFC), with each test including 20 repetitions of each of the stimuli in the tested series. American and Japanese participants

---

[33] The four approximants represent abstract phonological oppositions in American English, i.e. phonemes. However, Japanese language employs /w, j, r/ as abstract phonological oppositions, but does not employ a distinct /l/ phoneme. Phonetic realisation of /w, j, r/ differ in similarities between American English and Japanese. The phoneme /j/ is realised phonetically in a very similar way in both languages, being a glide palatal consonant articulated with neutral-to-spread lip posture. The phoneme /w/ is an approximant realised with lip rounding or protrusion (as in /uː/) in American English and with spread lips in Japanese (as in the back unrounded Japanese vowel /ɯ/). The phoneme /r/ is phonetically realised as a retroflex (post-alveolar) or palate-alveolar approximant ([ɹ]) in American English and typically a retroflex or alveolar flap in Japanese ([ɾ]) and occasionally as a lateral alveolar tap [ɺ].

proved similar discrimination and identification results, demonstrating sharp crossovers at category boundaries in discriminating and labelling within-category stimuli for the /w, j/ and /w, r/ series. However, Japanese participants proved remarkably less consistent in labelling within-category for the /r, l/ series. Best and Strange's (1995) findings confirmed their initial hypotheses and aligned with the PAM's predictions about the perceptual assimilation patterns: (a) /w, j/ contrasts assimilated to *TC* type; (b) /w, r/ contrasts assimilated to *CG* type; and (c) /r, l/ contrasts assimilated to *SC* type. A further statistical analysis showed that experienced Japanese participants had significantly intermediate discrimination and identification results as against American and inexperienced Japanese participants.

Conclusion

Regardless of the nature of arguments and explanations each of the three theoretical frameworks tries to employ, it is to emphasise that they are mutually compatible. While the NLM theory lends itself to early life stages of speech perception development, the PAM and the SLM equally employ L1 and L2 similarities and dissimilarities to account for adult L2 speech perception. As against the PAM that accounts for initial stage of L2 vowel perceptual development, the SLM accounts for ultimate development of L2 vowel perception. Indeed, neither the PAM nor the SLM does account for the course of development of L2 vowel sounds as a function of time. Therefore, the use of the PAM or SLM as a theoretical framework in our research work does not have any theoretical or empirical costs.

Chapter 3

Attitude: Literature Review

Introduction

Research in SLA has demonstrated the importance of several factors accounting for individual variation in attainment of an L2, involving maturational factors (Lenneberg, 1967; Kuhl, 1991; Best, 1995; Leather, 2003), language aptitude and input (Krashen, 1981, 1989; Flege, 1991), cognitive factors (O'Malley & Chamot, 1990; Oxford, 1990), and motivational and affective issues (Gardner & Lambert, 1972; Gardner, 1985; Belmechri & Hummel, 1998; Noels, 2001; Masgoret & Gardner, 2003). Among investigated factors, attitudes have been of particular interest for SLA researchers and language teachers because of their attested predictive power in determining the level of motivation, the amount of efforts required for effective and proficient L2 learning (Masgoret & Gardner, 2003), and classroom interaction. Interest in SLA research has resulted in identifying several types of individual, motivational and affective factors and their interplay in relation to language skills, mainly ultimate attainment in L2 pronunciation. The complex nature of these aforementioned determinants of L2 pronunciation has instigated several researchers to investigate the issue as a function of L2 learning goals and teaching methods. Indeed, attitude research in ultimate attainment of L2 pronunciation has demonstrated a complex dynamics of self-perceptions, beliefs, and affect in generating L2 pronunciation learning behaviours and their relationship to L2 pronunciation achievement and proficiency (Baran-Łucarz, 2017; Pawlak, Mystkowska-Wiertelak, & Bielak, 2015; Szyszka, 2017).

Given the significant relevance of the concept of attitude in L2 pronunciation learning, its multi-dimensional nature, and the massive accumulated available literature on the topic, we set this section to present guidelines required to (a) understand the concept, (b) clarify the inherent ambiguity in the topic, and (c) delineate its use in our present research. We will briefly offer (a) a synopsis of various conceptualisations of the concept of attitude as they have evolved in the history of social psychology, and (b) the relevant applications of the concept of attitude in L2 pronunciation learning in our research context.

3.1 The Concept of Attitude

Accumulated literature in social psychology on attitudes shows the multi-dimensionality of the concept, both its complexity and ambiguity that Kelley (1989) eloquently described as “a circus tent over a diverse set of sideshows” (p. vii). The inherent confusion in the attitude concept stems probably from the lack of consensus among social psychologists on both attitude theoretical conceptualisation and empirical operationalisation. In the absence of a consensual research framework among attitude researchers, social psychologists have largely investigated (a) the attitude concept as a latent, hypothetical construct involving, individually or in concert, affect, behaviour, and cognition; and (b) the way these structures are formed, represented in memory, and changed as a function of experience. According to Oskamp and Schultz (2005), the largely proposed conceptualisation of attitude in modern social psychology include (a) the tripartite view (b) the separate entities view, and (c) the latent process view. The tripartite view considers attitude as a single entity that is translated into an appraisal of an attitude object[34] expressed in terms of favourability or unfavourability with respect to the object’s affective, behavioural, and cognitive constructs, implying a consistent degree of correlation between the three constructs. The separate entities view considers attitude as an underlying affective construct likely to have a directive role in generating a “behavioural intention” with respect to the attitude object (Fishbein & Ajzen, 1975, p. 288). The latent process theory considers attitude an underlying response that may manifest, individually or in concert, in affective, cognitive, or behavioural responses, involving no specific relationship

---

[34] An attitude object is a generic expression used in attitude research to refer to physical objects, individuals, places, concepts, values, etc.

between the three constructs (Albarracín, Johnson, Zanna, & Kumkale, 2005). In the following section, we offer fundamental explanations about the nature of the attitude concept as outlined by the latent process view as it involves a much clearer framework for investigation at both theoretical and empirical levels. The advantages of the latent process view are that (a) it does not equate affect with attitude at the theoretical level, thus causing less conceptual confusion; and (b) it offers the possibility to measure individually the three constructs and makes no prior direction for probable relationships between the three constructs.

3.1.1 Constructs of attitude.

Fundamental knowledge of the composite nature of attitude concept in social psychology is a prerequisite for a clear understanding of its attested ambiguity and varied operationalisation in research. The three, largely investigated constructs that make the core attitude concept, are (a) affect, (b) behaviour, and (c) cognition (or simply the ABC's of psychology). Given the massive volume of available literature on the three aforementioned constructs of attitude and researchers' highly integrative approaches in investigating the concept, we need to make clear that our intention in the immediate section is to offer neither a review of the existing literature on the topic nor to weigh their epistemic contribution. Rather, our intention is to offer straightforward and clear working concepts, drawn from some authors in the field, within the purview of our research interests, while relevant empirical evidence from attitude research, particularly L2 pronunciation learning, will be subsequently discussed briefly.

3.1.1.1 Affect.

Associated with appraisal mechanisms that have evolved for their adaptive value in fundamental life tasks, affect has rarely, if ever, been defined unanimously by psychologists but has been rather approached from a functional perspective, being a significant probable behavioural drive in the individual's life. Nonetheless, affect can be defined as a first-hand experienced ephemeral state of mind, elicited (consciously or unconsciously) as response to an internal or external stimulus, whether it has physiological underpinnings or not (Ekman, 1994). In a comprehensive and clear review of research on affective experiences and their taxonomy, Schimmack and Crites, Jr. (2005) suggested several types of affective experiences and the way they would differ from each other in terms of causes and consequences. For Schimmack and Crites, Jr. (2005), there are three types of affective experiences: (a) emotions, (b) moods, and (c) sensory affective experiences. The three types can be described in terms of dimensions such as (a) quality, (b) intensity, (c) duration, (d) object-directedness (i.e. source of elicitation), and (e) awareness of the affective experience. Emotions appear to be conscious, object-directed affective experiences as opposed to mood that is considered an unconscious affective experience elicited by no apparent object, and is determined by internal bodily states and energetic arousal. Sensory experiences are rather affective states that (a) are elicited in response to sensory stimulation, regardless of the degree of consciousness involved; and (b) are based on simple associative learning mechanisms (Schimmack & Crites, Jr., 2005, p. 406). Yet, more importantly, all affective experiences can be measured against a bipolar continuum, the ends of which represent opposite ends of the five above-mentioned dimensions.

Direct consideration of the previous notes in the field of SLA research would reveal the relevance and importance of experienced emotions within the learning and social contexts in forming and shaping a learner's attitude towards L2 learning, involving language skills, learning context, professional career goals, and other related issues. Experienced emotions, with respect to some attitude objects such as use of RP English in classroom conversations and confidence in one's abilities in possessing skilled RP pronunciation habits, differ in the extent to which they reflect negative or positive attitudes about these attitude objects. For example, experienced emotions that imply positive attitudes towards use of RP English in classroom conversations may involve feelings of personal satisfaction, self-confidence, and comfort. However, emotions that imply negative attitudes towards use of RP English in classroom conversations may involve feelings of dissatisfaction, diffidence, and discomfort. It is commonplace among attitude researchers (a) to think of affect as being positive or negative with respect to an attitude object; and (b) to assume that the more positive an individual feels about the attitude object, the more likely the individual will form a positive attitude towards it and vice versa.

Conceptualisation of affect as being either positive or negative with respect to an attitude object is analogous with scaling experienced emotions onto the attitudinal dimension for that object in terms of favourability or unfavourability. An individual expresses the intensity of favourability or unfavourability in degree of negativity or positivity of experienced affect by assigning scale values on the underlying attitudinal dimension. For example, with respect to RP English use in classroom conversations, learners of English may express their degree of favourability or unfavourability towards use of RP English by assigning a scale value from an x-number of suggested values that represent the intensity of experienced emotion. The most typical ordinal attitudinal scale for attitude measurement is the Likert scale (Breckler, 2004).

3.1.1.2 Behaviour.

It is common practice in modern psychology to consider behaviour in its (a) simple form, i.e. a response that directly produces an observable effect; or (b) underlying form, i.e. an indirect cause of a series of mediated events leading consequently to an observable effect (Skinner, 1957, p. 1). The former (explicit response) is commonly categorised as an exemplar overt behaviour, while the latter (implicit response) is categorised as a typical covert behaviour (Jaccard & Blanton, 2005, p. 127). Although both forms of responses represent a specific type of behaviour, producing immediate and intermediate effects, their conceptualisation is guided by divergent theoretical standpoints, behaviourism and cognitivism, respectively. Consistent with the line of thought in differentiating the three components of the attitude concept as suggested by mainstream attitude researchers, it stands clearly that the behavioural construct in attitude involves explicit behaviours alone. Thus, there is much reason to believe that implicit behaviours, which take the form of mental reactions or judgements, would form an entity that best fits in with the cognitive construct of attitude, and would therefore be the object of focus subsequently.

Assumed significance of behaviour in attitude formation has proliferated interest in a large array of behaviours, including those that psychologists investigate for purposes of (a) testing to make theoretical advances and insights or (b) recognising personal, social, and societal significance. Considerable literature on attitude research reveals researchers' efforts to extend theoretical advances to address behavioural issues of social and applied significance as in politics, marketing, education, etc. (Jaccard & Blanton, 2005). Indeed, SLA researchers have regularly pursued valid extension of attitude theory to L2 learning behaviours touching upon voluntary actions taken by learners to achieve explicitly-stated goals with respect to the L2 learning and instruction. With respect to learning L2 skills and language elements, L2 learners' behaviours differ in the extent to which they reflect positive or negative attitudes towards that object. For example, learning behaviours that imply a positive attitude towards English pronunciation among learners are likely to include their efforts to train themselves at using correct pronunciation by means of articulation activities, perception training, and self-monitoring tasks. Conversely, learners' behaviours that imply their negative attitudes towards English pronunciation are likely to exclude English learning activities that do encourage the aforementioned pronunciation activities. Categorical distinction of English pronunciation learning behaviours supposes that the more positive attitude learners of English evaluate English pronunciation, the more likely they engage in pronunciation activities and vice versa. Furthermore, learners of English may self-report degree of favourability or unfavourability about English pronunciation on an x-number of values standing for the frequency of occurrence of the behaviour.

3.1.1.3 Cognition.

The cognitive construct in attitude research bears on the concept of perception being an essential element in the acquisition or formation of beliefs, their representation in memory, and potential change over time as a result of learning experiences (Fishbein & Ajzen, 1975; Wyer, Jr., & Albarracín, 2005). As beliefs relate to mental processes involved in interpreting and making sense of events and information available to individuals, they are probably the most salient feature of attitude and the most investigated one. Far from purely computational processes implicated in representation of information about an attitude object, the importance of belief formation and change in attitude research lies in the relevance of beliefs as a source of knowledge and their estimated strength. Considering individuals as rational organisms who use information at their disposal to make judgements and decisions, Fishbein and Ajzen (1975) emphasised the importance of beliefs as the cornerstone of the attitude concept, claiming that,

> On the basis of direct observation or information received from outside sources or by way of various inference processes, a person learns or forms a number of beliefs about an object. That is, he associates the object with various attributes. In this manner, he forms beliefs about himself, about other people, about institutions, behaviors, events, etc. (p. 14)

For Fishbein and Ajzen (1975), beliefs an individual has about an attitude object refer to the subjective probability judgements of the relationship between the attitude object of belief and some other object that is essentially an attribute. To put it differently, forming a belief about an attitude object implies the establishment of a likely relationship between that attitude object and some other attribute. For example, learners of English may assign different attributes to some attitude objects such as RP English, English vocabulary, and English grammar, leading them to form distinct beliefs about these language aspects. Learners, who may believe in the existence of a relationship between RP English and the concept of importance or triviality, may also believe in the existence of a relationship between English vocabulary and primacy or insignificance. With respect to wording beliefs, it is archetypal in attitude research to convey verbally the referent of beliefs in meaningful propositional statements that individuals verify as true or false, or simply assert or deny. Wyer, Jr., and Albarracín (2005) articulated the source of knowledge from which individuals might form about the world they live in, as deriving from direct experience with knowledge referents (i.e. first-hand experience) or cognitive operations individuals have already acquired. Thus, for example, learners of English as an L2 may believe that use of RP pronunciation is important to pass the language test from evidence that their English oral performance has been considerably graded down once because of their accented speech. Similarly, the L2 learners may believe that direct contact with the native speakers of RP English will improve their English pronunciation from statistical evidence demonstrating a significant relationship between skilled pronunciation habits and direct contact with native speakers of the language.

Stability and change are two topics of interest that add up to the relative difficulty of conceptualisation of beliefs that pertain to the individuals' knowledge. Significant attitude research has demonstrated the relatively stable nature of beliefs, which are knowledge-relevant representations that dwell in the long-term memory structure and appear to be retrieved rather than constructed on the spot from temporarily accessible information. Eagly and Chaiken (1993) argued in favour of this standpoint as they suggested, "in daily life direct retrieval may be the rule rather than the exception" (p. 112). However, some attitude researchers have argued against the presumed relative stability of the attitude constructs and argued in favour of their fluctuating nature as Potter (1998) suggested, "the same individual can be found offering different evaluations on different occasions, or even during different parts of a single conversation" (p. 244). However, regardless of either argument's evidence, there is reason to believe that, even if beliefs are supposedly long-term memory structures and are retrieved rather than constructed on the spot, new information derived from direct observation or novel first-hand experience is likely to cause the individual to review old beliefs and construct new ones to meet certain epistemic or expectation ends. On this ground, we may argue that (a) strong beliefs are most likely to be accessed as long-term memory structures; and (b) malleable beliefs are most likely to be constructed on the spot in light of temporarily available information, the latter being the cause of state of ambivalence.

An illustrative example with respect to an L2 learning will help understand the issues presented just above. Second language learners with strong beliefs in the importance of L2 pronunciation in learning the language are very likely to use memory to convey verbally their knowledge with respect to the referent of the belief, and learners with malleable beliefs are very likely to try to build on information available on the spot to evaluate the proposition. That is, if the learners happen to be thinking of the difficulty of the writing skill and the amount of time required to improve their skills, they are likely to scale down the importance of L2 pronunciation. However, at a second time, if the L2 learners happen to be thinking of the praise they have received from their teachers or classmates for their unaccented L2, they are very likely to scale up L2 pronunciation importance, demonstrating a considerable ambivalence.

3.1.2 Attitudes in second language pronunciation learning.

There is little doubt that success in L2 pronunciation learning is not only bound to teachers' pedagogical preferences, informed practices, and selected syllabi contents, but it extends to psychological predispositions of learners that go beyond the latter's natural aptitude. Interest in social psychological predispositions has generated several researches investigating the likely substantial effect of attitudes on L2 pronunciation learning, with equal respect to language elements or the learning context in general. However, given the inherent vagueness of the attitude concept, researchers' resort to interest-tailored approaches involving attitudes, without setting an agreed-upon conceptual framework for the concept, has resulted in extensive collection of papers and books explicating the role of attitude as an affective sub-variable accounting for motivation and individual variations in L2 pronunciation attainment and proficiency. Beliefs and behaviours in L2 pronunciation learning have been approached by researchers within a general psychological framework, bearing mostly on theoretical premises of cognitivism and least on the psychology of attitude.

The focus of research in L2 pronunciation has been put on the latter's teaching and learning, including pronunciation instruction at different levels of education, the L2 model to teach, teaching and instructional methods, prioritisation of segmental or supra-segmental aspects, L2 pronunciation difficulties, the role of corrective feedback, pronunciation learning strategies, teachers' perceptions of L2 pronunciation usefulness, etc. (Pawlak, Mystkowska-Wiertelak, & Bielak, 2015; Szyszka, 2017). Notwithstanding this varied interest, the focus in L2 pronunciation research has been unbalanced, and researchers have given much emphasis to pedagogic activities that take place inside the classroom and have neglected the importance of several individual factors in determining attainment and proficiency in L2 pronunciation. Indeed, with respect to L2 pronunciation learning at the tertiary level, researchers have done little (a) to investigate empirically the way L2 learners' beliefs about pronunciation are formed and (b) to determine the causes that generate learning behavioural patterns that likely influence L2 pronunciation attainment and equally determine the nature of classroom interaction (Dörnei, 2005). Cenoz and Lecumberri (1999) admitted the supplementary nature of L2 pronunciation instruction and teaching material in syllabi regardless of pronunciation relative importance in L2 learning. Cenoz and Lecumberri (1999) stated,

> Researchers have devoted less attention to the acquisition of other linguistic areas. As far as second language teaching is concerned, pronunciation is becoming more important in languages classes and teaching materials…, but it is still considered more a supplementary activity rather than a central part of the syllabus. (p. 4)

With respect to L2 learners' beliefs about L2 pronunciation learning and teaching, it is not until recently that researchers have started to place emphasis on the way learners perceive the importance of L2 pronunciation in their curriculum and their professional life as well. The scant interest in the latter is probably the result of received beliefs about the utility of native-like pronunciation when learning an L2 and the historically long-held belief in the supposedly maturational effects on L2 pronunciation learning. Furthermore, current practices and foci of the communicative approach in L2 learning have considerable trivialised pronunciation and its instruction, as they back away serious interest in defining clear objectives of teaching L2 pronunciation aside from intelligibility in communication.

In an exploratory research on the role of attitudes, awareness, and beliefs in learning pronunciation among 86 Basque and Spanish learners, Cenoz and Lecumberri (1999) revealed some interesting findings about learning English pronunciation independently of first language. Cenoz and Lecumberri (1999) found that the participants (a) were aware of both the difficulty and importance of English pronunciation as a skill; (b) considered contact with native-speakers of English and perceptual training as the most influential factors in English pronunciation; and (c) considered difficulty of some English accents to be a matter of attitude. The authors called for further investigations in other L2 learning contexts to confirm the observed trends in their research and provide further information about phonetic awareness, beliefs, and attitudes in L2 acquisition and their potential relationships with L1 of the learners.

In a similar vein of research exploring learners' beliefs about the learning and teaching of English grammar, pronunciation, and vocabulary at the tertiary level among native speakers of Dutch, Simon and Taverniers (2011) reported,

> Learners considered vocabulary to be significantly more important for efficient communication than pronunciation or grammar and reported that, in general, vocabulary errors are significantly more likely to lead to communication breakdown than errors in the other two components. (p. 912)

Moreover, Simon and Taverniers (2011) found that participants reported high-perceived confidence in their abilities to attain native-like proficiency in the three language components, and that there were statistically no significant differences between them. Confirming the just reported trend of findings with relation to L2 pronunciation learning, Pawlak et al. (2015) reported similar results regarding beliefs of Polish major learners of English about the need to speak the language in a native-like manner and the desire to speak English with no Polish accent, regardless of the difficulty of the task. Out of several interesting findings, Pawlak et al. (2015) found that participants in their study (a) had positive beliefs about the importance and utility of learning segmental and supra-segmental features of English pronunciation; (b) had positive attitudes towards the role of phonetic learning in effective communication; and (c) believed that skilled pronunciation would enhance their perceptual abilities.

In recent survey of attitudes towards English pronunciation learning among Polish learners, Waniek-Klimczak, Rojczk, and Porzuczek (2015) reported that learners showed significant positive attitudes towards the attitude object, and that learners' goals in learning pronunciation might change with experience, as MA learners demonstrated significantly less positive attitudes towards pronunciation than BA students. Likewise, Szczygłowska (2017) confirmed Polish learners' (a) favourable appraisal of proficient English pronunciation habits, and (b) indecisiveness as whether to value pronunciation more than the other language components.

Conclusion

The complex nature of the attitude concept is inherent to the lack of consensus among social psychologists, as it involves the ABC's of psychology, affect, behaviour and cognition. Therefore, any conceptualisation of attitude would not be without theoretical and empirical costs. For the purpose of our research work, we opt for the tripartite conception of attitude, for it allows us to investigate a plethora of phenomena capable of offering the chance to gain a better insight into affect, perceptions and behaviours within the field of L2 learning.

Chapter 4

Methodology

Introduction

In this chapter, we formulate the four research questions that motivated the present research. As mentioned throughout Chapters 1, 2, and 3, L2 phonological learning is subject to several factors affecting both perception and production. Acquisition of an L1 as a cognitive skill is highly responsible for the formation of automatic speech perception and production processes adapted perfectly for fast and efficient processing of speech input and generating speech output. Formation of automatic processes for L1 speech perception and production and formation of L1 sound inventory are very likely to be responsible for reduced attention in capturing full information available in L2 speech, leading to the use of other cognitive strategies to compensate for unrecovered information in the speech signal. This would-be process of cognitive compensation in L2 speech perception is most likely the cause for pronunciation problems encountered later by L2 learners. As reviewed in Chapter 2 in speech perception models, learning experience may improve learners' perceptual abilities in recovering phonetic detail that is presumably unattended to because of L1 automatic processes for discrimination and identification of linguistically relevant detail. Additionally, there is reason to believe that improvement in perceptual abilities of L2 speech is not enough to assure skilled pronunciation of an L2, as further social psychological factors such as attitudes may mediate practices required for proficient and unaccented L2 pronunciation.

Within this framework, the research questions of the present work address the question of perceptual abilities of Algerian learners in discriminating and identifying RP English prototypic monophthongs, as they are phonetically different from both learners' native language and French language to which they have been introduced in the primary school at a pre-pubescent age. We think that attributing difficulties and concerns in RP English pronunciation among Algerian learners to their wrong conception of letter-to-sound correspondence in English, likely originating in frequent regular associations between spelling and pronunciation in French (Beghoul, 2007), is not enough to account fully for pronunciation problems. Empirical evidence regarding whether Algerian learners of RP English have an undistorted perception of English sounds is very scant, and their beliefs about the importance of RP English pronunciation in learning the language are still unidentified.

The following beliefs motivate the present research interest in perception of RP English prototypic vowels:

(a) L1 vowel size inventory may not accurately predict non-native language perceptual difficulties.
(b) An interplay between spectral and temporal features of vowels may predict L2 vowel perception.
(c) Perception of L2 vowels may be less categorical.
(d) Isolated and in-context L2 vowels may exhibit different perceptual patterns.

4.1 Research Questions

The following questions guide our research work:

(a) What is the pattern of discrimination of some x-step synthetic continua of some selected RP English prototypic vowel contrasts among Algerian learners of English?
(b) What type of perceptual cues do Algerian learners use in identification of RP English prototypic monophthongs? Temporal or spectral?
(c) What type of effect do contextual constraints have on identification of RP English prototypic monophthongs among Algerian learners of English?
(d) What type of attitudes do Algerian learners of English hold about RP English pronunciation learning and instruction?

(e) Is there a relationship between Algerian learners of English attitudes towards pronunciation and their speaking proficiency?

4.2 Research Hypotheses

To provide answers to the above-mentioned research questions, we set the following hypotheses:

(a) If Algerian learners of English use spectral cues in the perception of English prototypic monophthongs, then they would discriminate between them efficiently (i.e. accurately and quickly).

(b) If Algerian learners of RP English employ equally both spectral and temporal cues in identification of RP English prototypic monophthongs, they would exhibit similar identification patterns across all monophthongs.

(c) If vowel identification improves with contextual constraints, then identification of RP English monophthongs would exhibit different perceptual patterns when isolated and in context.

(d) If there exists a relationship between Algerian learners of English reported attitudes towards English pronunciation and oral performance, then they would be positively significantly correlated.

4.3 Methodology

This section describes the methodology of the current research. It describes (a) the characteristics of participants and criteria for their selection, (b) presentation of the experiments speech stimuli designed to elicit participants' data, (c) the procedure adopted to carry out experiments and survey for data collection, and (d) the rationale for using specific tools for analysis.

4.3.1 Participants.

The sample of the research comprised initially 58 participants. All of the participants were 3rd year students at the *École Normale Supérieure Assia DJEBAR de Constantine (Higher College for Teachers Training of Constantine)*, enrolled on an English course for a degree of *Professeur de l'Enseignement Secondaire (Secondary School Teacher)* and *Professeur de l'Enseignement Moyen (Middle School Teacher)*. Five of the participants were excluded from the investigation for violation of eligibility requirement. Excluded participants failed the medical history screening procedure for self-reporting previous hearing and dyslexic problems. The final sample consisted of 50 female and 3 male students, with a mean age of 21.42 years (*SD* = 0.57). Table 1 below shows general characteristics of participants with respect to their biographic data, hearing medical history, previous linguistic and extra-linguistic experiences that we used as a screening procedure not to compromise results and analyses and make sure the participants would form a homogeneous sample in terms of language input.

We set our choice on third year students (a) because they received an introductory English phonetics and phonology course for two years, presumably sufficient to help them have explicit knowledge of English language sounds and phonological system; and (b) to enquire into phonetics and phonology course effectiveness in having particular impact on participants' perceptual abilities of RP English monophthongs.

Table 1

Participant Characteristics

Biographical data (N = 58)					
	Mean	*SD*	Male	Female	
Age	21.42	0.57	6	52	
Hearing medical history (N = 58)					
Hearing medical history	Positive	Negative			
	5	53			
Hemispheric dominance	R-handed	L-handed			
	51	7			
Dyslexia medical history	Positive	Negative			
	1	57			
Family dyslexia medical history	Positive	Negative			
	0	58			
Linguistic experience					
Native language	Algerian dia.	Berber	French	Other	
	52	6	0	0	
First language learned at school	Stan. Arabic	French	Berber	Other	
	58	0	0	0	
Second language learned at school	French	English	Spanish	German	
	58	0	0	0	
Living abroad experience	Applicable	Non-app.			
	0	58			
Language tutoring course	Applicable	Non-app.			
	15	43			
Onset of language tutoring assistance	L-limit	U-limit	Mean	SD	
	5	20	15.21	4.15	
More than 1 year of language tutoring (N. of app. = 15)	Arabic	French	English	Spanish	German
	3	11	6	1	2
Prior-university pronunciation training	Applicable	Non-app.			
	0	58			
Frequent exposure to English	NN Eng.	Am. Eng.	Br. Eng.		
	9	28	21		
Extra-linguistic experience					
Music instruction	Applicable	Non-app.			
	26	32			
Onset of music instruction (N. of app. = 26)	L-limit	U-limit	*M*	*SD*	
	11	22	12.23	3.34	
Type if instruction (N. of app. = 26)	School subj.	Private tut.	Both		
	22	1	3		

Note. Some words were shortened for word processing reasons. dia.: dialect. Non-app.: non-applicable. L-limit: lower limit. U-limit: upper limit. School subj.: school subject. *M*: mean. Private tut.: private tutoring. *SD*: standard deviation. NN Eng.: non-native English. Am. Eng.: American English. Br. Eng.: British English. N: number. R-handed: right-handed. L-handed: left-handed. N. of app.: number of applicable cases.

4.3.2 Research design.

We designed three experimental tests in this study to investigate participants' perceptual abilities, one discrimination test and two identification tests, same-different *AX* and *m*-alternative forced choice tests (*m-AFC*), respectively. We opted for these types of test for their simplicity in synthesising and administration. Same-different *AX* discrimination tests are self-evident and do not require complex instructions. Participants would not need to identify consciously either the similarity or difference between the presented stimuli, but simply to report it. In designing the experiments reported here, we used both paradigms developed by Kuhl (1991) and Best (1995) to assess participants' perceptual abilities of discrimination of vowel tokens varying in spectral distance from prototypic vowels and to find out empirical threshold of sensitivity and pooled sensitivity indices across compared vowel pairs.

We used a 20-item questionnaire developed to measure the three constructs making the attitude concept. We designed the questionnaire to explore Algerian learners of English attitudes about pronunciation learning and instruction. Questionnaire design was based on some personal reflections on the topic prior to finding out about existing questionnaires in the field within the Polish context as described in the literature review.

For spectral values of RP English pure prototypic vowels, we used Deterding's (2006) measurements. As against measurements of RP English pure vowels' spectral properties based on the '*The North Wind and the Sun*', as suggested by the International Phonetic Association since 1912, Deterding (2006) provided measurements based on recordings of '*The Boy who Cried Wolf*' passage. The latter allows for more comprehensive and representative measurements of spectral characteristics of RP English pure vowels in various phonetic contexts (Deterding, 2006, p. 187).

For empirical sensitivity threshold and sensitivity indices, we used Signal Detection Theory as developed by MacMillan and Creelman (2005). Our choice was set on this theory and not alternative theories, like that of the Choice Theory for Luce (1977), for its suitability in the field of automaticity and relative simplicity in calculations and availability of results. Empirical threshold of sensitivity and pooled sensitivity indices were obtained from methods discussed by MacMillan and Creelman (2005, Chapter 9, pp. 213-243, using Appendix 5, Table A5.4, pp. 401-419, and Chapter 10, pp. 245-266, using Appendix 5, Table A5.7, pp. 426-430).

4.3.3 Stimuli materials.

Stimuli materials in both Experiments 1 and 2 were modelled after Deterding's (2006) spectral measurements of F_1 and F_2 of RP prototypic monophthongs (see Appendix2[35] for synthesised vowels' spectral detail). Stimuli were synthesised utilising PRAAT's Vowel Editor, version 6.0.14 (Boersma, 2013; Boersma & Weenink, 2016). Stimuli in in-context vowel identification test were also synthesised utilising the previously stated version of PRAAT, using Speech Synthesizer, set at RP English with a natural male voice variant (m7). Pairwise combinations of stimuli in same different vowel discrimination test were created using WavPad Sound Editor by NCH Software, version 6.55. All utilised software were ran on a professional desktop computer, Hewlett-Packard HP 600B Series.

4.3.4 General procedure.

We carried out all experiments by providing participants with clear explanations of the research objectives, precise instructions on how to take the test, and full account of availability arrangements needed to carry required experiments. Prior to all experiments, we informed participants that they would be equally rewarded for their time in course evaluation score, with additional compensation available under mutually agreed-upon conditions, regarding completion of all scheduled experiments and personal commitment. Experiments took place in

---

[35] Due to word processing reasons and nature of stimuli materials used in the study, all appendices are made available in a digital support.

a language laboratory at the *École Normale Supérieure Assia DEJBAR de Constantine*, with the help of the laboratory staff.

Throughout all experiments, we allowed participant to work on individual workstations and presented stimuli for them via professional, high quality language earphones, TH-952, at a most comfortable listening level (MCL) for loudness, 70 dB (Hochberg, 1975, p. 30). Stimuli presentation and collection of response files were managed by the use of TP Perception Tests/Perception Training Tasks application, Version 3.1, running on laboratory workstations. The choice of TP software was made because (a) it does not require expertise in programming languages; (b) it has a user-friendly interface; and (c) it allows visual, auditory and audio-visual testing conditions (Rato, Rauber, Kluge, & Santos, 2015). We made sure participants logged in using their full names to appear in the test auto-saved result files and to ensure data authenticity, reliability and fast processing of collected data.

4.3.5 Tools of analysis.

Data were compiled and analysed using Microsoft Office Excel 2013 and IBM SPSS 25.0. Initially, we used Microsoft Office Excel 2013 to (a) process data collected from auto-saved result files for all experiments and (b) plot synthetic stimuli on the vowel two-dimensional grid. Then, we used IBM SPSS 25.0 to process data, perform calculations and report descriptive and inferential statistics results, including the arithmetic mean, the standard deviation, *t*-test, one- and two-way ANOVA tests with and without repeated measures, factor analysis and multiple regression.

Chapter 5

Same Different Vowel Discrimination Test

Introduction

Prior to same different vowel discrimination test, we administered a pilot study with the classical AB*X* test to 64 participants reading for a degree in English at the *École Normale Supérieure Assia DJEBAR de Constantine*. The test consisted of a series of adjacent pairs of vowel tokens, designed to investigate participants' discrimination ability of presented stimuli. We synthesised adjacent pairs of stimuli, starting at one end of the vowel continuum that forms a prototypic English vowel and heading towards the other end making the other prototypic vowel. That is, if stimuli 1 and 8 were the prototypic vowels, adjacent stimuli pairs would form this series of stimuli 1_2, 2_3, 3_4 up to 7_8, with intermediate vowel tokens synthesised at fixed spectral distances in F_1 and F_2 from each other for all adjacent pairs.

Obtained results of the experiment showed no performance variability among participants, and the latter reported to have found the experiment extremely difficult. Inconsistent data complicated our state of affair, for the data revealed no specific pattern and were inconsistent for further processing. Participants reported almost all stimuli pairs across vowel comparisons to be the same with very few exceptions. These reasons caused us to modify the initial experiment design and to opt for a manageable one that would cause less difficulty and yield reliable and valid data for analysis. Therefore, we used the A*X* discrimination test instead.

In same different vowel discrimination test, we examined participants' sensitivity to vowel tokens spaced at equal intervals in a two dimensional grid. These equal intervals were systematically created by varying the spectral differences in RP English vowel token pairs with equal steps, with stimuli ranging concurrently in F_1 and F_2 along a continuum, the end-points of which make two English prototypic vowels. Spectral distance between vowel tokens varied across different vowel pairs, following spectral properties inherent in prototypic vowels, while maintaining maximum spectral difference for either, F_1, or F_2 equal or less than 62 Hz for all vowel pairs (see *Figures 6* through *15* for detailed spectral properties of all vowel tokens across compared vowel pairs).

We designed same different vowel discrimination test to:

a) Investigate the use of spectral cues in discriminating RP prototypic monophthongs and other synthetic vowels among Algerian learners of English.
b) Determine empirical threshold of sensitivity (minimal spectral distance capable of eliciting more than 50% of correct responses), and
c) Examine whether the participants' empirical threshold of sensitivity would be constant at a fixed spectral distance (i.e., regardless of vowel stimuli pair), or would be a function of compared prototypic vowels.

5.1 Stimuli Materials

Stimuli in this experiment consisted of forty three 180-ms vowel sounds whose duration was set at 180 ms and a steady fundamental frequency, F_0 set at 125 Hz, while maintaining constant pitch for an adult man, as suggested by Raphael, Borden, and Harris (2011, p. 82). We saved individual stimuli in separate 32-bit WAV files and then paired them, using WavPad Sound Editor, such that there were an initial 150-ms silence duration, a 250-ms inter-stimulus interval (ISI) and no final silence. We created four types of vowel pairs (9 series comprising 77 pairwise vowel combinations):

a) Thirty different non-pairs of prototypic vowels (2_7, 28_24, 47_49, etc.);
b) Eighteen pairs containing RP prototypic monophthongs (1_8, 23_29, 30_35, etc.);
c) Twenty adjacent vowel pairs (19_20, 20_21, 25_26, 26_27, etc.); and
d) Nine identical vowel pairs (4_4, 11_11, 14_14, etc.)

We did not include a large or an equal number of identical stimulus pairs for practical and analytical reasons. First, equal '*Same*' stimulus pairs to '*Different*' ones in the same block of trials would be more time consuming and exhausting for participants, a fact of which we were afraid to affect participants' commitment to complete all scheduled experiments. Second, '*Same*' stimulus pairs' results would be but thrown away, for they are needed for calculation purposes to report sensitivity and are per se uninteresting for analysis and interpretation.

To demonstrate graphically the rationale and undertaking of same different vowel discrimination test, we plotted and presented prototypic vowels on F_1 and F_2 dimensions in *Figure 6* and the synthetic ones in *Figures 7* through *15* (for detailed spectral values for all vowel tokens, see Appendix 2).

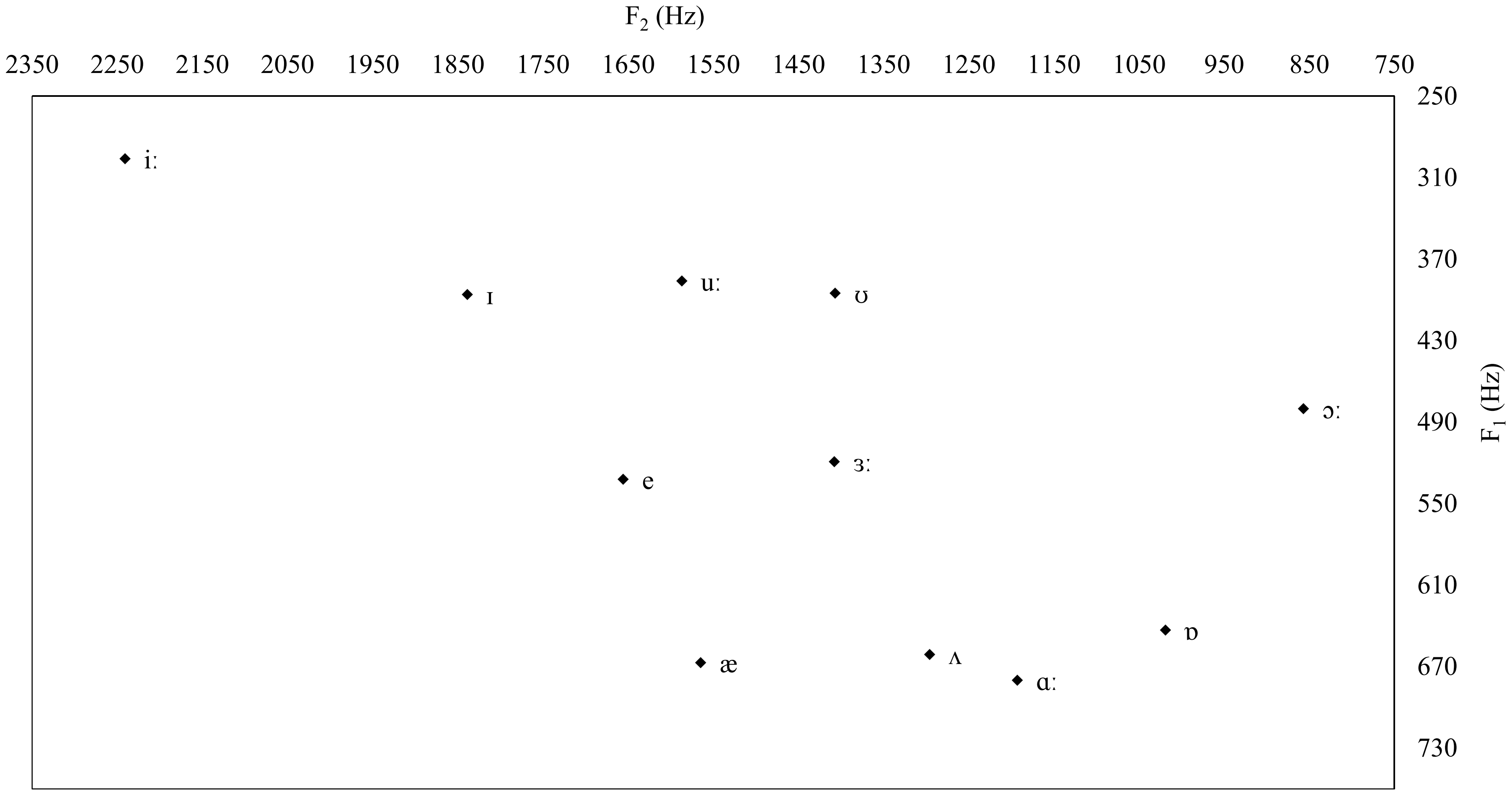


Figure 6. Plot of the first two formants for the RP pure prototypic monophthongs. From "The north wind versus a wolf: Short texts for the description and measurement of English pronunciation", by D. Deterding. Copyright 2006 by the International Phonetic Association. Adapted with permission.

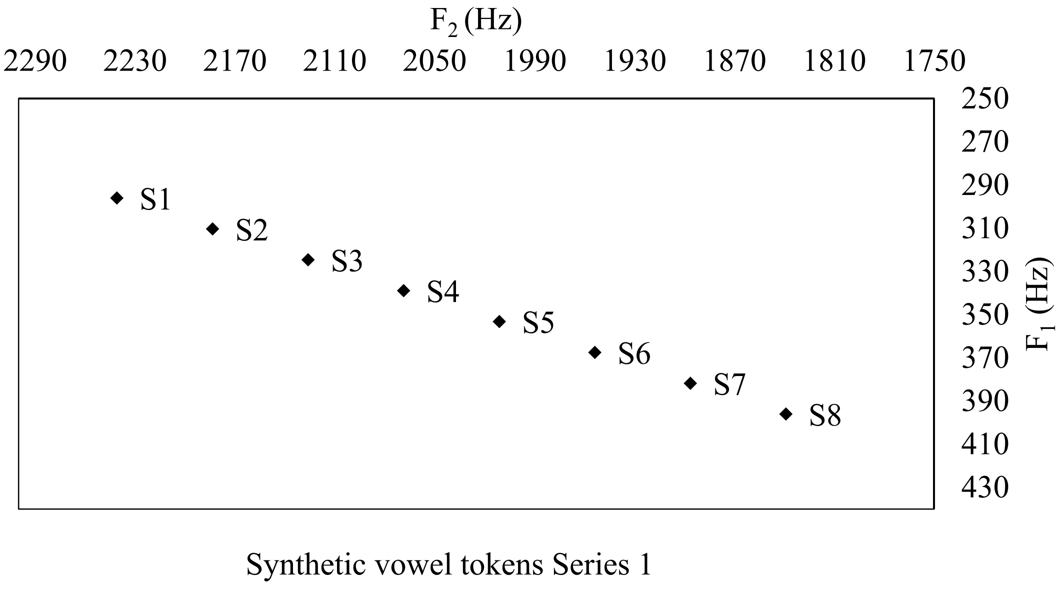


Figure 7. Plot of synthetic vowel tokens Series 1. Spectral distance: F$_1$ = 57 Hz; F$_2$ = 14 Hz. (S$_1$, S$_8$) = (/iː/, /ɪ/).

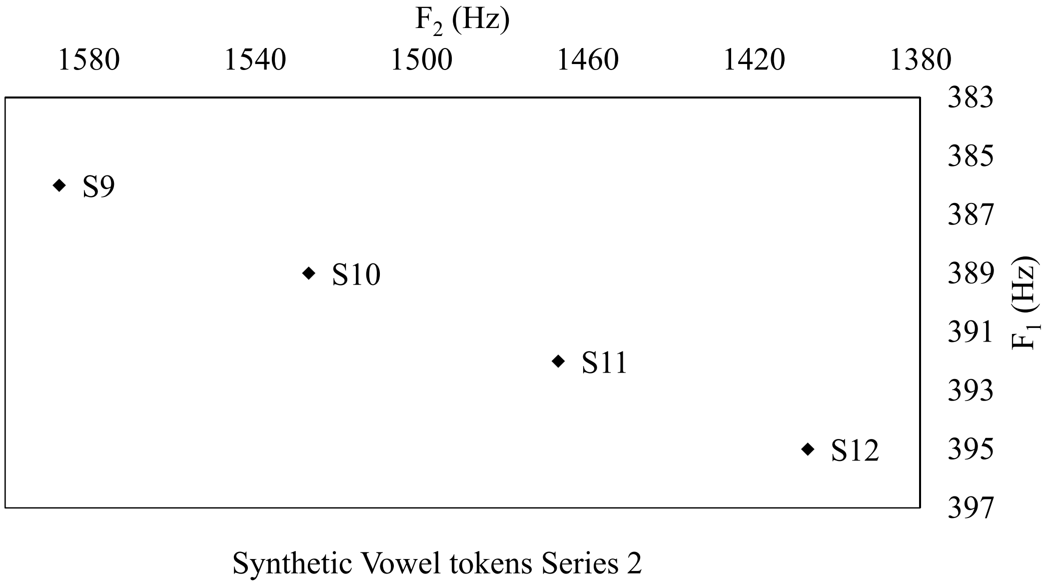


Figure 8. Plot of synthetic vowel tokens Series 2. Spectral distance: F$_1$ = 3 Hz; F$_2$ = 60 Hz. (S$_9$, S$_{12}$) = (/uː/, /ʊ/).

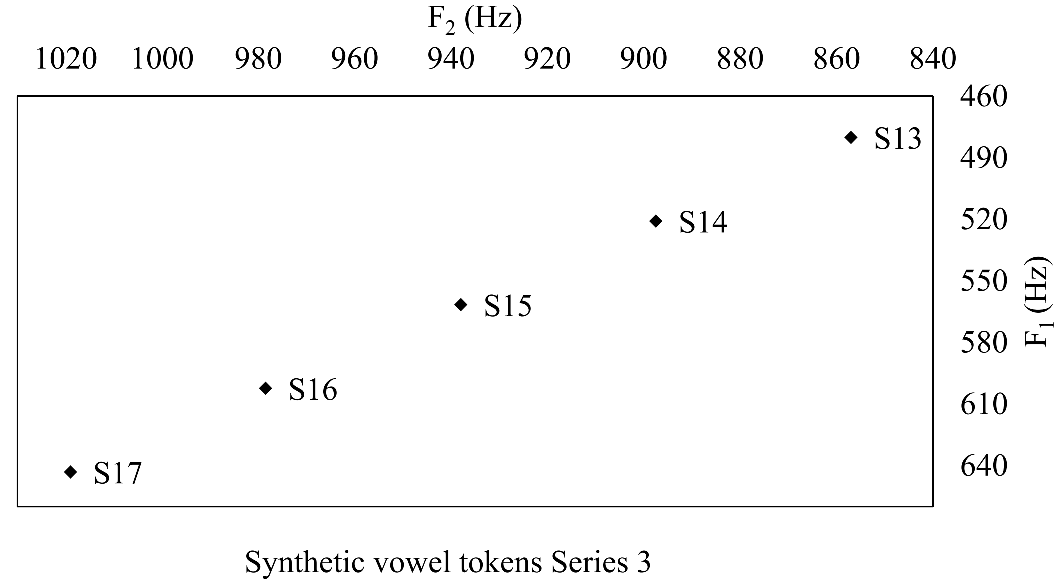


Figure 9. Plot of synthetic vowel tokens Series. Spectral distance: $F_1 = 41$ Hz; $F_2 = 40$ Hz. (S_{13}, S_{17}) = (/ɔː/, /ɒ/).

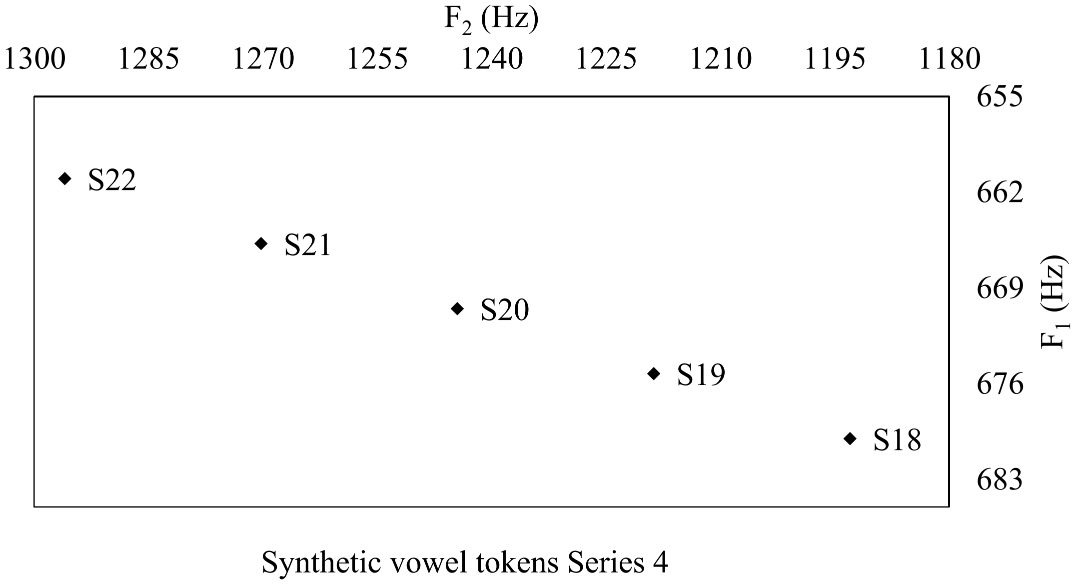


Figure 10. Plot of synthetic vowel tokens Series 4. Spectral distance: $F_1 = 5$ Hz; $F_2 = 26$ Hz. (S_{18}, S_{22}) = (/ɑː/, /ʌ/).

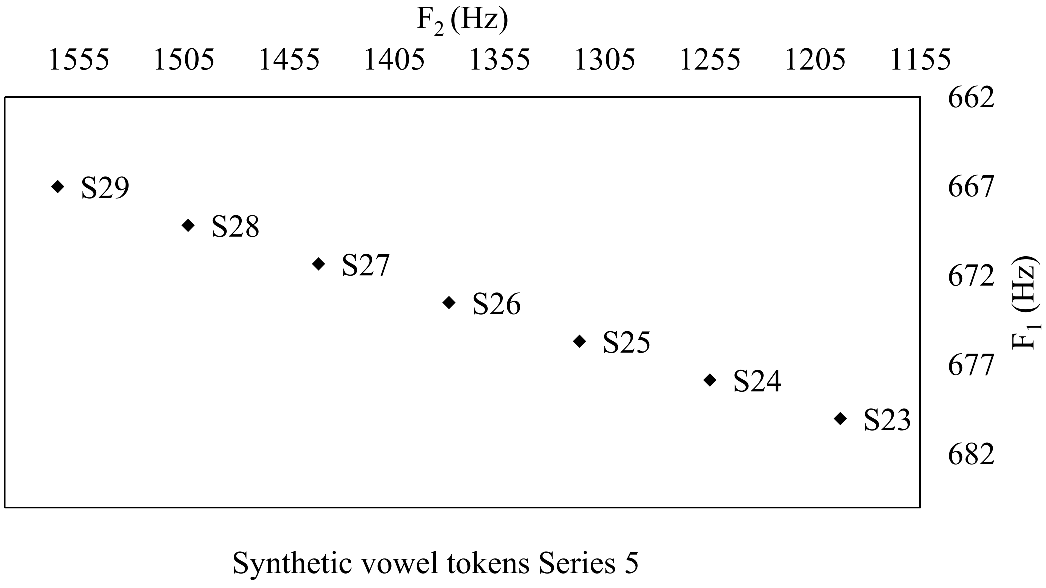


Figure 11. Plot of synthetic vowel tokens Series 5. Spectral distance: F_1 = 2 Hz, F_2 = 62 Hz. (S_{23}, S_{29}) = (/ɑː/, /æ/).

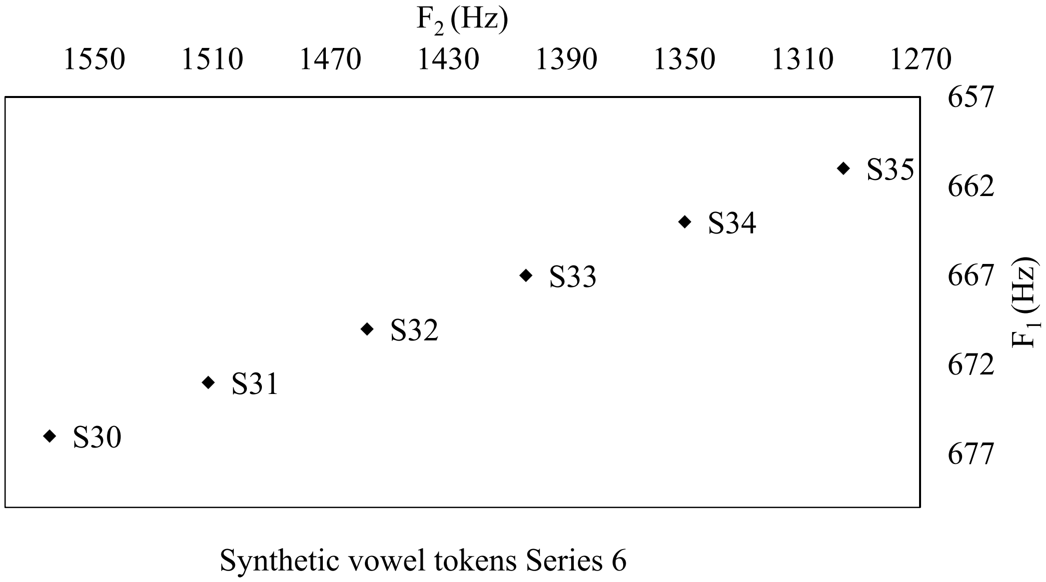
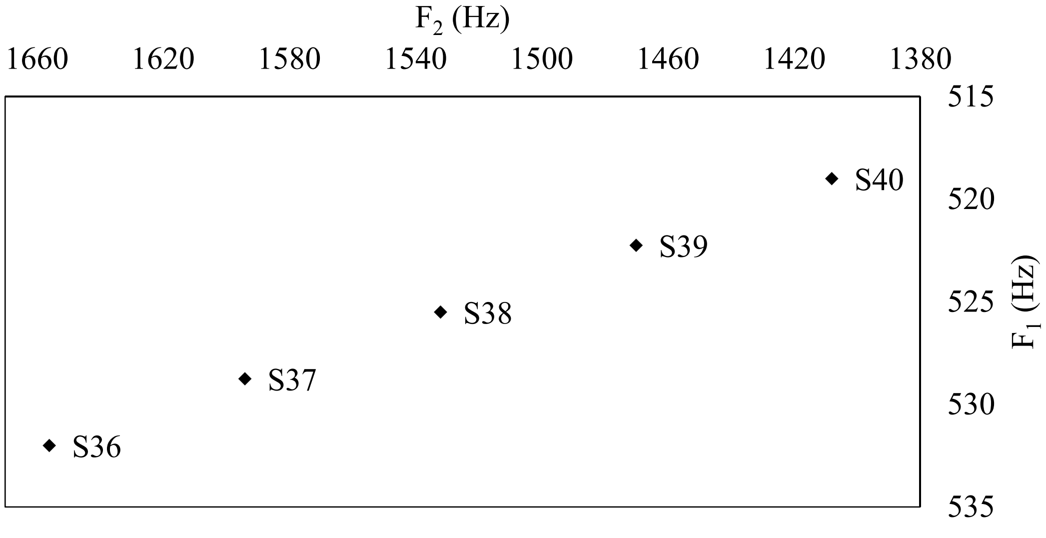
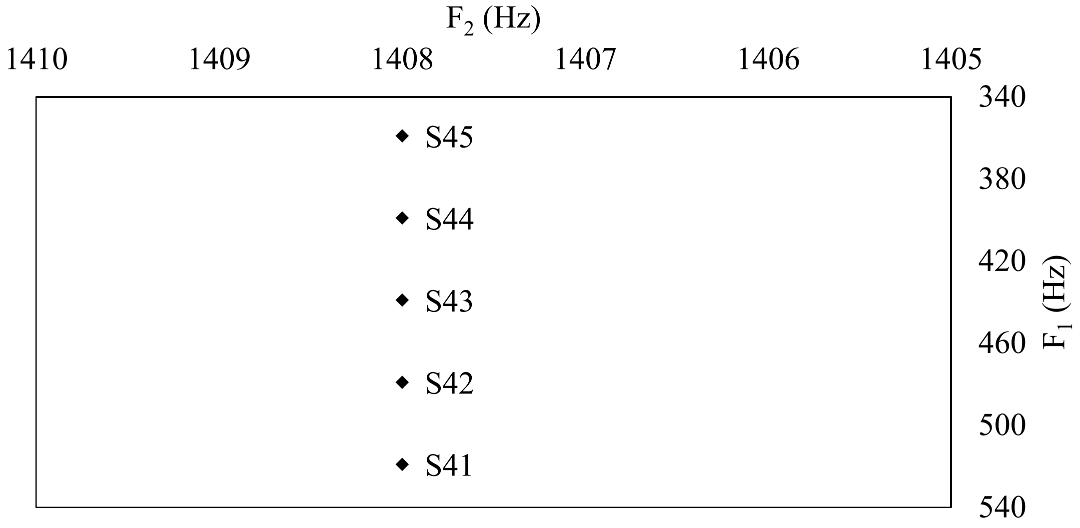


Figure 12. Plot of synthetic vowel tokens Series 6. Spectral distance F_1 = 3 Hz, F_2 = 54 Hz. (S_{30}, S_{35}) = (/æ/, /ʌ/)



Synthetic vowel tokens Series 7

Figure 13. Plot of synthetic vowel tokens Series 7. Spectral distance F_1 = 3 Hz, F_2 = 62 Hz. (S_{36}, S_{40}) = (/e/, /ɜː/).



Synthetic vowel tokens Series 8

Figure 14. Plot of synthetic vowel tokens Series 8. Spectral distance: F_1 = 40 Hz, F_2 = 00 Hz. (S_{41}, S_{45}) = (/ɜː/, /ʊ/).

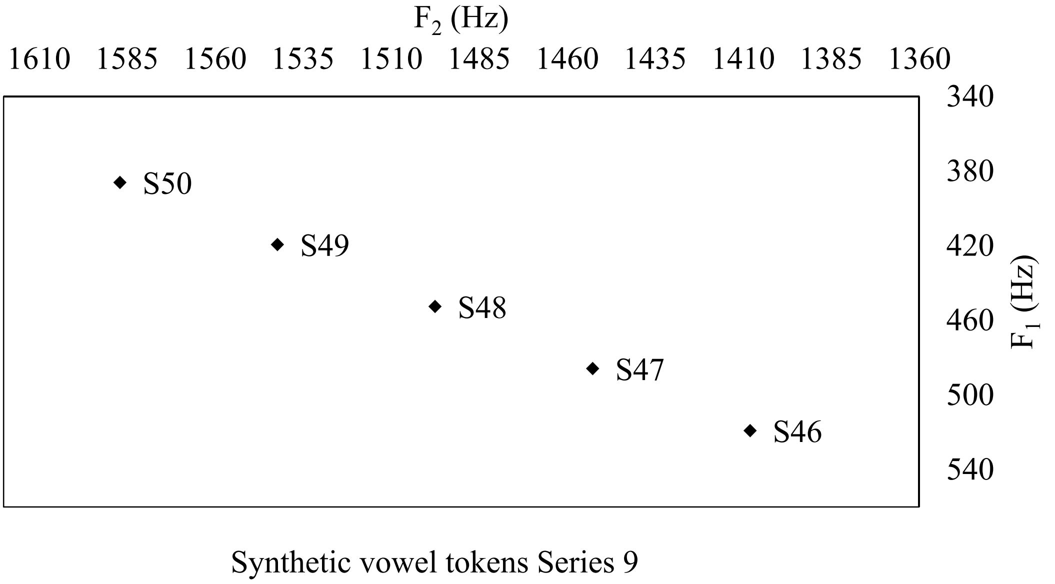


Figure 15. Plot of synthetic vowel tokens Series 9. Spectral distance: F_1 = 33 Hz; F_2 = 45 Hz. (S_{46}, S_{50}) = (/ɜː/, /uː/).

5.2 Procedure

Same different vowel discrimination test consisted of 16 blocks of trials, each with 50 trials. We did not count the first 30 trials, for we used them for familiarisation without informing the participants. Each trial began with the presentation of an alert message with task heading and instructions on the computer screen. Participants listened to a pair of vowel tokens in each trial and had to respond whether what they had heard was one same sound or two different sounds, simply by clicking on one of the buttons on the PC screen, '*Same'* or '*Different'*. The vowel tokens were paired such that, on any given trial, they were either the same or different. The button '*Same*' appeared to the left side of the screen, while the button '*Different*' appeared to the right side.

As posited by Pisoni (1973), there are two modes of sound perception, an auditory mode and a phonetic mode. The former is a highly detailed but quickly decaying trace memory of the sound, while the latter is a more abstract categorical representation of it. Upon responding to a trial, participants were given a 150-ms initial silence to hear the following vowel stimulus pair, and were orientated towards making a decision based on a highly short-term memory trace. In order to reduce memory load on participants and make them respond quickly, we encouraged them to consider their reaction time at the end of every block of trials and to try to have a shorter reaction time. Yet, we neither set a cut-off point nor favoured speed over accuracy. Vowel token pairs' presentation was randomised to avoid the carry-over effect and to prevent participants from developing any probable response strategy.

Upon completion of every block of trials, the *TP* software provided participants with immediate feedback and alerted them to their performance level for accuracy and reaction time. Participants had a systematic break after completion of every block of trials, during which we tried to alleviate the test pressure that had built up by encouraging them to report on any issue relating to the test. Participants resumed the test of their own volition. It took participants almost

4 hours to complete the test. Upon completion of the test, we collected test auto-saved result files of participants' data from their respective workstations.

5.3 Results

As data in same different vowel discrimination test were processed twice, we present results in two separate sections, *Raw data results* (Section 5.3.1) and *Modified data results* (Section 5.3.2). The former section displays detailed pooled results per stimulus pair and the latter displays detailed results for combined stimulus pairs and includes comments and basic analyses. In Section 5.3.1, '*Same*' and '*Different*' responses are presented as percentages. In the Section 5.3.2, hits and misses are presented as rates. Pooled sensitivity index *d'* values were based on the mean hit and false alarm rates for all participants, using roving (*differencing model*) methods as discussed by MacMillan and Creelman (2005, Chapter 9, pp. 213-243, using Appendix 5, Table A5.4, pp. 401-419, and Chapter 10, pp. 245-266, using Appendix 5, Table A5.7, pp. 426-430).

5.3.1 Raw data results.

5.3.1.1 Vowel series.

5.3.1.1.1 Series 1 (/iː/ vs. /ɪ/).

Table 2

Same-Different Rates for Vowel Stimuli Series 1 (/iː/ vs. /ɪ/)

Original data					
	Responses			Reaction time (*sec.*)	
Stimulus	'*Different*'	'*Same*'	*p(H)*	*M*	*SD*
1_8	97	3	.97	5.05	3.99
8_1	99	1	.99	5.31	3.67
2_7	95	5	.95	5.27	3.40
7_2	96	4	.96	4.79	3.60
3_6	83	17	.83	6.46	4.38
6_3	87	13	.87	5.36	3.99
4_5	25	75	.25	5.47	3.60
5_4	23	77	.23	6.52	4.22
4_4	8	92	.92	5.15	3.15

Note. N = 52. '*Different*' and '*Same*' values are mean percentages. Percentages are rounded to second decimal. *p(H)* is hit rate. ∑ Trials per stimulus = 10. Sound 1 corresponds to /iː/ and sound 8 to /ɪ/.

5.3.1.1.2 Series 2 (/uː/ vs. /ʊ/).

Table 3

Same-Different Rates for Vowel Stimuli Series 2 (/uː/ vs. /ʊ/)

Original data					
	Responses			Reaction time (*sec.*)	
Stimulus	‘*Different*’	‘*Same*’	*p(H)*	*M*	*SD*
9_12	73	27	.73	4.64	3.21
12_ 9	68	32	.68	4.57	3.50
10_11	39	61	.39	5.12	4.15
11_10	37	63	.37	5.13	3.79
11_11	18	82	.82	4.60	3.29

Note. N = 52[36]. ‘*Different*’ and ‘*Same*’ values are mean percentages. Percentages are rounded to second decimal. *p(H)* is hit rate. ∑ Trials per stimulus = 10. Sound 9 corresponds to /uː/ and sound 12 to /ʊ/.

5.3.1.1.3 Series 3 (/ɔː/ vs. /ɒ/).

Table 4

Same-Different Rates for Vowel Stimuli Series 3 (/ɔː/ vs. /ɒ/)

Original data					
	Responses			Reaction time (*sec.*)	
Stimulus	‘*Different*’	‘*Same*’	*p(H)*	*M*	*SD*
13_17	98	2	.98	3.96	2.77
17_13	98	2	.98	3.98	2.63
14_15	28	72	.28	4.42	2.96
15_14	28	72	.28	4.93	3.72
14_16	89	11	.89	5.11	3.53
16_14	87	13	.87	5.03	3.79
15_16	73	27	.73	3.59	2.86
16_15	64	36	.64	3.91	2.72
14_14	10	90	.90	5.25	3.52

Note. N = 52. ‘*Different*’ and ‘*Same*’ values are mean percentages. Percentages are rounded to second decimal. *p(H)* is hit rate. ∑ Trials per stimulus = 10. Sound 13 corresponds to /ɔː/ and sound 17 to /ɒ/.

---

[36] It is to note that one of the participants in the present research missed the first experiment.

5.3.1.1.4 Series 4 (/aː/ vs. /ʌ/).

Table 5

Same-Different Rates for Vowel Stimuli Series 4 (/aː/ vs. /ʌ/)

Original data					
	Responses			Reaction time (*sec.*)	
Stimulus	'*Different*'	'*Same*'	*p(H)*	*M*	*SD*
18_22	45	55	.45	5.32	3.93
22_18	45	55	.45	5.51	4.27
19_20	29	71	.29	5.26	3.22
20_19	28	73	.28	5.33	3.52
19_21	29	71	.29	5.66	4.69
21_19	32	68	.32	5.19	3.23
20_21	31	69	.31	6.44	4.73
21_20	33	67	.33	6.07	4.53
21_21	32	68	.68	6.13	4.18

Note. N = 52. '*Different*' and '*Same*' values are mean percentages. Percentages are rounded to second decimal. *P(H)* is hit rate. ∑ Trials per stimulus = 10. Sound 18 corresponds to /ɑː/ and sound 22 to /ʌ/.

5.3.1.1.5 Series 5 (/aː/ vs. /æ/).

Table 6

Same-Different Rates for Vowel Stimuli Series 5 (/aː/ vs. /æ/)

Original data					
	Responses			Reaction time (*sec.*)	
Stimulus	'*Different*'	'*Same*'	*p(H)*	*M*	*SD*
23_29	98	2	.98	4.57	3.52
29_23	97	3	.97	4.48	3.07
24_28	95	5	.95	4.23	3.39
28_24	96	4	.96	4.83	4.11
25_26	38	62	.38	4.47	3.08
26_25	38	62	.38	3.95	2.66
26_27	34	66	.34	4.94	3.34
27_26	30	70	.30	4.24	2.61
27_25	70	30	.70	4.17	3.31
25_27	70	30	.70	4.48	3.04
25_25	14	86	.86	4.50	2.94

Note. N = 52. '*Different*' and '*Same*' values are mean percentages. Percentages are rounded to second decimal. *P(H)* is hit rate. ∑ Trials per stimulus = 10. Sound 23 corresponds to /ɑː/ and sound 29 to /æ/.

5.3.1.1.6 Series 6 (/æ/ vs. /ʌ/).

Table 7

Same-Different Rates for Vowel Stimuli Series 6 (/æ/ vs. /ʌ/)

Original data					
	Responses			Reaction time (*sec.*)	
Stimulus	'*Different*'	'*Same*'	*p(H)*	*M*	*SD*
30_35	87	13	.87	4.99	3.94
35_30	89	14	.86	4.90	3.56
31_34	78	22	.78	4.30	2.59
34_31	72	28	.72	4.30	2.83
32_33	34	66	.34	6.33	4.11
33_32	36	64	.36	4.94	3.01
33_33	20	80	.80	4.53	3.27

Note. N = 52. '*Different*' and '*Same*' values are mean percentages. Percentages are rounded to second decimal. *p(H)* is hit rate. ∑ Trials per stimulus = 10. Sound 30 corresponds to /æ/ and sound 12 to /ʌ/.

5.3.1.1.7 Series 7 (/e/ vs. /ɜː/).

Table 8

Same-Different Rates for Vowel Stimuli Series 7 (/e/ vs. /ɜː/)

Original data					
	Responses			Reaction time (*sec.*)	
Stimulus	'*Different*'	'*Same*'	*p(H)*	*M*	*SD*
36_40	77	23	.77	5.25	4.40
40_36	75	25	.75	5.34	3.57
37_38	31	69	.31	5.64	4.02
38_37	34	66	.34	5.79	3.44
37_39	51	49	.51	5.28	3.72
39_37	49	51	.49	5.24	3.65
38_39	32	68	.32	5.42	4.30
39_38	33	67	.33	4.81	3.17
39_39	54	46	.54	5.43	3.29

Note. N = 52. '*Different*' and '*Same*' values are mean percentages. Percentages are rounded to second decimal. *P(H)* is hit rate. ∑ Trials per stimulus = 10. Sound 36 corresponds to /e/ and sound 40 to /ɜː/.

5.3.1.1.8 Series 8 (/ɜː/ vs. /ʊ/).

Table 9

Same-Different Rates for Vowel Stimuli Series 8 (/ɜː/ vs. /ʊ/)

Original data					
	Responses			Reaction time (*sec.*)	
Stimulus	‘*Different*’	‘*Same*’	*p(H)*	*M*	*SD*
41_45	98	2	.98	3.60	2.67
45_41	99	1	.99	4.38	2.99
42_43	83	17	.83	4.77	3.70
43_42	89	11	.89	3.95	2.78
42_44	93	7	.93	4.86	3.29
44_42	93	7	.93	4.01	2.61
43_44	13	87	.13	4.72	3.04
44_43	12	88	.12	5.69	4.01
44_44	7	93	.93	3.87	2.42

Note. N = 52. ‘*Different*’ and ‘*Same*’ values are mean percentages. Percentages are rounded to second decimal. *P*(H) is hit rate. ∑ Trials per stimulus = 10. Sound 41 corresponds to /ɜː/ and sound 45 to /ʊ/.

5.3.1.1.9 Series 9 (/ɜː/ vs. /uː/).

Table 10

Same-Different Rates for Vowel Stimuli Series 9 (/ɜː/ vs. /uː/)

Original data					
	Responses			Reaction time (*sec.*)	
Stimulus	‘*Different*’	‘*Same*’	*p(H)*	*M*	*SD*
46_50	98	2	.98	4.26	3.51
50_46	98	2	.98	4.35	3.38
47_48	91	9	.91	4.56	3.23
48_47	94	6	.94	4.46	3.40
47_49	93	7	.93	4.63	2.86
49_47	97	3	.97	5.47	3.52
48_49	12	88	.12	4.54	3.64
49_48	12	88	.12	4.88	3.57
48_48	7	93	.93	4.72	3.06

Note. N = 52. ‘*Different*’ and ‘*Same*’ values are mean percentages. Percentages are rounded to second decimal. *P(H)* is hit rate. ∑ Trials per stimulus = 10. Sound 46 corresponds to /ɜː/ and sound 50 to /uː/.

Tables 2 through 10 display descriptive statistics for obtained data for same different vowel discrimination test before combining stimulus pairs’ results, including sample mean hit rate, mean false alarm, and mean reaction time. The mean percent score of perceptual discrimination ability (*hit rate*) for participants ranged from 12% to 98% across all vowel pairs. Participants were strongly inclined towards responding '*Different*' across several vowel pairs, except for Series 4 where discrimination level did not reach chance level, i.e. 50%. False alarm rates of participants ranged from 7% to 32% across many vowel pairs, except for Series 7, where the false alarm rate was 54%.

Mean reaction times varied between 3 and 7 seconds for all series. The reported mean reaction time of 6.46 seconds in Series 1 is reasonably understandable, for the participants came to take the test for the very first time and they were probably focusing on making sure their decisions contained correct responses. Again, the standard deviation of mean reaction times for pairs in Series 4 stands quite differently from the other vowel series, suggesting probable difficulty encountered by participants to signal presence of any difference present in stimulus pairs. The highest mean score of discrimination (hit rate) in Series 4 adds support to this explanation ($p(H) = .45$).

5.3.1.2 Order presentation effect.

To make sure there were neither order presentation effect nor a carry-over effect that might have resulted in improving perceptual discrimination ability and reduced reaction time, a series of paired samples *t*-test was performed on participants' mean hit rates and reaction times.

5.3.1.2.1 Hit rate.

Table 11

Paired Samples t-Test on the Effect of Presentation Order on Mean Hit Rate

	Paired differences (hit rate)			
Pair	*M*	*SD*	*t*	*Sig. (2-tailed)*
1_8 & 8_1	-.02	.07	-2.11	.040
2_7 & 7_2	-.01	.08	-0.68	.498
3_6 & 6_3	-.03	.16	-1.56	.124
4_5 & 5_4	.03	.19	0.93	.357
9_12 & 12_9	.05	.18	2.00	.050
10_11 & 11_10	.02	.20	0.70	.485
13_17 & 17_13	.01	.05	0.77	.444
14_16 & 16_14	.02	.15	1.13	.265
14_15 & 15_14	.00	.22	-0.06	.949
15_16 & 16_15	.08*	.23	2.63	.011
18_22 & 22_18	.00	.24	-0.06	.955
19_20 & 20_19	.02	.17	0.73	.466
19_21 & 21_19	-.03	.23	-0.78	.438
20_21 & 21_20	.02	.16	0.93	.359
23_29 & 29_23	.01	.06	0.97	.334
24_28 & 28_24	-.01	.11	-0.56	.581
26_25 & 25_26	.00	.23	-0.05	.958
25_27 & 27_25	-.01	.24	-0.16	.874
26_27 & 27_26	.04	.24	1.25	.219
30_35 & 35_30	.01	.14	0.51	.613
31_34 & 34_31	.06*	.18	2.57	.013
32_33 & 33_32	-.02	.20	-0.75	.458
36_40 & 40_36	.02	.15	1.09	.282
38_37 & 37_38	.03	.21	1.04	.304
37_39 & 39_37	.02	.23	0.77	.443
38_39 & 39_38	-.01	.22	-0.19	.848

Note. N = 52. Confidence interval (CI) = 95%. Mean differences with an asterisk are significant results. Values are rounded to second decimals, except for 2-tailed probability. *df* = 51.

Table 11 (*continued*)

Paired Samples t-Test on the Effect of Presentation Order on Mean Hit Rate

	Paired differences (hit rate)			
Pair	*M*	*SD*	*t*	*Sig. (2-tailed)*
41_45 & 45_41	-.01	.05	-1.40	.168
42_43 & 43_42	-.06*	.19	-2.44	.018
42_44 & 44_42	.00	.10	-0.27	.788
43_44 & 44_43	.02	.14	0.90	.371
46_50 & 50_46	.01	.05	0.77	.444
47_49 & 49_47	-.03	.11	-2.18	.034
47_48 & 48_47	-.03	.15	-1.42	.160
49_48 & 48_49	.00	.11	0.00	1.000

Note. N = 52. Confidence interval (CI) = 95%. Mean differences with an asterisk are significant results. Values are rounded to second decimals, except for 2-tailed probability. *df* = 51.

Results in Table 11 show no significant order presentation effect on mean hit rates overall, except for three pairs (31_34 & 34_31, 15_16 & 16_15, 42_43 & 43_42) across all series. These significant differences in mean hit rates are likely a carry-over effect or participants' increased mobilisation of selective attention to stimuli.

5.3.1.2.2 Reaction time.

Table 12

Paired Samples t-Test on the Effect of Presentation Order on Mean Reaction Time

	Paired differences (reaction time in seconds)			
Pair	*M*	*SD*	*t*	*Sig. (2-tailed)*
1_8 & 8_1	-0.26	5.36	-0.35	.726
2_7 & 7_2	0.48	4.81	0.72	.475
3_6 & 6_3	1.10	4.86	1.64	.108
4_5 & 5_4	-1.05	5.12	-1.47	.147
9_12 & 12_9	0.07	4.55	0.12	.908
10_11 & 11_10	-0.01	5.87	-0.01	.990
13_17 & 17_13	-0.01	3.91	-0.03	.978
14_16 & 16_14	0.07	4.73	0.11	.912
14_15 & 15_14	-0.51	4.84	-0.76	.448
15_16 & 16_15	-0.31	3.71	-0.61	.544
18_22 & 22_18	-0.19	5.51	-0.25	.807
19_20 & 20_19	-0.07	4.41	-0.11	.914
19_21 & 21_19	0.47	5.77	0.58	.563
20_21 & 21_20	0.37	5.74	0.46	.645
23_29 & 29_23	0.09	4.84	0.13	.894
24_28 & 28_24	-0.60	6.00	-0.72	.476
26_25 & 25_26	-0.52	4.28	-0.87	.387
25_27 & 27_25	0.32	4.25	0.54	.590
26_27 & 27_26	0.70	4.36	1.16	.253
30_35 & 35_30	0.09	5.45	0.11	.910
31_34 & 34_31	0.00	3.71	0.01	.994
32_33 & 33_32	1.39	5.73	1.75	.085
36_40 & 40_36	-0.09	6.19	-0.11	.916
38_37 & 37_38	0.15	5.16	0.21	.836
37_39 & 39_37	0.04	5.31	0.06	.956
38_39 & 39_38	0.61	5.50	0.81	.424

Note. N = 52. Confidence interval (CI) = 95%. Mean differences with an asterisk are significant results. Values are rounded to second decimals, except for 2-tailed probability. *df* = 51.

Table 12 (*continued*)

Paired Samples t-Test on the Effect of Presentation Order on Mean Reaction Time

	Paired differences (reaction time in seconds)			
Pair	*M*	*SD*	*t*	*Sig. (2-tailed)*
41_45 & 45_41	-0.78	4.23	-1.33	.188
42_43 & 43_42	0.82	4.34	1.36	.180
42_44 & 44_42	0.85	3.75	1.63	.108
43_44 & 44_43	-0.97	5.31	-1.32	.193
46_50 & 50_46	-0.08	4.81	-0.13	.899
47_49 & 49_47	-0.72	4.58	-1.13	.264
47_48 & 48_47	0.10	5.07	0.15	.882
49_48 & 48_49	0.34	5.10	0.48	.636

Note. N = 52. Confidence interval (CI) = 95%. Mean differences with an asterisk are significant results. Values are rounded to second decimals, except for 2-tailed probability. *df* = 51.

In a similar fashion to previously reported findings, pairwise comparisons of order presentation effect on mean reaction time showed no significant differences across all presented pairs of stimuli, indicating likely participants' use of a uniform strategy in making a decision regardless of the presented vowel pair. This raises an interesting question of whether native speakers of RP English would show a similar or a different pattern of responses with respect to discrimination and reaction time.

5.3.2 Modified data results.

5.3.2.1 Sensitivity measurements.

Tables 13 through 21 display descriptive statistics for obtained data for same-different vowel discrimination test, after combining stimulus pairs' results, including sample mean hit rate, mean false alarm, and mean reaction time. For our research purpose, we set a confirmed empirical threshold sensitivity at a hit rate above .50 to invoke the concept of categorical perception and .80 to invoke potentially the concept of automaticity.

5.3.2.1.1 Series 1 (/iː/ vs. /ɪ/).

Table 13

Sensitivity Measurements for Same-Different Test for Stimuli Series 1 (/iː/ vs. /ɪ/)

Modified data			
Combined stimulus pairs	*p(H)*	*p*(M)	*p*(FA)
1_8 & 8_1	.98	.02	.08
2_7 & 7_2	.96	.04	.08
3_6 & 6_3	.85	.15	.08
4_5 & 5_4	.24	.76	.08

Note. N = 52. *P(H)* is hit rate. *P*(M) is mean miss rate. *P*(FA) is mean false-alarm rate. Sound 1 corresponds to /iː/ and sound 8 to /ɪ/.

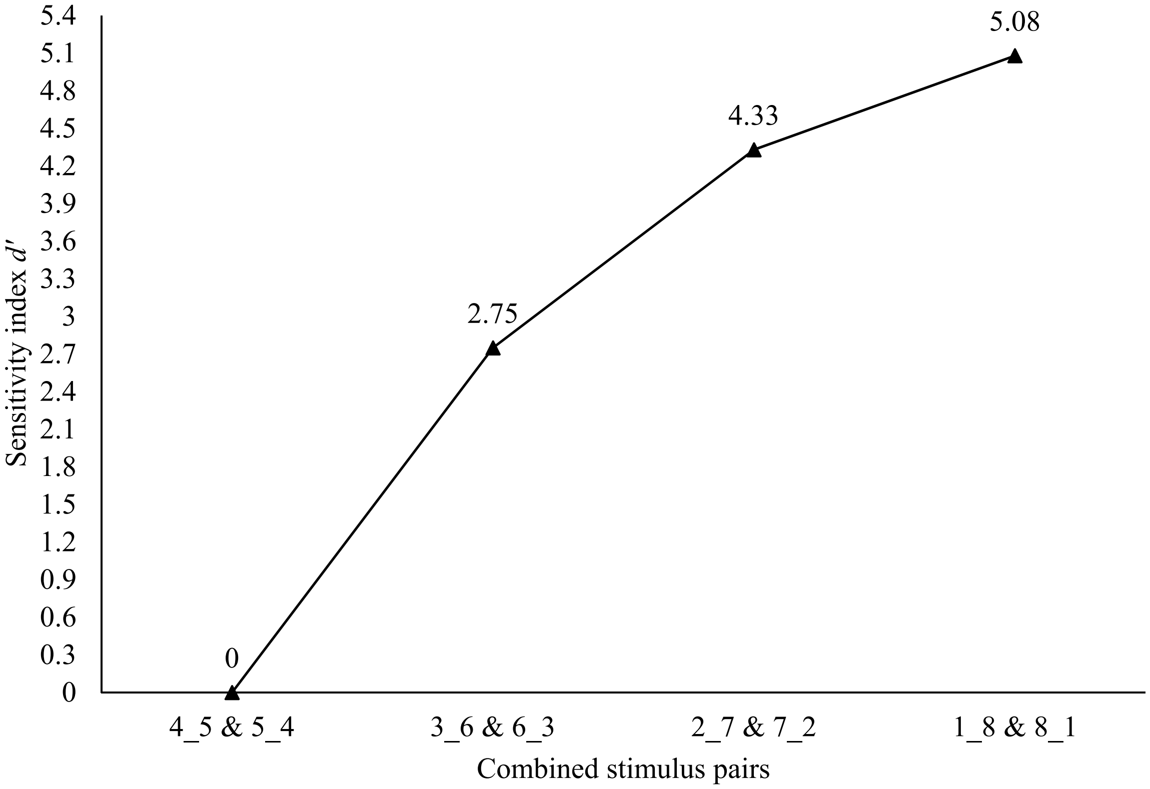


Figure 16. Discrimination of stimulus pairs across /iː/ vs. /ɪ/ spectral continuum. N = 52. *P*(FA) = .08. Sound 1 corresponds to /iː/ and sound 8 to /ɪ/.Sensitivity index *d'* is measured for combined stimulus pairs modified data, using differencing model.

Table 13 shows participants' remarkable abilities in discriminating differences in the presented vowel stimuli, except for the combined pairs 4_5 & 5_4. We find a similar perceptual capacity in their low false alarm rate, estimated at .08. For sensitivity measurements, *Figure* 16 show high significant pooled sensitivity indices across the continuum for the combined stimulus pairs across /iː/ vs. /ɪ/ spectral continuum, with pooled *d'* ranging from 2.75 to 5.08, except for the stimulus pairs 4_5 & 5_4 (*d'* = 0). This latter fact is likely to represent an empirical threshold of sensitivity to the pair /iː/ vs. /ɪ/. This result suggests the probable existence of a perceptual category boundary for /iː/ and /ɪ/ prototypes, likely delimited by sounds 3 and 6, and out of the perceptual space between sound 4 and 5. The latter sounds are discriminatorily imperceptible, indicating empirically a probable existence of two distinct phonetic vowel categories serving as prototypes, perceptually discriminable above sound 4 (F_1 339 Hz, F_2 2069 Hz) towards /iː/ or below sound 5 (F_1 353 Hz, F_2 2011 Hz) towards /ɪ/ on the perceptual map. The empirical threshold of sensitivity to the pair /iː/ vs. /ɪ/ is difficult to determine here, indicating a larger perceptual space to be occupied by either /iː/ or /ɪ/ as described before. That is, if the perceptual category boundary, as defined by a mean discrimination hit rate above .50 or chance level, lies between sounds 3 and 4, then the vowel /ɪ/ may be said to occupy a larger perceptual than /iː/. Similarly, if the perceptual category boundary lies between sounds 5 and 6, then the vowel /iː/ may be said to occupy a larger perceptual space than /ɪ/.

5.3.2.1.2 Series 2 (/uː/ vs. /ʊ/).

Table 14

Sensitivity Measurements for Same-Different Test for Stimuli Series 2 (/uː/ vs. /ʊ/)

Modified data			
Combined stimulus pairs	*p(H)*	*p*(M)	*p*(FA)
9_12 & 12_9	.71	.29	.18
10_11 & 11_10	.38	.62	.18

Note. N = 52. *P(H)* is hit rate. *P*(M) is mean miss rate. *P*(FA) is mean false-alarm rate. Sound 9 corresponds to /uː/ and sound 12 to /ʊ/.

Compared to Series 1, results in Table 14 display less notable abilities in discriminating the presented stimuli, with less mean hit rates and a slightly important false-alarm rate of .18. With respect to sensitivity measurements, *Figure* 17 shows relatively less important pooled sensitivity indices for combined stimulus pairs across /uː/ vs. /ʊ/ spectral distance continuum, with pooled *d'* ranging from 0 to 1.87, suggesting limited discrimination ability and less perceptual category boundary of these two RP English prototypic vowels. The boundary line delimiting categorical perception lies empirically above sound 10 (F_1 389 Hz, F_2 1527 Hz) or below sound 11 (F_1 392 Hz, F_2 1467 Hz). The position of the perceptual category boundary will determine both the perceptual space occupied by each prototypic vowel and the empirical threshold of sensitivity to the pair /uː/vs. /ʊ/. In other words, if the perceptual category boundary lies between sounds 9 and 10, then the vowel /ʊ/ may be said to occupy a larger perceptual space than /uː/. Likewise, if the perceptual category boundary lies between sounds 11 and 12, then the vowel /uː/ may be said to occupy a larger perceptual space than /ʊ/.

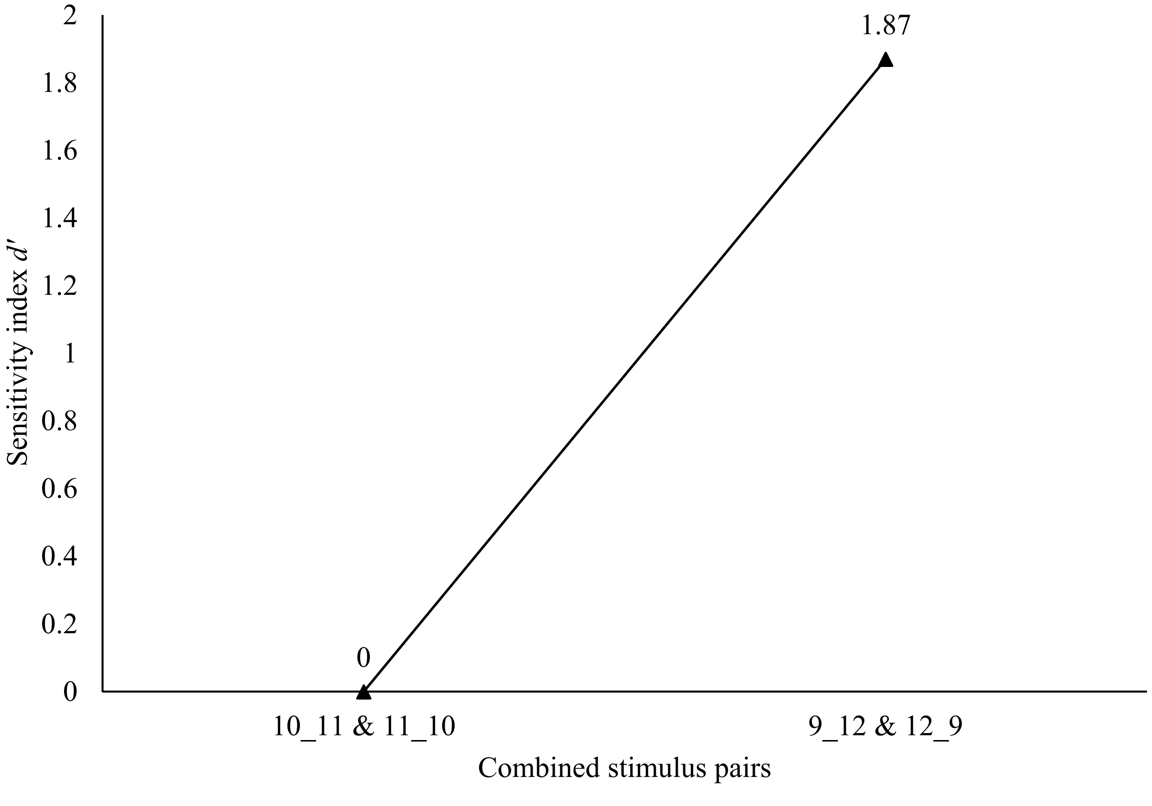


Figure 17. Discrimination of stimulus pairs across /uː/ vs. /ʊ/ spectral continuum. N = 52. *P*(FA) = .18. Sound 9 corresponds to /uː/ and sound 12 to /ʊ/.Sensitivity index *d'* is measured for combined stimulus pairs modified data, using differencing model.

5.3.2.1.3 Series 3 (/ɔː/ vs. /ɒ/).

Table 15

Sensitivity Measurements for Same-Different Test for Stimuli Series 3 (/ɔː/ vs. /ɒ/)

Modified data			
Combined stimulus pairs	*p(H)*	*p*(M)	*p*(FA)
13_17 & 17_13	.98	.02	.08
14_15 & 15_14	.28	.72	.08
15_16 & 16_15	.68	.32	.08
14_16 & 16_14	.88	.12	.08

Note. N = 52. *P(H)* is hit rate. *P*(M) is mean miss rate. *P*(FA) is mean false-alarm rate. Sound 13 corresponds to /ɔː/ and sound 17 to /ɒ/.

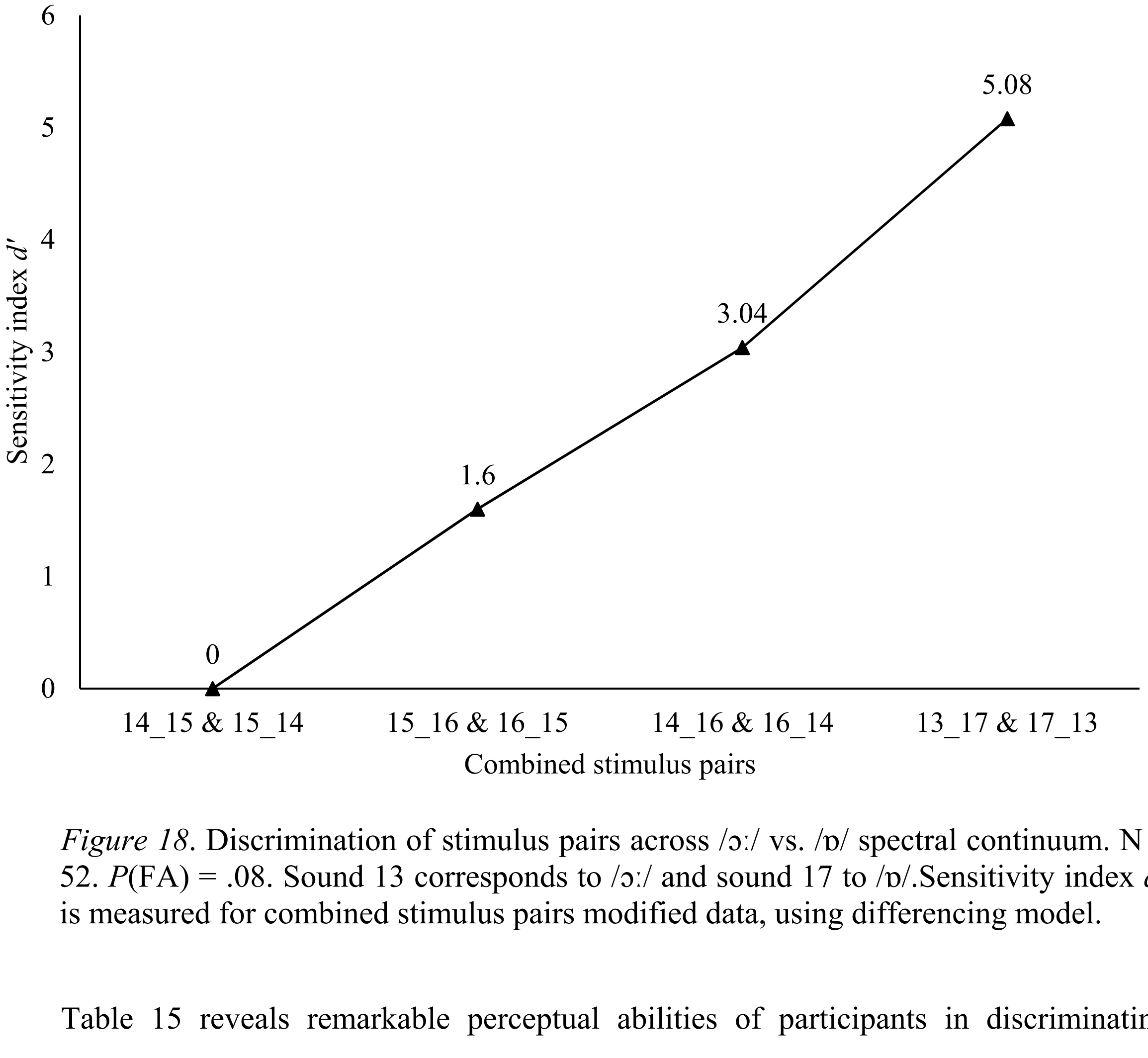


Figure 18. Discrimination of stimulus pairs across /ɔː/ vs. /ɒ/ spectral continuum. N = 52. *P*(FA) = .08. Sound 13 corresponds to /ɔː/ and sound 17 to /ɒ/.Sensitivity index *d'* is measured for combined stimulus pairs modified data, using differencing model.

Table 15 reveals remarkable perceptual abilities of participants in discriminating differences in the presented stimuli and a low false-alarm rate of .08. Sensitivity measurements in *Figure* 18 show interesting results regarding participants' perceptual abilities of discrimination for combined stimulus pairs across /ɔː/ vs. /ɒ/ spectral distance continuum. Pooled sensitivity indices are significantly high for the two combined stimulus pairs 13_17 & 17_13 and 14_16 & 16_14 (pooled sensitivity index *d'* ranging from 1.6 to 5.08), suggesting a high level of categorical perception. Nonetheless, the spectral distance of about 40 Hz in both F_1 and F_2 near the vowel /ɒ/ for the combined stimulus pair 15_16 & 16_15 is significantly more perceptible than an equal spectral distance near /ɔː/ for the combined pair 14_15 & 15_14. This fact suggests the existence of a demarcating perceptual category boundary for the two distinct phonetic vowel categories lying between sound 15 (F_1 562 Hz, F_2 938 Hz) and sound 16 (F_1 602 Hz, F_2 979 Hz), with the vowel /ɔː/ occupying a larger perceptual space on the perceptual map towards /ɒ/. For the pair /ɔː/ vs. /ɒ/, empirical threshold of sensitivity lies considerably below a spectral distance of 41 Hz in both F_1 and F_2 towards /ɒ/, and very probably above 41 Hz in both F_1 and F_2 towards /ɔː/.

5.3.2.1.4 Series 4 (/ɑː/ vs. /ʌ/).

Table 16

Sensitivity Measurements for Same-Different Test for Stimuli Series 4 (/ɑː/ vs. /ʌ/)

Modified data			
Combined stimulus pairs	*p(H)*	*p*(M)	*p*(FA)
18_22 & 22_18	.45	.55	.32
19_20 & 20_19	.28	.72	.32
20_21 & 21_20	.32	.68	.32
21_19 & 19_21	.30	.70	.32

Note. N = 52. *P(H)* is hit rate. *P*(M) is mean miss rate. *P*(FA) is mean false-alarm rate. Sound 18 corresponds to /ɑː/ and sound 22 to /ʌ/.

Table 16 reveals a completely different pattern of discrimination abilities for Series 4 compared to the three previous series, with low hit rates, not exceeding .45, and a high false-alarm rate of .32. Equally, sensitivity measurements in *Figure* 19 demonstrate that the pair/ɑː/ vs. /ʌ/ seems to be significantly difficult to discriminate among all examined vowel pairs in the experiment. Discrimination hit rates for combined stimulus pairs across /ɑː/ vs. /ʌ/ spectral continuum fall far below our defined empirical threshold of sensitivity that is indicative of potential categorical perception (*P(H)* = .50) or chance level, with all pooled sensitivity indices being equal to 0. This finding suggests a significant perceptual difficulty to discriminate these two vowels at least at the spectral level alone and provides empirical support for the absence of sensitivity to vowels in this area on the perceptual map. Accordingly, a spectral distance of 21 Hz in F_1 and 103 Hz in F_2 in the area of /ɑː/ vs. /ʌ/ prototypic vowels is perceptibly indiscriminable, posing empirical difficulty to invoke the concept of categorical perception. Empirical threshold of sensitivity to the pair /ɑː/ vs. /ʌ/ lies considerably above a spectral distance of 19 Hz in F_1 and 103 Hz in F_2.

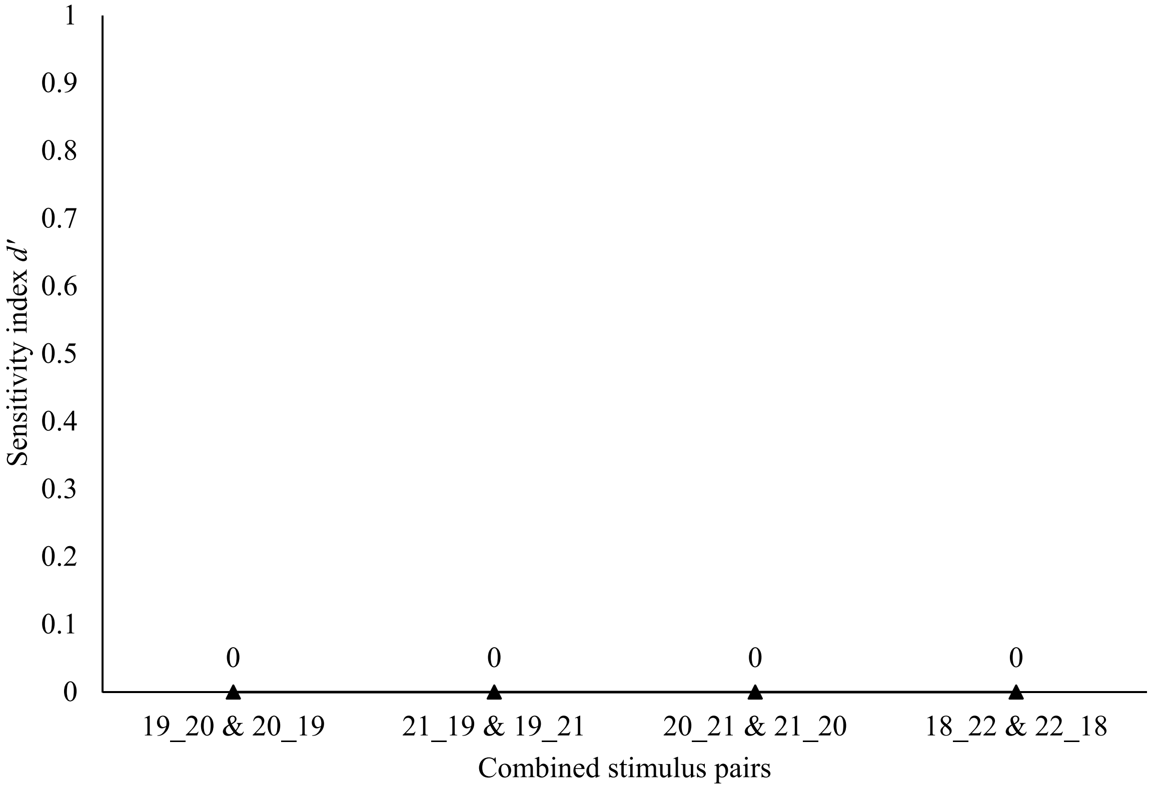


Figure 19. Discrimination of stimulus pairs across /ɑː/ vs. /ʌ/ spectral continuum. N = 52. *P*(FA) = .32. Sound 18 corresponds to /ɑː/ and sound 22 to /ʌ/.Sensitivity index *d'* is measured for combined stimulus pairs modified data, using differencing model.

5.3.2.1.5 Series 5 (/aː/ vs. /æ/).

Table 17

Sensitivity Measurements for Same-Different Test for Stimuli Series 5 (/aː/ vs. /æ/)

Modified data			
Combined stimulus pairs	*p(H)*	*p*(M)	*p*(FA)
23_29 & 29_23	.97	.03	.14
24_28 & 28_24	.96	.04	.14
25_27 & 27_25	.70	.30	.14
25_26 & 26_25	.38	.62	.14
26_27 & 27_26	.32	.68	.14

Note. N = 52. *P(H)* is hit rate. *P*(M) is mean miss rate. *P*(FA) is false-alarm rate. Sound 23 corresponds to /ɑː/ and sound 29 to /æ/.

Table 17 displays important discrimination hit rates for vowel pairs near the prototypic ones and a slightly important false-alarm rate of .14. As shown below in *Figure 20*, pooled sensitivity indices for combined stimulus pairs across /ɑː/ vs. /æ/ spectral continuum are notably higher, with *d'* ranging from 1.72 to 4.63 for the three combined stimulus pairs (23_29 & 29_23, 24_28 & 28_24, and 25_27 & 27_25), and a null sensitivity index for the remaining ones. Compared to the vowel pair /ɔː/ vs. /ɒ/ just above in Series 4, a spectral distance of 2 Hz in F_1 and 62 Hz in F_2 between /ɑː/ vs. /æ/ yielded similar sensitivity indices for both pairs 25_26 & 26_25 and 26_27 & 27_26. This empirical fact suggests the existence of a perceptual category boundary for the two vowel categories between sound 25 (F_1 676 Hz, F_2 1317 Hz) and sound 27 (F_1 671 Hz, F_2 1441 Hz), indicating probably an equal perceptual space occupied by both prototypic vowels. Empirical threshold of sensitivity to the pair /æ/ vs. /ʌ/ lies considerably above a spectral distance of 2 Hz in F_1 and 62 Hz in F_2 delimited by sounds 25 and 27.

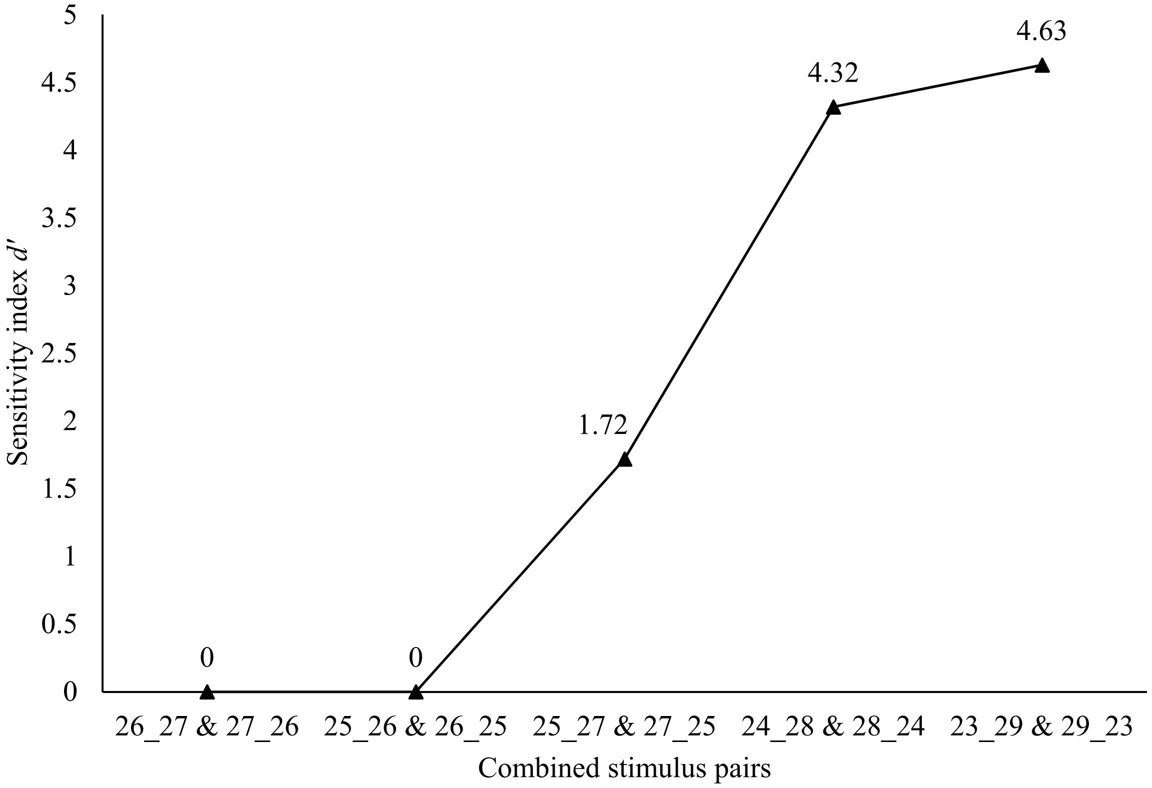


Figure 20. Discrimination of stimulus pairs across /ɑː/ vs. /æ/ spectral continuum. N = 52. *P*(FA) = .14. Sound 23 corresponds to /ɑː/ and sound 29 to /æ/.Sensitivity index *d'* is measured for combined stimulus pairs modified data, using differencing model.

5.3.2.1.6 Series 6 (/œ/ vs. /ʌ/).

Table 18

Sensitivity Measurements for Same-Different Test for Stimuli Series 6 (/œ/ vs. /ʌ/)

Modified data			
Combined stimulus pairs	*p(H)*	*p*(M)	*p*(FA)
30_35 & 35_30	.86	.14	.20
31_34 & 34_31	.75	.25	.20
32_33 & 33_32	.35	.65	.20

Note. N = 52. *P(H)* is hit rate. *P*(M) is mean miss rate. *P*(FA) is mean false-alarm rate. Sound 30 corresponds to /æ/ and sound 12 to /ʌ/.

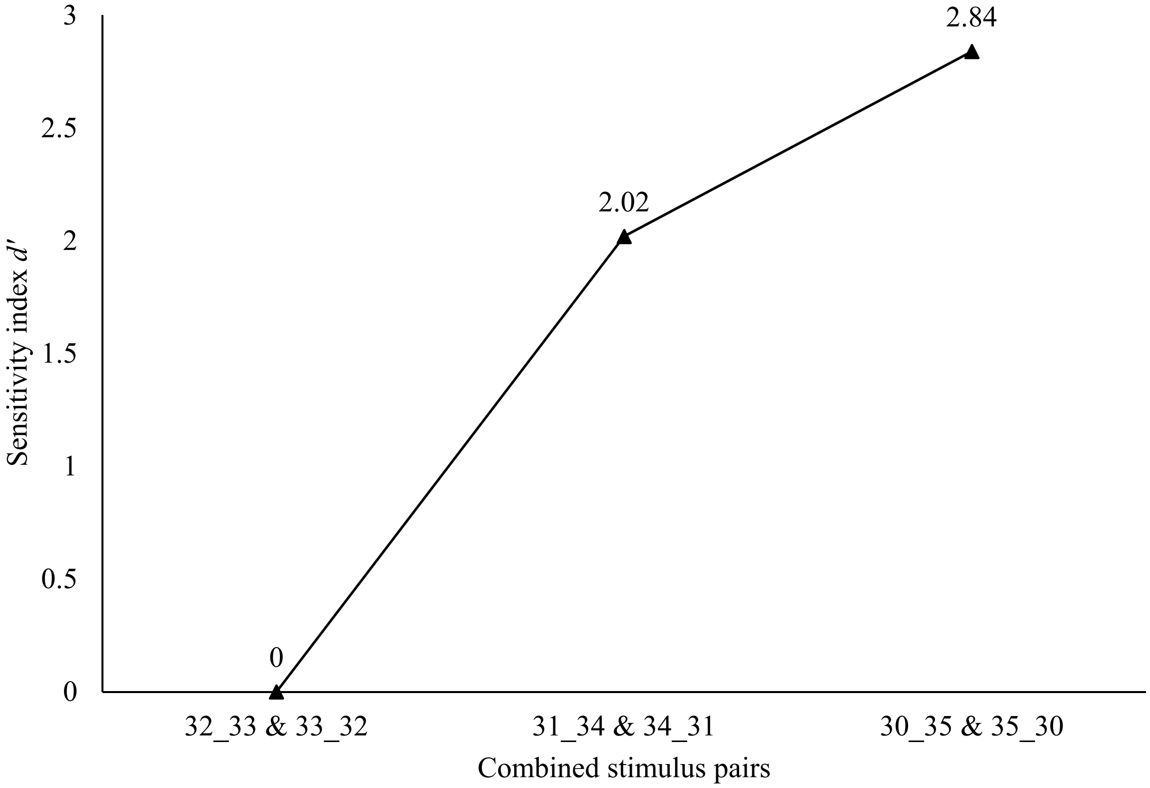


Figure 21. Discrimination of stimulus pairs across /æ/ vs. /ʌ/ spectral continuum. N = 52. *P*(FA) = .20. Sound 30 corresponds to /æ/ and sound 35 to /ʌ/.Sensitivity index *d'* is measured for combined stimulus pairs modified data, using differencing model.

Table 18 shows a similar pattern of discrimination across the prototypic vowels' spectral continuum to the previous results in Series 5, with also an important false-alarm rate of .20. *Figure 21* shows participants' substantial discrimination perceptual ability for combined stimulus pairs across /æ/ vs. /ʌ/ spectral continuum (practically less than the one reported for the pair /ɑː/ vs. /æ/), with pooled sensitivity indices *d'* ranging from 0 to 2.84. The spectral distance of 3 Hz in F_1 and 53 Hz in F_2, in the low area of the perceptual map between /æ/ vs. /ʌ/ prototypes, seems hard to discriminate, suggesting the existence of a categorical perception boundary for /æ/ vs. /ʌ/ delimited by the sounds 33 and 34 near /ʌ/ and the sounds 31 and 32 near /æ/. The assumed perceptual category boundary may indicate that either of the prototypic vowels /æ/ or /ʌ/ occupies a larger perceptual space from the other. That is, if the perceptual category boundary lies above sounds 33 and 34, then the vowel /æ/ will be said to occupy a larger perceptual space. However, if the perceptual category boundary falls between sounds 32 and 31, then the vowel /ʌ/ will be said be said to occupy a larger perceptual space. Based on this result, empirical threshold of sensitivity to this pair of prototypic vowels depends on the establishment of categorical perception boundary.

5.3.2.1.7 Series 7 (/e/ vs. /ɜː/).

Table 19

Sensitivity Measurements for Same-Different Test for Stimuli Series 7 (/e/ vs. /ɜː/)

Modified data			
Combined stimulus pairs	*p(H)*	*p*(M)	*p*(FA)
36_40 & 40_36	.76	.24	.54
37_39 & 39_37	.50	.50	.54
37_38 & 38_37	.33	.67	.54
38_39 & 39_38	.32	.68	.54

Note. N = 52. *P(H)* is hit rate. *P*(M) is miss rate. *P*(FA) is mean false-alarm rate. Sound 36 corresponds to /e/ and sound 40 to /ɜː/.

In addition to the most noticeable finding in Table 19 regarding the extremely high false-alarm rate of .54, there seems to be a well-delimited perceptual category boundary for the prototypic vowels /e/ and /ɜː/. *Figure* 22 shows significant pooled sensitivity indices for combined stimulus pair 36_40 & 40_36, with pooled *d'* of 2.08, while pooled sensitivity indices for other combined stimulus pairs were null, with a hit rate of .50 for combined stimulus pair 37_39 & 39_70 (*d'* = 0.15). Such a result suggests participants' loose sensitivity (below chance level) to this vowel pair, with an empirical threshold of sensitivity lying above a spectral distance of 7 Hz in F_1 and 124 Hz in F_2, that is, twice the set spectral distance for experimentation. Although the prototypic vowels /e/ and /ɜː/ were discriminated above chance level, they seem to occupy a small perceptual space similar to that occupied by /uː/ and /ʊ/ in Series 2.

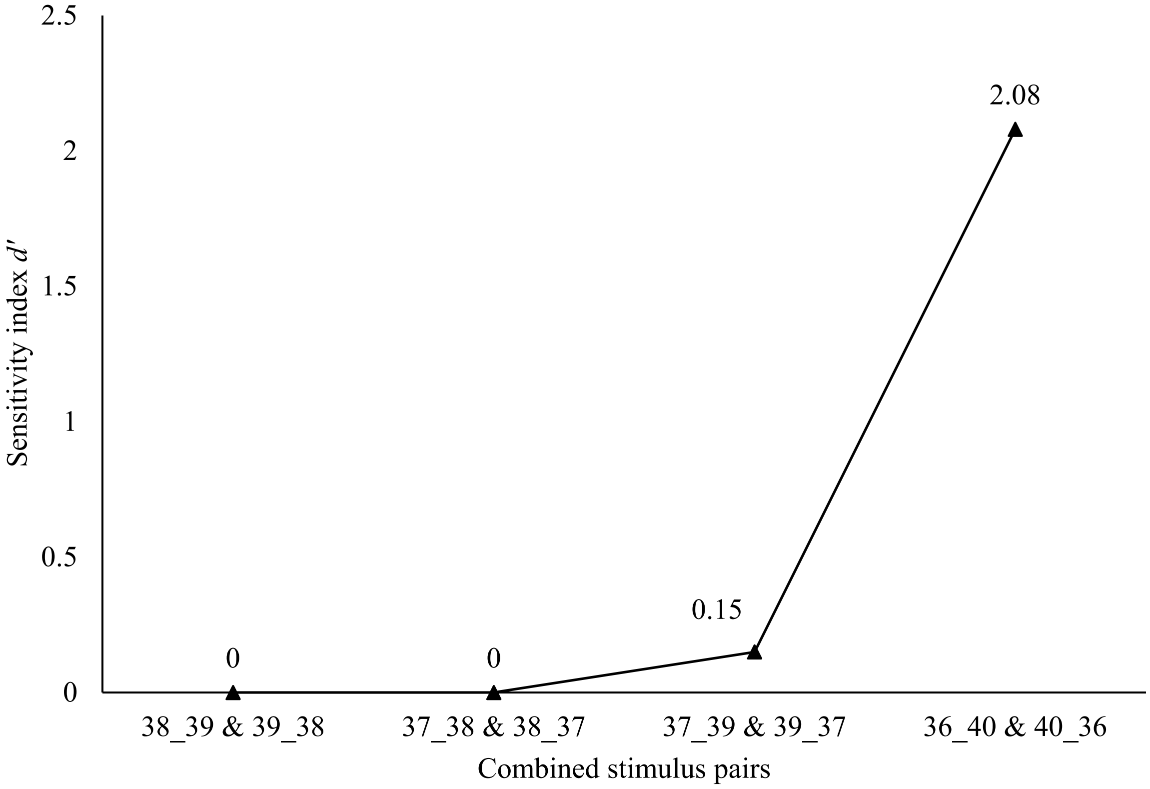


Figure 22. Discrimination of stimulus pairs across /e/ vs. /ɜː/ spectral continuum. N = 52. *P*(FA) = .54. Sound 36 corresponds to /e/ and sound 40 to /ɜː/.Sensitivity index *d'* is measured for combined stimulus pairs modified data, using differencing model.

5.3.2.1.8 Series 8 (/ɜː/ vs. /ʊ/).

Table 20

Sensitivity Measurements for Same-Different Test for Stimuli Series 8 (/ɜː/ vs. /ʊ/)

Modified data			
Combined stimulus pairs	*p(H)*	*p*(M)	*p*(FA)
41_45 & 45_41	.99	.01	.07
42_44 & 44_42	.93	.07	.07
42_43 & 43_42	.86	.14	.07
43_44 & 44_43	.13	.87	.07

Note. N = 52. *P(H)* is hit rate. *P*(M) is mean miss rate. *P*(FA) is mean false-alarm rate. Sound 41 corresponds to /ɜː/ and sound 45 to /ʊ/.

Table 20 above shows almost perfect discrimination perceptual abilities for combined stimulus pairs across the spectral continuum of the prototypic vowels /ɜː/ and /ʊ/ and the lowest false-alarm rate of .07 compared with the previous series. Pooled sensitivity measurements in

Figure 23 are significantly high, with pooled *d'* ranging from 2.84 to 5.76, suggesting a strongly attested discrimination perceptual ability among participants. Nonetheless, the discrimination perceptual ability between sound 42 (F_1 439 Hz, F_2 1408 Hz) and 43 (F_1 399 Hz, F_2 1408 Hz) suggests the existence of a boundary line for categorical perception for these prototypic vowels, with the vowel /ʊ/ likely to occupy a more perceptual space. For the pair /ɜː/ vs. /ʊ/, empirical threshold of sensitivity lies considerably above a spectral distance of 40 Hz in F_1 and 00 Hz in F_2 towards /ʊ/, and below 40 Hz in F_1 and 00 Hz in F_2 towards /ɜː/.

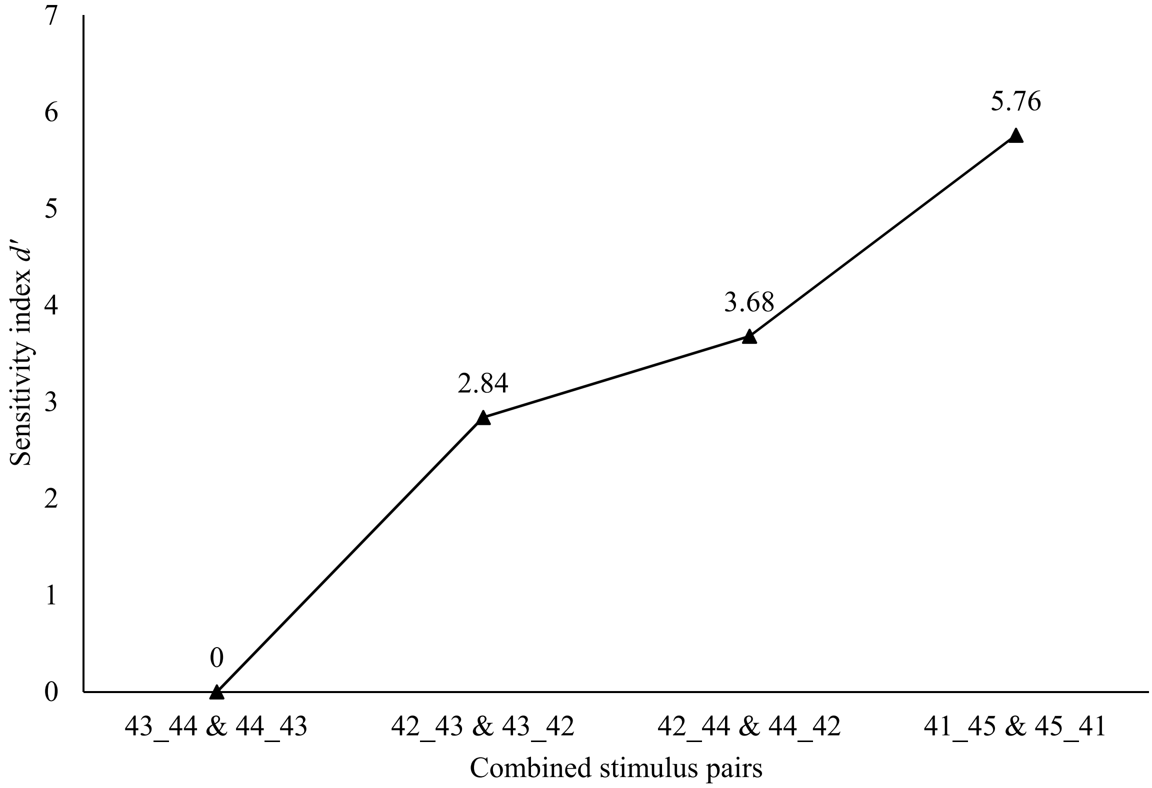


Figure 23. Discrimination of stimulus pairs across /ɜː/ vs. /ʊ/ spectral continuum. N = 52. *P*(FA) = .07. Sound 41 corresponds to /ɜː/ and sound 45 to /ʊ/.Sensitivity index *d'* is measured for combined stimulus pairs modified data, using differencing model.

5.3.2.1.9 Series 9 (/ɜː/ vs. /uː/).

Table 21

Sensitivity Measurements for Same-Different Test for Stimuli Series 9 (/ɜː/ vs. /uː/)

Modified data			
Combined stimulus pairs	*p(H)*	*p*(M)	*p*(FA)
46_50 & 50_46	.98	.02	.07
47_49 & 49_47	.95	.05	.07
47_48 & 48_47	.93	.07	.07
48_49 & 49_48	.12	.88	.07

Note. N = 52. *P(H)* is hit rate. *P*(M) is mean miss rate. *P*(FA) is mean false-alarm rate. Sound 46 corresponds to /ɜː/ and sound 50 to /uː/.

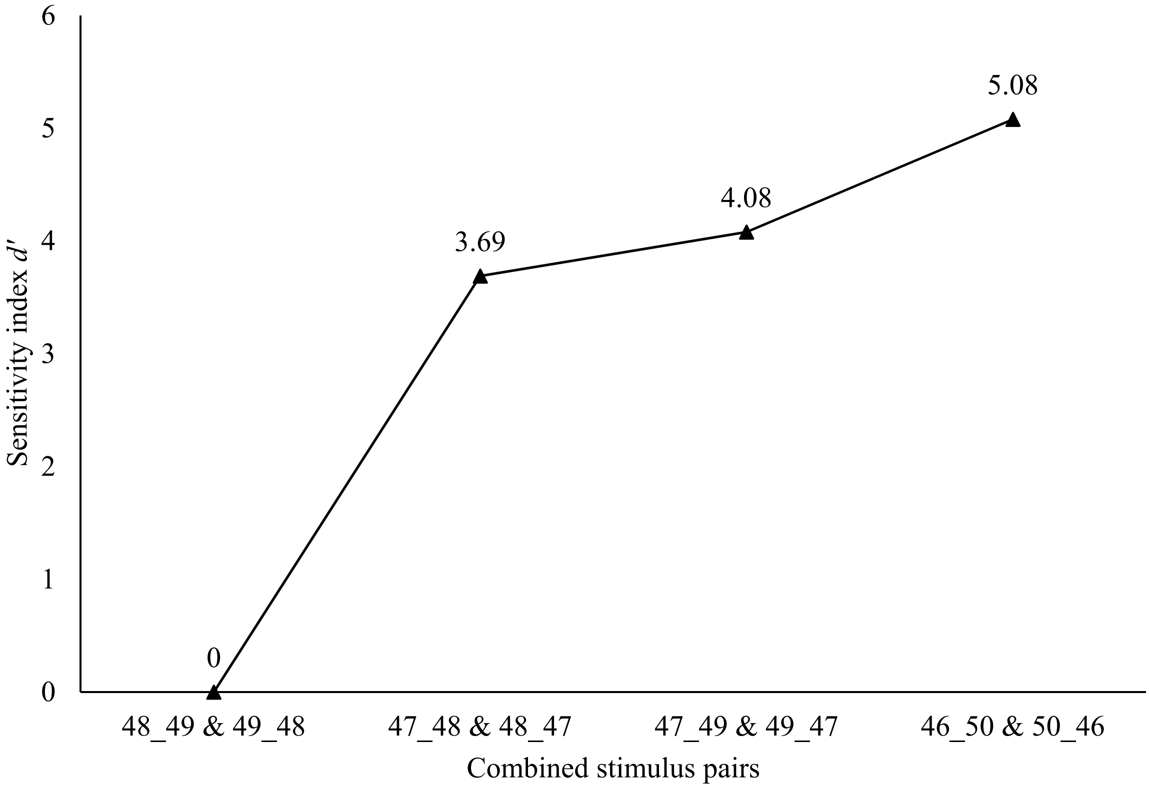


Figure 24. Discrimination of stimulus pairs across /ɜː/ vs. /uː/ spectral continuum. N = 52. *P*(FA) = .07. Sound 46 corresponds to /ɜː/ and sound 50 to /uː/.Sensitivity index *d'* is measured for combined stimulus pairs modified data, using differencing model.

Coming to the last vowel pair in this experiment, Table 21 shows perfect discrimination hit rates for three combined stimulus pairs and the lowest false alarm rate of .07 that is similar to that obtained for Series 8. *Figure* 24 shows high pooled sensitivity indices for three combined stimulus pairs, with *d’* ranging from 3.69 to 5.08, indicating high, significant perceptual ability. Interestingly, a spectral distance of 34 Hz in F_2 and 45 Hz in F_2 between /ɜː/ vs. /uː/ is significantly imperceptible near /uː/, and is largely and significantly perceptible near /ɜː/, suggesting, as in the previous vowel pair /ɜː/ vs. /ʊ/, that the prototypic vowel /uː/ may occupy a larger perceptual space compared to /ɜː/. Empirical threshold of sensitivity to the pair /ɜː/ vs. /uː/ lies considerably above 33 Hz in F_1 and 45 Hz in F_2 towards /uː/ and below 33 Hz in F_1 and 45 Hz in F_2 towards /ɜː/.

Overall, the obtained results suggest that most RP prototypic monophthongs are considerably discriminated beyond chance level (50%) and at varying discrimination levels across compared prototypes’ continua. This experientially based fact may be indicative of (a) a biologically inherent auditory capacity to discriminate these prototypic vowels based on spectral qualities alone, and (b) the presence of distinct phonetic categories representing these prototypic vowels due to participants’ experience with English language vowel system, being sufficiently exposed to English language input and phonetic instruction.

5.3.2.2 Variation in discrimination rate and reaction time.

In order to investigate potential significant effect of spectral distance on discrimination mean hit rate and reaction time across all series and for main pairs of prototypic vowels, we ran one-way repeated measures ANOVA on respective data. The rationale for this analysis was that mean discrimination hit rate and reaction time were dependent variables whose values would be a function of vowel token quality and spectral distance from prototypic vowel centres. For pairwise comparisons of means in ANOVA were calculated on estimated marginal means, we present first the latter, repeated measure ANOVA table second, and pairwise comparisons third. Again, as no homogeneous variances among dependent samples were found and sphericity assumption was repeatedly violated, the lower-bound correction was used with all ANOVA analyses. We adopted this very conservative and unconventionally recommended method to make sure generalisation of potential conclusions would be statistically valid in the worst probabilistic scenario (for detail on sphericity tests, see Appendix 6).

5.3.2.2.1 Discrimination rate.

A one-way ANOVA with repeated measures conducted on mean discrimination hit rates between vowel token pairs across all series revealed the following results.

5.3.2.2.1.1 Series 1 (/iː/ vs. /ɪ/).

Table 22

Mean Hit Rate for Combined Pairs Series 1

Combined pair	*M*	*SE*
1_8 & 8_1	.98	.01
2_7 & 7_2	.96	.01
3_6 & 6_3	.85	.02
4_5 & 5_4	.24	.03

Note. N = 52.

One-way repeated measures ANOVA with the lower-bound correction determined that mean discrimination hit rate differed statistically significantly between combined pairs of vowel tokens, $F(1, 51) = 308.56$, $p < .000$, and a large effect size $\eta^2 = .86$. Post hoc tests using the Bonferroni correction revealed that reduction in spectral distance elicited significant reduction in mean discrimination hit rate between all combined pairs, except for the combined pairs 1_8 & 8_1 and 2_7 & 7_2, which was statistically non-significant, *MD*[37] = .02, $p = .12$. The most noticeable difference was observed between the prototypic vowel pair 1_8 & 8_1 and the token pair 4_5 & 5_4, *MD* = .74, $p < .000$. Based on these empirical findings, we can conclude that reduction in spectral distance along the two first vowel formants (i.e. simultaneous reduction in F_1 and F_2) between prototypic vowels /iː/ and /ɪ/ elicits statistically a significant drop in perceptual discrimination ability. The present empirical finding suggests that participants could

---

[37] *MD* denotes mean difference

use their attentional resources to detect spectral differences in the presented vowel tokens and could employ them in a significant manner.

Table 23

Within-Subjects Effects for Mean Hit Rate for Combined Pairs Series 1

Source: Series 1	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	3	308.562	.000	.858	1.000
Greenhouse-Geisser	1.356	308.562	.000	.858	1.000
Huynh-Feldt	1.380	308.562	.000	.858	1.000
Lower-bound	1	308.562	.000	.858	1.000

[a]. Computed using alpha = .05

Table 24

Pairwise Comparisons of Mean Hit Rate for Combined Pairs Series 1

Combined pair (I)	Combined pair (J)	Mean difference (I – J)	*Sig.* [b]
1_8 & 8_1	2_7 & 7_2	.02	.122
	3_6 & 6_6	.13*	.000
	4_5 & 5_4	.74*	.000
2_7 & 7_2	3_6 & 6_3	.11*	.000
	4_5 & 5_4	.72*	.000
3_6 & 6_3	4_5 & 5_4	.61*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

5.3.2.2.1.2 Series 2 (/uː/ vs. /ʊ/).

Table 25

Mean Hit Rate for Combined Pairs Series 2

Combined pair	*M*	*SE*
9_12 & 12_9	.71	.04
10_11 & 11_10	.38	.03

Note. N = 52.

One-way repeated measures ANOVA with the lower-bound correction determined that mean discrimination hit rate differed statistically significantly between the two combined pairs of vowel tokens, $F(1, 51) = 115.792$, $p < .000$, and a large effect size $\eta^2 = .69$. Post hoc tests using the Bonferroni correction revealed again that two different spectral distances elicited significantly different mean discrimination hit rates between the combined pairs 9_12 & 12_9 and 10_11 & 11_10, *MD* = .33, $p < .000$, indicating a similar pattern of perceptual capacity observed in Series 1.

Table 26

Within-Subjects Effects for Mean Hit Rate for Combined Pairs Series 2

Source: Series 2	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	1	115.792	.000	.694	1.000
Greenhouse-Geisser	1.000	115.792	.000	.694	1.000
Huynh-Feldt	1.000	115.792	.000	.694	1.000
Lower-bound	1	115.792	.000	.694	1.000

[a] Computed using alpha = .05

Table 27

Pairwise Comparisons of Mean Hit Rate for Combined Pairs Series 2

Combined pair (I)	Combined pair (J)	Mean difference (I – J)	*Sig.* [b]
9_12 & 12_9	10_11 & 11_10	.33*	.000
10_11 & 11_10	9_12 & 12_9	-.33*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

5.3.2.2.1.3 Series 3 (/ɔː/ vs. /ɒ/).

Table 28

Mean Hit Rate for Combined Pairs Series 3

Combined pair	*M*	*SE*
13_17 & 17_13	.98	.01
14_15 & 15_14	.88	.02
15_16 & 16_15	.28	.03
15_16 & 16_15	.68	.04

Note. N = 52.

One-way repeated measures ANOVA with the lower-bound correction showed a similar pattern of effect of spectral distance on mean discrimination hit rate of vowel tokens, $F(1, 51) = 155.854$, $p < .000$, and a large effect size $\eta^2 = .75$. Post hoc analyses using the Bonferroni correction revealed consistently the observation that reduction in spectral distance elicited significant reduction in mean discrimination hit rate across all combined pairs. The two most noticeable differences were observed between the prototypic vowel pair 13_17 & 17_13 and the combined pair 14_15 & 15_14 (*MD* = .70, $p < .000$), and the combined pair 14_16 & 16_14 and 14_15 & 15_14 (*MD* = .60, $p <.000$).

Table 29

Within-Subjects Effects for Mean Hit Rate for Combined Pairs Series 3

Source: Series 3	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	3	155.854	.000	.753	1.000
Greenhouse-Geisser	1.734	155.854	.000	.753	1.000
Huynh-Feldt	1.789	155.854	.000	.753	1.000
Lower-bound	1	155.854	.000	.753	1.000

[a]. Computed using alpha = .05

Table 30

Pairwise Comparisons of Mean Hit Rate for Combined Pairs Series 3

Combined pair (I)	Combined pair (J)	Mean difference (I – J)	*Sig.* [b]
13_17 & 17_13	14_16 & 16_14	.10*	.000
	14_15 & 15_14	.70*	.000
	15_16 & 16_15	.30*	.000
14_16 & 16_14	14_15 & 15_14	.60*	.000
	15_16 & 16_15	.19*	.000
14_15 & 15_14	15_16 & 16_15	-.41*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

5.3.2.2.1.4 Series 4 (/ɑ:/ vs. /ʌ/).

Table 31

Mean Hit Rate for Combined Pairs Series 4

Combined pair	*M*	*SE*
18_22 & 22_18	.45	.03
19_20 & 20_19	.28	.03
21_19 & 19_21	.30	.03
20_21 & 21_20	.32	.03

Note. N = 52.

Consistent with previously demonstrated findings, one-way repeated measures ANOVA with the lower-bound correction revealed the effect of spectral distance on mean discriminated hit rate between combined pairs of vowel tokens, $F(1, 51) = 19.993$, $p < .000$, and a large effect size $\eta^2 = .28$, though much less than the reported ones for previous series. However, post hoc analyses using the Bonferroni correction revealed a slightly different pattern of discrimination rate with regard to spectral distance, eliciting significant reduction only between the prototypic vowel pair (18_22 & 22_18) and other combined pairs of vowel tokens. An interesting finding with this series is that, irrespective of observed difficulty in reaching a chance level in discriminating the prototypic vowels /ɑ:/ and /ʌ/, participants could use spectral differences in F_1 and F_2 to report statistically significantly different mean discrimination hit rate among the main prototypic pair and the remaining pairs.

Table 32

Within-Subjects Effects for Mean Hit Rate for Combined Pairs Series 4

Source: Series 4	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	3	19.993	.000	.282	1.000
Greenhouse-Geisser	2.720	19.993	.000	.282	1.000
Huynh-Feldt	2.889	19.993	.000	.282	1.000
Lower-bound	1	19.993	.000	.282	.992

[a]. Computed using alpha = .05

Table 33

Pairwise Comparisons of Mean Hit Rate for Combined Pairs Series 4

Combined pair (I)	Combined pair (J)	Mean difference (I – J)	*Sig.* ^{b}
18_22 & 22_18	19_20 & 20_19	.16*	.000
	21_19 & 19_21	.14*	.000
	20_21 & 21_20	.13*	.000
19_20 & 20_19	21_19 & 19_21	-.02	1.000
	20_21 & 21_20	-.03	.911
21_19 & 19_21	20_21 & 21_20	-.01	1.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
^{b}. Adjustment for multiple comparisons: Bonferroni.

5.3.2.2.1.5 Series 5 (/a:/ vs. /æ/).

Table 34

Mean Hit Rate for Combined Pairs Series 5

Combined pair	*M*	*SE*
23_29 & 29_23	.98	.01
24_28 & 28_24	.96	.01
25_26 & 26_25	.38	.03
25_27 & 27_25	.70	.03
26_27 & 27_26	.32	.03

Note. N = 52.

In line with the previous findings, one-way repeated measures ANOVA with the lower-bound correction revealed a significant effect of spectral distance on mean discrimination hit rate, $F(1, 51) = 198.459$, $p < .000$, and a very large effect size $\eta^2 = .80$. Furthermore, as might reasonably be expected, identical spectral distance for the combined pairs 25_26 & 26_25 and 26_27 & 27_26 yielded non-significant difference in mean discrimination hit rate. Remarkable differences were observed among the prototypic vowel pair and the pair 25_26 & 26_25 (*MD* = .60, $p < .000$), and the pair 26_27 & 27_26 (*MD* = .66, $p < .00$). A similar finding was found between the combined pairs 24_28 & 28_24 and 26_27 & 27_26 (*MD* = .64, $p < .000$). Based on these results, we may suggest that there is empirical evidence that a perceptual category boundary for the prototypic vowels /ɑː/ and /æ/ would exist midway between sounds 25 and 27.

Table 35

Within-Subjects Effects for Mean Hit Rate for Combined Pairs Series 5

Source: Series 5	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	4	198.459	.000	.796	1.000
Greenhouse-Geisser	2.764	198.459	.000	.796	1.000
Huynh-Feldt	2.938	198.459	.000	.796	1.000
Lower-bound	1	198.459	.000	.796	1.000

[a]. Computed using alpha = .05

Table 36

Pairwise Comparisons of Mean Hit Rate for Combined Pairs Series 5

Combined pair (I)	Combined pair (J)	Mean difference (I – J)	*Sig.* [b]
23_29 & 29_23	24_28 & 28_24	.02	.827
	25_26 & 26_25	.60*	.000
	25_27 & 27_25	.28*	.000
	26_27 & 27_26	.66*	.000
24_28 & 28_24	25_26 & 26_25	.58*	.000
	25_27 & 27_25	.26*	.000
	26_27 & 27_26	.64*	.000
25_26 & 26_25	25_27 & 27_25	.32*	.000
	26_27 & 27_26	.06	1.000
25_27 & 27_25	26_27 & 27_26	.38*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

5.3.2.2.1.6 Series 6 (/æ/ vs. /ʌ/).

Table 37

Mean Hit Rate for Combined Pairs Series 6

Combined pairs	*M*	*SE*
30_35 & 35_30	.86	.03
31_34 & 34_31	.75	.03
32_33 & 33_32	.35	.02

Note. N = 52.

The significant effect of spectral distance from the prototypic vowels /æ/ and /ʌ/ was demonstrated by results of one-way repeated measures ANOVA with the lower-bound correction, $F(1, 51) = 195.285$, $p < .000$, $\eta^2 = .793$. Mean discrimination hit rate between prototypic vowels was statistically significantly higher than for the two other vowel token pairs, 31_34 & 34_31 (*MD* = .12, *p* < .000), and 32_33 & 33_32 (*MD* = .52, *p* < .000). That is, the more the spectral distance near the /æ/ vs. /ʌ/ on the perceptual map, the more the discrimination and vice versa.

Table 38

Within-Subjects Effects for Mean Hit Rate for Combined Pairs Series 6

Source: Series 6	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	2	195.285	.000	.793	1.000
Greenhouse-Geisser	1.568	195.285	.000	.793	1.000
Huynh-Feldt	1.609	195.285	.000	.793	1.000
Lower-bound	1	195.285	.000	.793	1.000

[a]. Computed using alpha = .05

Table 39

Pairwise Comparisons of Mean Hit Rate for Combined Pairs Series 6

Combined pair (I)	Combined pair (J)	Mean difference (I – J)	*Sig.* [b]
30_35 & 35_30	31_34 & 34_31	.12*	.000
	32_33 & 33_32	.52*	.000
31_34 & 34_31	32_33 & 33_32	.40*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

5.3.2.2.1.7 Series 7 (/e/ vs. /ɜː/).

Table 40

Mean Hit Rate for Combined Pairs Series 7

Combined pair	*M*	*SE*
36_40 & 40_36	.76	.03
37_38 & 38_37	.32	.02
37_39 & 39_37	.50	.03
38_39 & 39_38	.32	.03

Note. N = 52.

As with previous findings, one-way repeated measures ANOVA with the lower-bound correction proved statistically significant effect of spectral distance on mean discrimination hit rates of synthetic vowel tokens, *F*(1, 51) = 82.568, *p* <.000, and a large effect size $\eta^2 = .618$. An equal spectral distance in combined vowel pairs 37_38 & 38_37 and 38_39 & 39_38 showed identical discrimination rate. The perceptual relevance in discrimination hit rate among the combined pairs 37_39 & 39_37 and 38_39 & 39_38, and equal discrimination hit rate among pairs 38_39 & 39_38 and 37_38 & 38_37, may suggest the existence of a perceptual category boundary between prototypic representations midway between sounds 37 and 39.

Table 41

Within-Subjects Effects for Mean Hit Rate for Combined Pairs Series 7

Source: Series 7	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	3	82.568	.000	.618	1.000
Greenhouse-Geisser	2.607	82.568	.000	.618	1.000
Huynh-Feldt	2.607	82.568	.000	.618	1.000
Lower-bound	1	82.568	.000	.618	1.000

[a]. Computed using alpha = .05

Table 42

Pairwise Comparisons of Mean Hit Rate for Combined Pairs Series 7

Combined pair (I)	Combined pair (J)	Mean difference (I – J)	*Sig.* [b]
36_40 & 40_36	37_38 & 38_37	.43*	.000
	37_39 & 39_37	.26*	.000
	38_39 & 39_38	.43*	.000
37_38 & 38_37	37_39 & 39_37	-.18*	.000
	38_39 & 39_38	.00	1.000
37_39 & 39_37	38_39 & 39_38	.18*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

5.3.2.2.1.8 Series 8 (/ɜː/ vs. /ʊ/).

Table 43

Mean Hit Rate for Combined Pairs Series 8

Combined pairs	*M*	*SE*
41_45 & 45_41	.99	.01
42_43 & 43_42	.86	.02
42_44 & 44_42	.93	.02
43_44 & 44_43	.13	.02

Note. N = 52.

One-way repeated measures ANOVA with the use of the lower-bound correction performed on mean discrimination hit rates among combined vowel token pairs in Series 8 revealed exceptionally interesting findings. Varying the spectral difference in terms of the first formant alone showed statistically significant differences in mean discrimination hit rates among combined vowel pairs, $F(1, 51) = 705.009$, $p < .000$, and a large effect size $\eta^2 = .933$. That is, based on F_1 alone, participants could discriminate statistically significantly between combined pairs. The large discrimination difference between the pairs 42_44 & 44_42 and 44_43 & 43_44 (.80, $p < .000$) compared to the small discrimination difference among pairs 42_44 & 44_42 and 42_43 & 43_43 (.07, $p < .000$) may confirm our previous suggestion of the existence of a perceptual category boundary between representations of /ɜː/ and /ʊ/. This perceptual category line is likely to exist between sounds 42 and 43 and reveals a larger perceptual space for /ʊ/ over /ɜː/.

Table 44

Within-Subjects Effects for Mean Hit Rate for Combined Pairs Series 8

Source: Series 8	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	3	705.009	.000	.933	1.000
Greenhouse-Geisser	1.702	705.009	.000	.933	1.000
Huynh-Feldt	1.755	705.009	.000	.933	1.000
Lower-bound	1	705.009	.000	.933	1.000

[a]. Computed using alpha = .05

Table 45

Pairwise Comparisons of Mean Hit Rate for Combined Pairs Series 8

Combined pair (I)	Combined pair (J)	Mean difference (I – J)	*Sig.* [b]
41_45 & 45_41	42_43 & 43_42	.13*	.000
	42_44 & 44_42	.06*	.000
	43_44 & 44_43	.86*	.000
42_43 & 43_42	42_44 & 44_42	-.07*	.000
	43_44 & 44_43	.73*	.000
42_44 & 44_42	43_44 & 44_43	.80*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

5.3.2.2.1.9 Series 9 (/ɜː/ vs. /uː/).

Table 46

Mean Hit Rate for Combined Pairs Series 9

Combined pairs	*M*	*SE*
46_50 & 50_46	.98	.01
47_49 & 49_47	.95	.01
47_48 & 48_47	.93	.02
48_49 & 49_48	.12	.02

Note. N = 52.

A similar pattern of previous findings was observed for Series 9. One-way repeated measures ANOVA with the lower-bound correction showed statistically significant effect of reduction in spectral distance on mean discrimination hit rates among combined vowel pairs, F (1, 51) = 828.826, $p < .000$, and a very large effect size $\eta^2 = .942$. The large discrimination difference between the combined pairs 47_49 & 49_47 and 48_49 & 49_48 (.81, $p < .000$) and the small non-significant difference among the combined pairs 47_49 & 49_47 and 47_48 & 48_47 (.02, $p = .456$) may also confirm our previous suggestion that the vowels /ɜː/ and /uː/ were discriminated quite well, and that the vowel /uː/ proves to occupy a larger space on the vowel perceptual map. The last findings with respect to Series 8 and 9 are likely to be indicative of the relatively small space allocated for the vowel /ɜː/ on the perceptual space.

Table 47

Within-Subjects Effects for Mean Hit Rate for Combined Pairs Series 9

Source: Series 9	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	3	828.826	.000	.942	1.000
Greenhouse-Geisser	1.372	828.826	.000	.942	1.000
Huynh-Feldt	1.397	828.826	.000	.942	1.000
Lower-bound	1	828.826	.000	.942	1.000

[a]. Computed using alpha = .05

Table 48

Pairwise Comparisons of Mean Hit Rate for Combined Pairs Series 9

Combined pair (I)	Combined pairs (J)	Mean difference (I – J)	*Sig.* [b]
46_50 & 50_46	47_49 & 49_47	.03*	.015
	47_48 & 48_47	.05*	.002
	48_49 & 49_48	.86*	.000
47_49 & 49_47	47_48 & 48_47	.02	.456
	48_49 & 49_48	.83*	.000
47_48 & 48_47	48_49 & 49_48	.81*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

5.3.2.2.1.10 *Prototypic Vowel Pairs.*

Table 49

Mean Hit Rate for Prototypic Vowels Pairs and Spectral Distance

			Spectral difference (in Hz)	
Combined pair	*M*	*SE*	F_1	F_2
/iː/ vs. /ɪ/	.98	.01	100	402
/uː/ vs. /ʊ/	.71	.04	9	180
/ɔː/ vs. /ɒ/	.98	.01	163	162
/ɑː/ vs. /ʌ/	.45	.03	19	103
/ɑː/ vs. /æ/	.98	.01	13	372
/æ/ vs. /ʌ/	.86	.03	15	269
/e/ vs. /ɜː/	.76	.03	13	248
/ɜː/ vs. /ʊ/	.99	.01	124	0
/ɜː/ vs. /uː/	.98	.01	133	179

Note. N = 52.

Table 50

Within-Subjects Effects for Mean Hit Rate for Pairs of Prototypic Vowels

Source: Vowel pair	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	8	69.357	.000	.576	1.000
Greenhouse-Geisser	3.271	69.357	.000	.576	1.000
Huynh-Feldt	3.522	69.357	.000	.576	1.000
Lower-bound	1	69.357	.000	.576	1.000

[a]. Computed using alpha = .05

To investigate differences in discrimination among RP English prototypic monophthong pairs, one-way ANOVA with repeated measures was performed on mean discrimination hit rates among pairs of prototypic vowels. With the lower-bound correction as shown below in Table 50, mean discrimination hit rates differed statistically significantly between prototypic vowels, $F(1, 51) = 69.357$, $p < .000$, and a significantly large effect size $\eta^2 = .576$. The obtained findings depict a complex picture of participants' reliance on spectral cues in discriminating differences among the presented stimuli. Overall, the larger the spectral distance the more the discrimination hit rate rule was respected, with one remarkable exception. As shown in Table 49, RP English prototypic vowels with large spectral distances in both F_1 and F_2 were discriminated at high rates compared to spectrally close vowels such as /æ/ vs. /ʌ/, /uː/ vs. /ʊ/, and /e/ vs. /ɜː/. However, the exception was with the vowel pair /ɜː/ vs. /ʊ/ that was highly discriminated regardless of the identical similarity between /ɜː/ and /ʊ/ along the F_2 scale.

Table 51 shows that participants could report accurately the difference in the presented vowels, relying exclusively on variation along the F_1 scale. It is to note that pairs of prototypic vowels with higher spectral distance along F_2 were relatively discriminated at lower rates than those with higher spectral distance along F_1, suggesting a perceptual preference among participants to rely on F_1 in their decision. For instance, the pair /ɑː/ vs. /ʌ/ was statistically significantly discriminated at a lower rate than the pairs /iː/ vs. /ɪ/ ($MD = -.53$, $p < .000$), /ɜː/ vs. /uː/ ($MD = -.54$, $p < .000$), and /ɔː/ vs. /ɒ/ ($MD = -.53$, $p < .000$). Participants showed the probable pattern of preference to using F_1 in prototypic vowel discrimination also with the pairs /ɜː/ vs. /uː/ and /æ/ vs. /ʌ/ (*MD* = -.12, *p* < .01), where F_2 spectral distance in the latter pair did not likely help in detecting difference. Pairs of prototypic vowels with considerably similar spectral distance in F_1 and large spectral distance in F2 further confirmed this pattern of preference. They were discriminated at higher rates and there were statistically no significant differences in their mean discrimination hit rates: the pairs /iː/ vs. /ɪ/, /ɑː/ vs. /æ/, /ɜː/ vs. /uː/, and /ɜː/ vs. /ʊ/ were discriminated at significantly equal rates.

Table 51

Pairwise Comparisons of Mean Hit Rate for Pairs of Prototypic Vowels

Combined pair (I)	Combined pair (J)	Mean difference (I – J)	*Sig.* [b]
/iː/ vs. /ɪ/	/uː/ vs. /ʊ/	.27*	.000
	/ɔː/ vs. /ɒ/	-.00	1.000
	/ɑː/ vs. /ʌ/	.53*	.000
	/ɑː/ vs. /æ/	.00*	1.000
	/æ/ vs. /ʌ/	.11*	.005
	/e/ vs. /ɜː/	.22*	.000
	/ɜː/ vs. /ʊ/	-.01	1.000
	/ɜː/ vs. /uː/	-.01	1.000
/uː/ vs. /ʊ/	/ɔː/ vs. /ɒ/	-.27*	.000
	/ɑː/ vs. /ʌ/	.26*	.000
	/ɑː/ vs. /æ/	.27*	.000
	/æ/ vs. /ʌ/	-.16*	.017
	/e/ vs. /ɜː/	-.05	1.000
	/ɜː/ vs. /ʊ/	-.28*	.000
	/ɜː/ vs. /uː/	-.28*	.000
/ɔː/ vs. /ɒ/	/ɑː/ vs. /ʌ/	.53*	.000
	/ɑː/ vs. /æ/	.00	1.000
	/æ/ vs. /ʌ/	.12*	.003
	/e/ vs. /ɜː/	.22*	.000
	/ɜː/ vs. /ʊ/	-.01	1.000
	/ɜː/ vs. /uː/	-.00	1.000
/ɑː/ vs. /ʌ/	/ɑː/ vs. /æ/	-.53*	.000
	/æ/ vs. /ʌ/	-.42*	.000
	/e/ vs. /ɜː/	-.31*	.000
	/ɜː/ vs. /ʊ/	-.54*	.000
	/ɜː/ vs. /uː/	-.54*	.000
/ɑː/ vs. /æ/	/æ/ vs. /ʌ/	.11*	.003
	/e/ vs. /ɜː/	.22*	.000
	/ɜː/ vs. /ʊ/	-.01	1.000
	/ɜː/ vs. /uː/	-.01	1.000
/æ/ vs. /ʌ/	/e/ vs. /ɜː/	.11	.091
	/ɜː/ vs. /ʊ/	-.12*	.001
	/ɜː/ vs. /uː/	-.12*	.003

Note. Pairwise comparisons are based on estimated marginal means.

*. The mean difference is significant at the .05 level.

[b]. Adjustment for multiple comparisons: Bonferroni.

Table 51 (*continued*)

Pairwise Comparisons of Mean Hit Rate of Pairs of Prototypic Vowels

Combined pair (I)	Combined pair (J)	Mean difference (I – J)	*Sig.* [b]
/e/ vs. /ɜː/	/ɜː/ vs. /ʊ/	-.23*	.000
	/ɜː/ vs. /uː/	-.22*	.000
/ɜː/ vs. /ʊ/	/ɜː/ vs. /uː/	.01	1.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

5.3.2.2.2 Reaction time.

To investigate probable significant differences in reaction time, we ran one-way ANOVA with repeated measures on mean reaction time for vowel token pairs per series. With the lower-bound correction, there were statistically non-significant results for all series including comparisons among pairs of prototypic vowels, except for Series 6, $F(1, 51) = 4.150$, $p < .05$, $\eta^2 = .075$. Detailed results results for Series 6 are presented here.

Table 52

Mean Reaction Time Series 6

Combined pairs	*M*	*SE*
30_35 & 35_30	4.95	0.36
31_34 & 34_31	4.30	0.28
32_33 & 33_32	5.63	0.30

Note. N = 52.

Table 53

Within-Subjects Effects for Mean Reaction Time Series 6

	df	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	2	4.150	.019	.075	.721
Greenhouse-Geisser	1.888	4.150	.021	.075	.703
Huynh-Feldt	1.958	4.150	.019	.075	.714
Lower-bound	1	4.150	.047	.075	.516

[a]. Computed using alpha = .05

Indeed, there is little to explain about the effect of spectral distances and compared vowel token pairs on mean reaction time, as it is hard to attribute the very few significant differences to a specific reason or manipulation during the experiment. We were tempted to trivialise the importance of these results and skip their analysis, therefore.

Table 54

Pairwise Comparisons of Mean Reaction Time Series 6

Combined pair (I)	Combined pair (J)	Mean difference (I – J)	*Sig.* [b]
30_35 & 35_30	31_34 & 34_31	.65	.559
	32_33 & 33_32	-.69	.520
31_34 & 34_31	32_33 & 33_32	-1.33*	.005

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

5.3.2.3 Variation in discrimination among proficiency groups.

To investigate possible discrimination variation among proficiency groups in our sample, we divided participants into three groups as a function of their achievement level in *Speaking and Phonetics* course: (a) 19 poor (10 ≤ course score < 12), (b) 21 mediocre (12 ≤ course score < 14), and (c) 12 high (course score ≥ 14). Prior to grouping of participants, we equally considered methodological issues relating to robustness of the ANOVA test and scoring issues relating to variability and significance. We avoided (a) having groups of largely unequal size and (b) trivialising existing course scoring differentials, as the range of course scores was substantially narrow. Below is Table 55 for descriptive statistics of discrimination test data between proficiency groups.

Table 55

Descriptive Statistics of Discrimination Test Data between Proficiency Groups

Vowel Series	Combined stimulus pair	Proficiency group	*M*	*SD*	*SE*	95% CI Lower bound	95% CI Upper bound
Series 1 (/iː/ vs. /ɪ/)	1_8 & 8_1	Poor	.98	.04	.01	.96	1.00
		Mediocre	.97	.07	.01	.94	1.00
		High	.99	.02	.01	.98	1.00
	2_7 & 7_2	Poor	.96	.07	.02	.92	.99
		Mediocre	.95	.09	.02	.91	.99
		High	.98	.09	.03	.92	1.03
	3_6 & 6_3	Poor	.86	.15	.03	.79	.93
		Mediocre	.81	.13	.03	.75	.87
		High	.91	.13	.04	.83	.99
	4_5 & 5_4	Poor	.26	.27	.06	.13	.39
		Mediocre	.23	.19	.04	.14	.32
		High	.23	.29	.08	.04	.41
Series 2 (/uː/ vs. /ʊ/)	9_12 & 12_9	Poor	.69	.33	.07	.53	.85
		Mediocre	.67	.31	.07	.53	.81
		High	.80	.30	.09	.61	.99
	10_11 & 11_10	Poor	.42	.29	.07	.28	.56
		Mediocre	.36	.21	.05	.27	.46
		High	.33	.17	.05	.23	.44
Series 3 (/ɔː/ vs. /ɒ/)	13_17 & 17_13	Poor	.97	.08	.02	.93	1.01
		Mediocre	.99	.03	.01	.98	1.00
		High	.98	.03	.01	.96	1.00
	14_16 & 16_14	Poor	.89	.16	.04	.82	.97
		Mediocre	.88	.15	.03	.81	.95
		High	.85	.21	.06	.72	.98
	14_15 & 15_14	Poor	.33	.26	.06	.20	.46
		Mediocre	.24	.16	.04	.17	.32
		High	.25	.21	.06	.12	.38
	15_16 & 16_15	Poor	.69	.27	.06	.56	.82
		Mediocre	.68	.27	.06	.56	.80
		High	.68	.24	.07	.52	.83

Note. N = 52.

Table 55 (*continued*)

Descriptive Statistics of Discrimination Test Data between Proficiency Groups

Vowel Series	Combined stimulus pair	Proficiency group	*M*	*SD*	*SE*	95% CI Lower bound	95% CI Upper bound
Series 4 (/ɑː/ vs. /ʌ/)	18_22 & 22_18	Poor	.42	.29	.07	.28	.56
		Mediocre	.43	.24	.05	.32	.54
		High	.53	.18	.05	.42	.64
	19_20 & 20_19	Poor	.26	.19	.04	.17	.35
		Mediocre	.30	.21	.05	.21	.40
		High	.28	.19	.05	.16	.40
	19_21 & 21_19	Poor	.30	.20	.05	.21	.40
		Mediocre	.32	.19	.04	.24	.41
		High	.28	.14	.04	.18	.37
	20_21 & 21_20	Poor	.30	.22	.05	.19	.40
		Mediocre	.33	.23	.05	.23	.43
		High	.33	.17	.05	.22	.43
Series 5 (/ɑː/ vs. /æ/)	23_29 & 29_23	Poor	.97	.05	.01	.94	.99
		Mediocre	.99	.03	.01	.97	1.00
		High	.98	.04	.01	.95	1.00
	24_28 & 28_24	Poor	.94	.10	.02	.89	.99
		Mediocre	.97	.06	.01	.95	1.00
		High	.96	.07	.02	.92	1.01
	26_25 & 25_26	Poor	.42	.21	.05	.32	.53
		Mediocre	.41	.26	.06	.28	.53
		High	.26	.18	.05	.14	.38
	25_27 & 27_25	Poor	.71	.23	.05	.60	.82
		Mediocre	.71	.24	.05	.59	.82
		High	.67	.23	.07	.53	.81
	26_27 & 27_26	Poor	.34	.24	.06	.23	.46
		Mediocre	.31	.20	.04	.22	.40
		High	.31	.18	.05	.20	.43

Note. N = 52.

Table 55 (*continued*)

Descriptive Statistics of Discrimination Test Data between Proficiency Groups

Vowel Series	Combined stimulus pair	Proficiency group	*M*	*SD*	*SE*	95% CI Lower bound	95% CI Upper bound
Series 6 (/æ/ vs. /ʌ/)	30_35 & 35_30	Poor	.89	.13	.03	.83	.95
		Mediocre	.82	.26	.06	.70	.94
		High	.90	.14	.04	.82	.99
	31_34 & 34_31	Poor	.77	.16	.04	.70	.85
		Mediocre	.69	.28	.06	.56	.82
		High	.80	.21	.06	.67	.94
	32_33 & 33_32	Poor	.37	.21	.05	.27	.47
		Mediocre	.30	.14	.03	.24	.36
		High	.40	.15	.04	.30	.50
Series 7 (/e/ vs. /ɜ:/)	36_40 & 40_36	Poor	.80	.22	.05	.70	.91
		Mediocre	.73	.25	.05	.62	.84
		High	.74	.29	.08	.55	.92
	38_37 & 37_38	Poor	.36	.18	.04	.27	.44
		Mediocre	.28	.14	.03	.22	.35
		High	.35	.19	.05	.23	.46
	37_39 & 39_37	Poor	.59	.23	.05	.48	.70
		Mediocre	.44	.21	.05	.34	.53
		High	.47	.30	.09	.27	.66
	38_39 & 39_38	Poor	.42	.21	.05	.32	.52
		Mediocre	.28	.18	.04	.20	.36
		High	.24	.20	.06	.12	.37

Note. N = 52.

Participants among the three proficiency groups demonstrated little variation in discrimination across all series with the absence of any systematic pattern to establish a general tendency. As for Series 7, poor achievers discriminated among all synthetic stimulus pairs at a higher rate than high achievers. For Series 6, high achievers discriminated among all synthetic stimulus pairs at a higher rate than poor achievers. However, for Series 8, discrimination varied among poor and high achievers across synthetic stimuli pairs.

Table 55 (*continued*)

Descriptive Statistics of Discrimination Test Data between Proficiency Groups

Vowel Series	Combined stimulus pair	Proficiency group	*M*	*SD*	*SE*	95% CI Lower bound	95% CI Upper bound
Series 8 (/ɜː/ vs. /ʊ/)	41_45 & 45_41	Poor	.98	.06	.01	.95	1.00
		Mediocre	.99	.02	.00	.98	1.00
		High	1.00	.00	.00	1.00	1.00
	42_43 & 43_42	Poor	.87	.14	.03	.80	.94
		Mediocre	.86	.12	.03	.80	.91
		High	.85	.18	.05	.73	.96
	42_44 & 44_42	Poor	.92	.14	.03	.85	.99
		Mediocre	.93	.10	.02	.88	.97
		High	.93	.11	.03	.86	1.01
	43_44 & 44_43	Poor	.15	.18	.04	.07	.23
		Mediocre	.11	.11	.02	.06	.17
		High	.11	.09	.03	.05	.17
Series 9 (/ɜː/ vs. /uː/)	46_50 & 50_46	Poor	.97	.07	.02	.93	1.00
		Mediocre	.99	.02	.01	.98	1.00
		High	1.00	.00	.00	1.00	1.00
	47_49 & 49_47	Poor	.94	.11	.03	.89	.99
		Mediocre	.95	.09	.02	.91	.99
		High	.97	.06	.02	.93	1.00
	47_48 & 48_47	Poor	.92	.13	.03	.86	.98
		Mediocre	.91	.13	.03	.86	.97
		High	.96	.07	.02	.92	1.01
	49_48 & 48_49	Poor	.13	.17	.04	.05	.21
		Mediocre	.14	.13	.03	.08	.20
		High	.07	.09	.03	.01	.13

Note. N = 52.

We performed a one-way between subjects ANOVA to investigate and compare the effect of proficiency in *Speaking and Phonetics* course on discrimination among the three groups of achievers across all experimental manipulations, i.e. series. As displayed in Table 56 below, there were no significant effects of level of speaking proficiency on discrimination of synthetic stimulus pairs across all series, except for one negligible combined stimulus pair in Series 7 (/e/ vs. /ɜː/), with $F(2, 49) = 3.91$, $p < .05$. Post hoc comparisons with the Bonferroni test indicated that the mean discrimination for the combined stimulus pair (38_39 & 39_38) for the poor group was significantly higher than the high achievers, with $MD = .18$, $p < .05$. Taken together, these findings suggest that level of proficiency in *Speaking and Phonetics* course does not have an effect on discrimination among synthetic stimulus pairs. Specifically, our findings suggest that speaking proficiency may not offer discrimination perceptual advantage.

Table 56

Variation in Discrimination Abilities between Proficiency Groups

Vowel Series	Combined stimulus pair	*F*	*Sig.*	ES η^2
Series 1 (/iː/ vs. /ɪ/)	1_8 & 8_1	0.94	0.40	0.04
	2_7 & 7_2	0.43	0.65	0.02
	3_6 & 6_3	2.18	0.12	0.08
	4_5 & 5_4	0.10	0.91	0.00
Series 2 (/uː/ vs. /ʊ/)	9_12 & 12_9	0.68	0.51	0.03
	10_11 & 11_10	0.55	0.58	0.02
Series 3 (/ɔː/ vs. /ɒ/)	13_17 & 17_13	0.50	0.61	0.02
	14_16 & 16_14	0.25	0.78	0.01
	14_15 & 15_14	0.93	0.40	0.04
	15_16 & 16_15	0.01	0.99	0.00
Series 4 (/ɑː/ vs. /ʌ/)	18_22 & 22_18	0.86	0.43	0.03
	19_20 & 20_19	0.23	0.80	0.01
	19_21 & 21_19	0.27	0.77	0.01
	20_21 & 21_20	0.12	0.89	0.00
Series 5 (/ɑː/ vs. /æ/)	23_29 & 29_23	1.50	0.23	0.06
	24_28 & 28_24	0.81	0.45	0.03
	26_25 & 25_26	2.15	0.13	0.08
	25_27 & 27_25	0.12	0.89	0.00
	26_27 & 27_26	0.15	0.87	0.01
Series 6 (/æ/ vs. /ʌ/)	30_35 & 35_30	0.95	0.39	0.04
	31_34 & 34_31	1.16	0.32	0.05
	32_33 & 33_32	1.50	0.23	0.06

Note. N = 52
df (between groups) = 2
df (within groups) = 49

Table 56 (*continued*)

Variation in Discrimination Abilities between Proficiency Groups

Vowel Series	Combined stimulus pair	*F*	*Sig.*	ES η^2
Series 7 (/e/ vs. /ɜː/)	36_40 & 40_36	0.49	0.62	0.02
	38_37 & 37_38	1.15	0.33	0.04
	37_39 & 39_37	2.21	0.12	0.08
	38_39 & 39_38	3.91	0.03	0.14
Series 8 (/ɜː/ vs. /ʊ/)	41_45 & 45_41	1.55	0.22	0.06
	42_43 & 43_42	0.09	0.91	0.00
	42_44 & 44_42	0.04	0.96	0.00
	43_44 & 44_43	0.48	0.62	0.02
Series 9 (/ɜː/ vs. /uː/)	46_50 & 50_46	2.22	0.12	0.08
	47_49 & 49_47	0.31	0.73	0.01
	47_48 & 48_47	0.68	0.51	0.03
	49_48 & 48_49	1.02	0.37	0.04

Note. N = 52
df (between groups) = 2
df (within groups) = 49

5.3.2.4 Discrimination abilities and speaking proficiency.

Pearson product-moment correlation was performed to determine the relationship between the participants' discrimination among combined stimulus pairs and their speaking proficiency. Computed correlations varied between negative and positive correlation in the absence of statistical significance and an established pattern[38]. Obtained findings suggest discrimination abilities and speaking proficiency to be unrelated. Further analysis of isolated vowel tokens identification will further prove whether perceptual abilities relate to speaking proficiency.

---

[38] Given the large number of computed relationships between discrimination and speaking proficiency, the correlation matrix was not inserted for word processing reasons. Detailed results appear in Appendix 11.

Chapter 6

Isolated Vowel Tokens Identification Test

Introduction

Isolated vowel tokens identification test was motivated by an interest in finding the optimal decision in accurately identifying a vowel as a function of its temporal properties, i.e. duration. The choice of different durations for the same vowel is justified by contextual variability of the occurrence of the vowel among adjacent sounds and position within the word (Gimson, 1979, p. 95).

We designed the second experiment to

a) Investigate Algerian learners' perceptual abilities in identifying various RP English prototypic monophthongs as a function of temporal change,

b) Provide empirical evidence for perceptual difficulties of RP English prototypic monophthongs among Algerian learners of English, and

c) Reveal any systematic misidentification of presented stimuli.

6.1 Stimuli Materials

Stimuli in isolated vowel tokens identification test consisted of 76 vowel tokens, described as follows:

a) Five long prototypic vowels /iː/, /uː/, /ɜː/, /ɑː/ and /ɔː/, each with 8 stimuli, whose durations ranged from 150 to 290 ms, and

b) Six short prototypic vowels /ɪ/, /ʊ/, /e/, /æ/, /ʌ/ and /ɒ/, each with 6 stimuli, whose durations ranged from 90 to 190 ms. The vowel /ə/ was not included for its high variability and unavailable spectral measurements.

c) All vowel tokens were saved in 32-bit WAV files for test administration.

6.2 Procedure

The second experiment is an *m-alternative* forced choice test (*m-AFC*) consisting of 16 blocks of trials, each with 50 trials. We did not count the first 40 trials, for we treated them as familiarisation. In fact, this was done systematically without informing the participants. Each trial began with the presentation of an alert message with task heading and instructions on the computer screen. Participants listened to one vowel token in each trial, and had to identify the heard vowel among suggested choices. Each trial was preceded by a 150-ms initial silence and followed by no final silence.

There were two series of suggested choices,

a) Test 1, 7 choices /iː/, /ɪ/, /ʊ/, /uː/, /e/, /ɜː/ and /ə/

b) Test 2, 6 choices /ɒ/, /ʌ/, /ɑː/, /æ/, /ɔː/ and /e/

Participants had to respond simply by clicking on one of the buttons on the PC screen, with the IPA symbol of the chosen vowel on it. There was no logical order for the presentation of vowels on the PC screen, and vowel tokens were randomised in both identification tests. Upon completion of a block of trials, the *TP* software provided participants with immediate feedback and alerted them to their performance level for accuracy and reaction time. Participants had a systematic break after completion of every block of trials, during which we tried to alleviate the test pressure that would be building by encouraging them to report on any issue relating to the test. Though participants were encouraged to break enough before resuming the test of their own volition and not to respond too quickly, they reported the identification test to have been more difficult than the discrimination test. It took participants 4 hours, as in same different vowel discrimination experiment, to complete the test. Upon completion of the test, we collected test auto-saved result files of participants' data from their workstations.

6.3 Results

Analyses were performed on data to report on (a) vowel identification hit rates, (b) confusion matrices, and (c) hit rate variation across vowels. Reaction time analysis was not reported herein in detail as tests' results showed no statistical significance (see Table 88). Hit rates were calculated based on mean correct identification of vowels among participants, and pooled results are presented as mean identification hit rates plotted as a function of temporal manipulations. Confusion matrices are based on aggregate mean identification hit rate of vowels. For convenience purposes only, initial presentation of results will follow the quantitative principle of vowels' categorisation.

6.3.1 Vowel identification across temporal manipulations.

6.3.1.1 Short vowels.

As shown in *Figure* 25 below, RP English prototypic monophthongs were identified differently along their temporal properties. Surprisingly however, pooled mean identification hit rates did not reach chance level. At 90 ms, participants identified the vowels at varying rates across the various temporal manipulations. For the vowel /æ/, the identification rate improved from .06 at 90 ms, .08 at 110 ms, .12 at 130 ms, .20 ms at 150 ms, .22 at 17 ms, to .30 ms at 190 ms, suggesting considerable difficulty of the vowel identification at short durations. A similar case was observed with /ʊ/ that was identified at higher rates with increasing duration of the vowel, from .34 at 90 ms, .41 at 110 ms, 43 at 130 ms, .46 at 170 ms, to .42 at 190 ms, except for the sudden fall in its identification at 150 ms (.13). This change in identification of the vowel /ʊ/ is singular and is hard to explain, aside from presumed participants' fatigue and attentional collapse.

Other vowels demonstrated different identification rate patterns across temporal manipulations. The vowel /e/ showed a particular pattern of identification. It was identified at an increasing rate from 90 to 150 ms, then at a decreasing rate starting from 170 ms, suggesting a special use of the temporal cue for this vowel. In a rather consistent manner, the vowel /ɒ/ showed a consistent drop in its identification rate across temporal increase. The vowel /ɒ/ was best identified at 90 ms with a mean hit rate of .12 that dropped continuously with duration augmentation. For the vowel /ɪ/, the identification rate pattern was quite constant across increasing duration, though it dropped at 190 ms, marking probably a new pattern for identification. Likewise, the vowel /ʌ/ revealed a similar identification rate pattern to that of /ɪ/, with quite constant identification rate of .26 at 90 ms to .25 at 170 ms, then a drop at 190 ms with an identification rate of .20.

Overall, there seems to be no specific common duration at which all vowels are highly, accurately identified. The vowels /ɪ/, /ʌ/ and /ɒ/ are accurately identified better at shorter durations ranging from 90 to 110 ms. To the contrary, the vowels /ʊ, æ/ are accurately better identified at longer durations, ranging from 170 to 190 ms, and the vowel /e/ at a midway duration of 150 ms.

6.3.1.2 Long vowels.

Figure 26 below shows a general picture about the way participants identified RP English prototypic monophthongs accurately, with an increasing rate as a function of temporal augmentation. *Figure* 26 shows a uniform pattern in identification hit rates of monophthongs as a function of duration. These pooled results are relatively higher than those observed with short vowels, ranging from .03 to .08 at 150 ms to .96 at 290 ms for almost all vowels, except for /uː/. As the figure shows, identification hit rates of long vowels reached more than half of the trials at 210 ms, except for /uː/ whose identification rate was comparatively lower than the other vowels (.40). As opposed to /iː/ that was identified at a hit rate of .60 at 190 ms, identification hit rate of /uː/ did not reach .60 at a duration of 290 ms, at which all other vowels' identification hit rates were beyond .80. Compared to its counterpart on the temporal scale, /ʊ/, which was constantly identified at a higher rate compared to short vowels, the vowel /uː/ was the least accurately identified among long vowels.

As *Figures* 25 and 26 show, crossover point in identification of both short and long vowels was remarkably different for different durations. For this purpose, a one way ANOVA with repeated measures will be performed on mean identification rates to report statistically significant differences in identification across temporal manipulations in Section 5.4.1. However, consistent with our adopted methodology and prior to performing one-way repeated measures ANOVA, sensitivity measurements and confusion matrices will be presented to find out probable systematic sources of identification, misidentification and confusion.

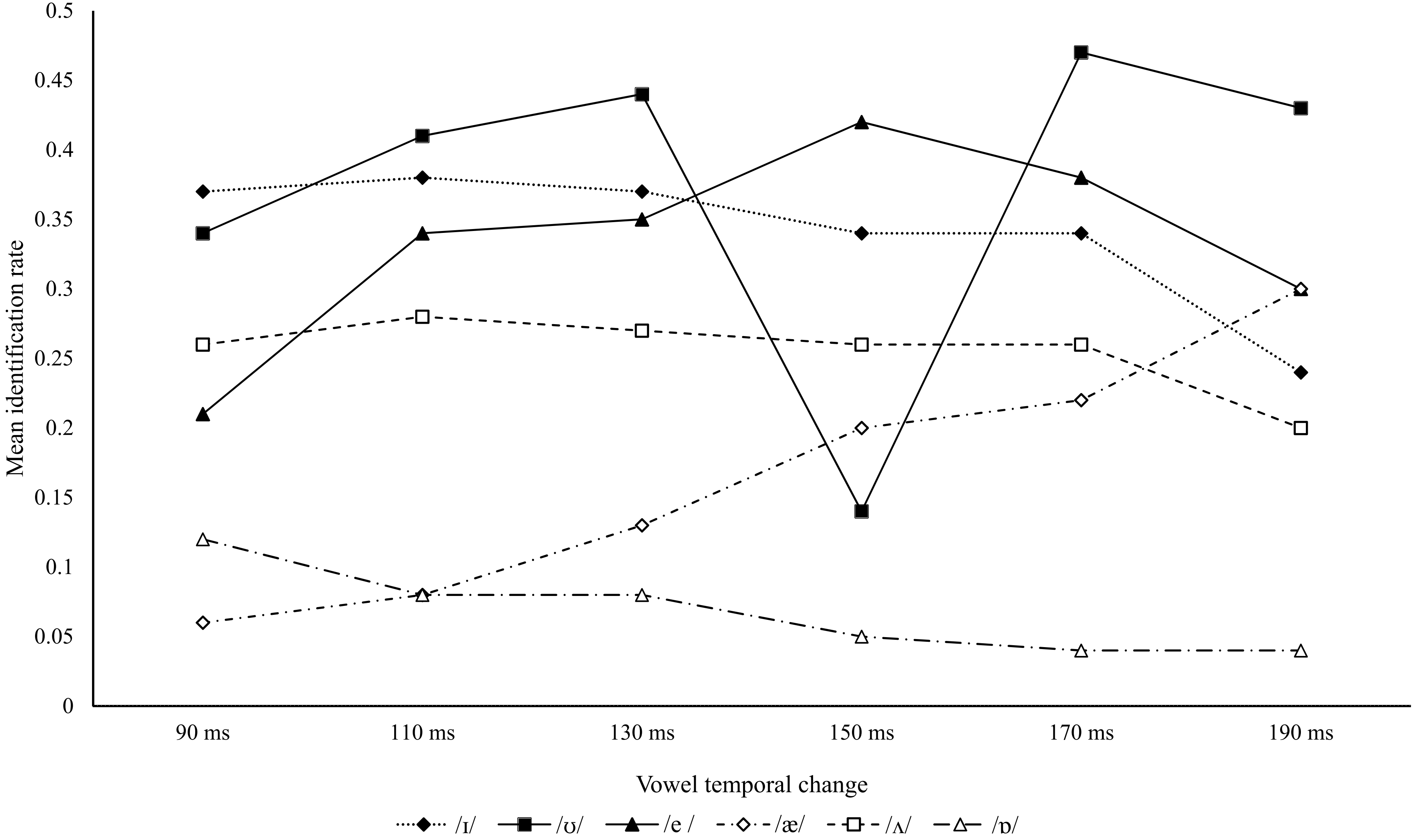


Figure 25. Mean identification rates of short vowels across temporal change

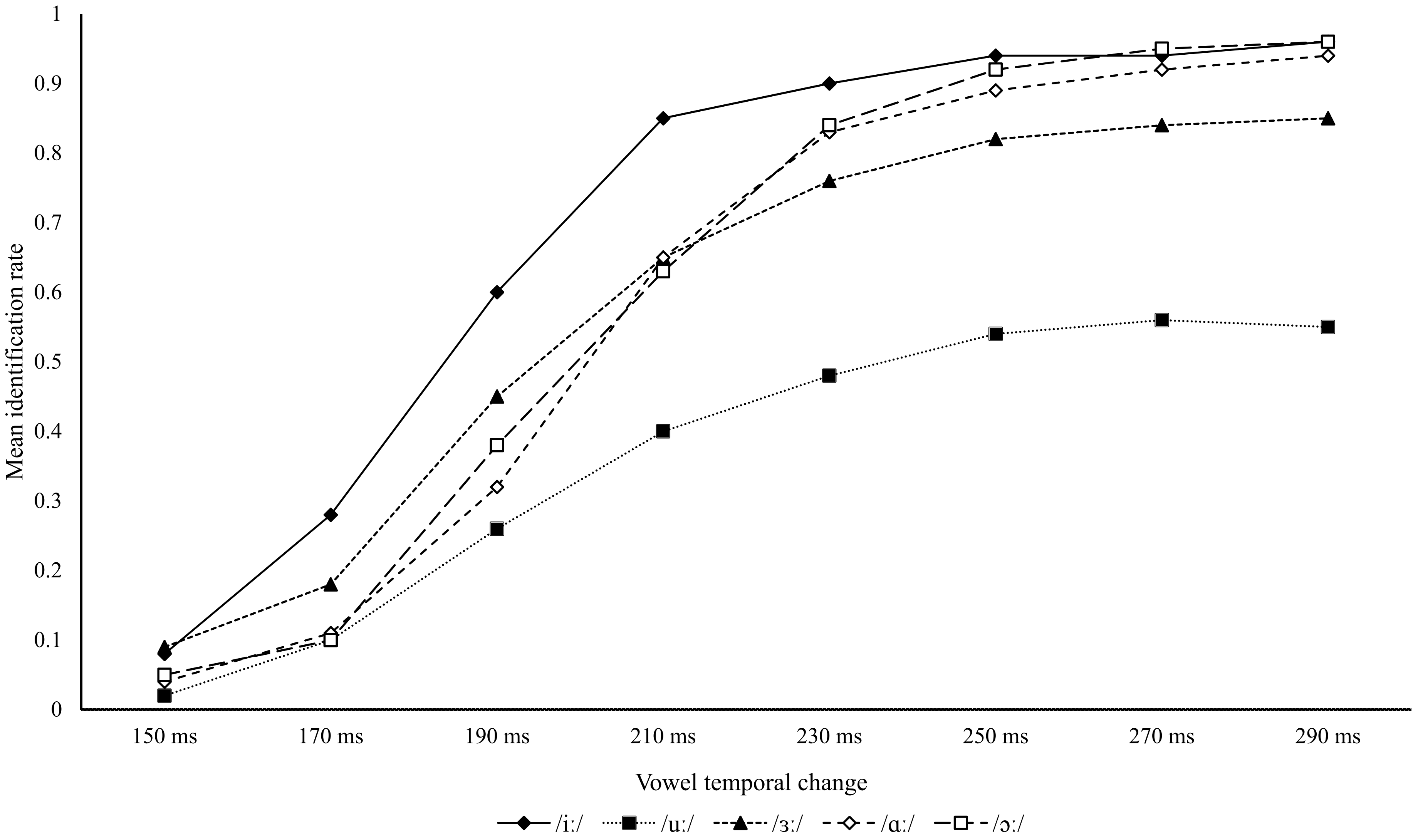


Figure 26. Mean identification rates of long vowels across temporal change

6.3.2 Sensitivity measurements and confusion matrices.

Confusion matrices in Tables 57 and 58 reveal considerably higher mean identification hit rates for long vowels compared to short ones, with sensitivity indices ranging from -0.60 for /ɒ/ to 1.85 for /iː/. High significant sensitivity measurements are found with long vowels /iː/, /ɔː/, /ɜː/, and /ɑː/ in a descending pattern, except for /uː/. However, the vowels /ɒ/, /æ/, /ʌ/, /e/, /ɪ/, /uː/, and /ʊ/ are found to be poorly identified, with participants being almost insensitive to /ɒ/ (*d'* = -0.60) and /æ/ (*d'* = 0.02). For cases of insensitivity and low sensitivity, vowels prove to be misidentified, not only with their opposites on the temporal continuum, but also to near vowels on the perceptual map. For temporal confusion, the vowels /iː/, /ɑː/, and /ɔː/ were misidentified with their short counterparts, /ɪ/, /æ/, and /ɒ/, respectively. The vowels /uː/ and /ɜː/ were highly misidentified with /ʊ/ and /ɜː/, and /e/, respectively, and the vowel /ɒ/ with both /ʌ/ and /æ/. The short /ɪ/ was highly misidentified with /ə/ and /e/ and unexpectedly with /ʊ/, given probably the position of the latter on the perceptual map. These findings suggest perceptual identification difficulty of isolated vowel tokens across temporal manipulations, with the short vowels demonstrating rather unchanging hit rates compared to long ones.

Table 57

Aggregate Mean Identification of Isolated Vowel Tokens (7-AFC Confusion Matrix)

	Vowel choice *(m-alternative forced choice)*						
Vowel	/iː/	/ɪ/	/ʊ/	/uː/	/e/	/ɜː/	/ə/
/iː/	**.69**	*.26*	.01	.03	.00	.00	.00
/ɪ/	.03	**.34**	.15	.03	.19	.02	*.24*
/ʊ/	.03	.03	**.37**	.08	*.29*	.07	.15
/uː/	.02	.02	*.23*	**.36**	.12	*.22*	.04
/e/	.00	.03	.04	.01	**.33**	.10	*.48*
/ɜː/	.00	.00	.05	.07	*.18*	**.58**	.12

Note. N = 53. *M* = 7. Number of stimuli = 6. Values in boldface are mean hit rates for vowels. Underlined values represent high misidentification. Values are rounded to second decimal. ∑ Trials per short vowel = 60. ∑ Trials per long vowel = 80. Sum of rates per vowel may not equal 1 due to rounding.

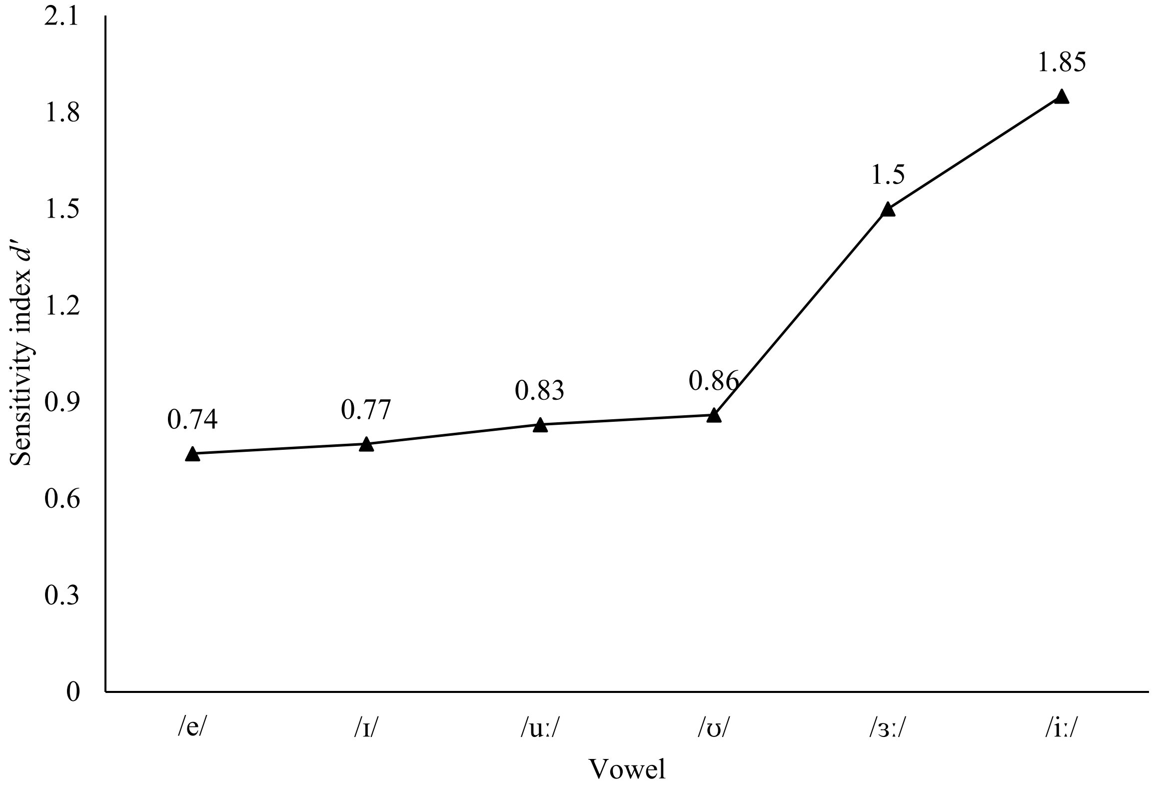


Figure 27. Identification of isolated vowel tokens in an *m-AFC* test. *m* = 7. N = 53

Table 58

Aggregate Mean Identification of Isolated Vowel Tokens (6-AFC Confusion Matrix)

	Vowel choice *(m-alternative forced choice)*					
Vowel	/ɒ/	/ʌ/	/ɑː/	/æ/	/ɔː/	/e/
/ɒ/	**.07**	*.45*	.08	*.35*	.00	.04
/ʌ/	.01	**.25**	.04	*.27*	.00	*.42*
/ɑː/	.00	.12	**.58**	*.25*	.00	.03
/æ/	.01	*.16*	.02	**.17**	.00	*.65*
/ɔː/	*.39*	.00	.01	.00	**.60**	.00

Note. N = 53. *M* = 6. Number of stimuli = 5. Values in boldface are mean hit rates for vowels. Underlined values represent high misidentification. Values are rounded to second decimal. ∑ Trials per short vowel = 60. ∑ Trials per long vowel = 80. Sum of rates per vowel may not equal 1 due to rounding.

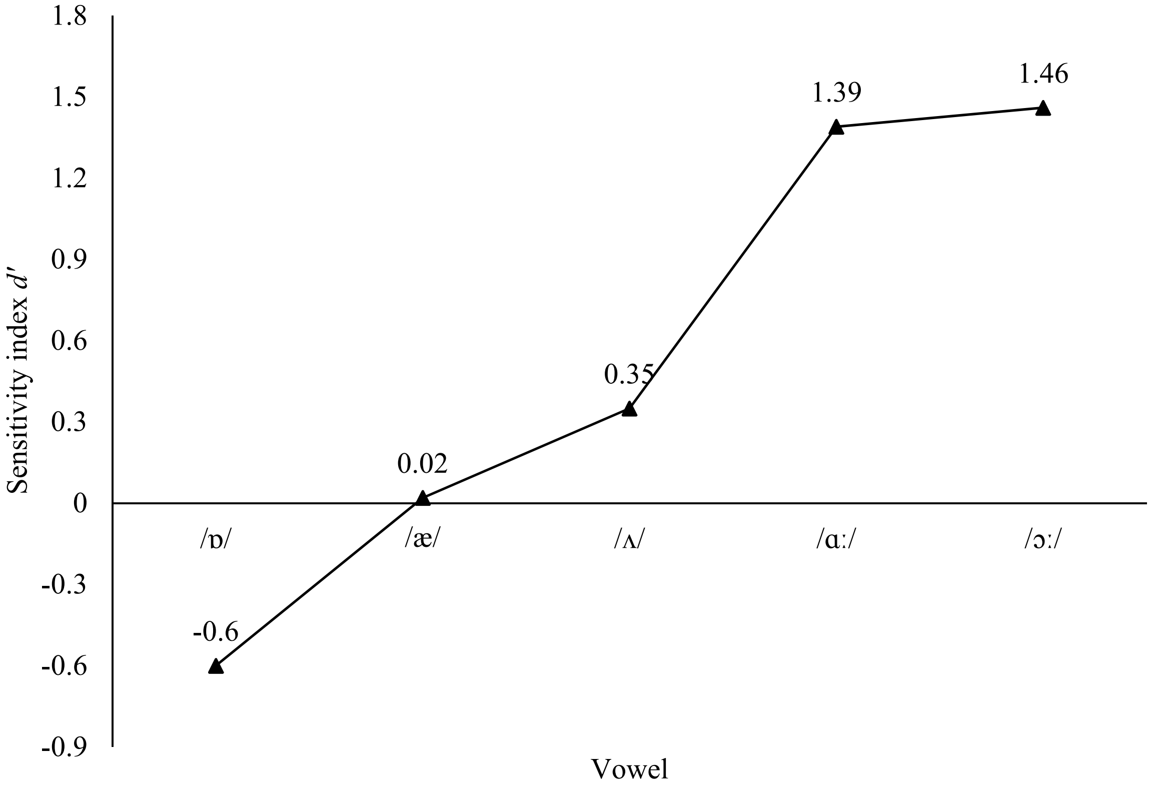


Figure 28. Identification of isolated vowel tokens in an *m-AFC* test. *m* = 6. N = 53

6.3.3 Variation in vowel identification.

To check whether identification differed significantly among short and long vowels across various temporal manipulations, we ran a one-way ANOVA with repeated measures on obtained data as shown in Tables 59 below, we present results in the following sections.

6.3.3.1 Short vowels.

Table 59

Mean Identification of Short Vowels across Temporal Manipulation

	Vowel					
	/æ/	/e/	/ɪ/	/ɒ/	/ʊ/	/ʌ/
Duration	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
90 ms	.06 (.10)	.21 (.23)	.38 (.34)	.12 (.19)	.34 (.28)	.26 (.27)
110 ms	.08 (.12)	.34 (.24)	.39 (.34)	.08 (.17)	.41 (.29)	.29 (.28)
130 ms	.12 (.17)	.35 (.27)	.37 (.34)	.08 (.16)	.43 (.30)	.28 (.26)
150 ms	.20 (.20)	.43 (.25)	.36 (.35)	.05 (.14)	.13 (.13)	.26 (.24)
170 ms	.22 (.25)	.37 (.28)	.33 (.32)	.04 (.11)	.46 (.28)	.25 (.24)
190 ms	.30 (.32)	.30 (.25)	.24 (.26)	.04 (.08)	.42 (.26)	.20 (.22)

Note. N = 53

6.3.3.1.1 Identification at 90 ms.

Table 60

Within-Subjects Effects for Mean Identification of Short Vowels at 90 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	5	13.226	.000	.203	1.000
Greenhouse-Geisser	3.758	13.226	.000	.203	1.000
Huynh-Feldt	4.088	13.226	.000	.203	1.000
Lower-bound	1.000	13.226	.001	.203	.946

[a]. Computed using alpha = .05

One-way repeated measures ANOVA with the lower-bound correction, as displayed in Table 60 above, determined that identification rates differed statistically significantly between short vowels at 90 ms duration, $F(1, 52) = 13.226$, $p < .000$, $\eta^2 = .203$. The vowel /ɪ/ was identified at a significantly higher rate than /æ/ (*MD* = .32, $p < .000$), /ɒ/ (*MD* = .26, $p < .000$) and /e/ (*MD* = .16, $p < .030$). The vowel /ʌ/ was identified at a statistically significantly higher rate than /æ/ (*MD* = .20, $p < .000$), and /ʊ/ was highly identified than /ɒ/ (*MD* = .22, $p < .002$).

Table 61

Pairwise Comparisons of Mean Identification of Short Vowels at 90 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/ɪ/	/æ/	.32*	.000
	/ʌ/	.12	.824
	/ɒ/	.26*	.000
	/ʊ/	.04	1.000
	/e/	.16*	.022
/æ/	/ʌ/	-.20*	.000
	/ɒ/	-.06	1.000
	/ʊ/	-.28*	.000
	/e/	-.16*	.001
/ʌ/	/ɒ/	.14*	.044
	/ʊ/	-.08	1.000
	/e/	.05	1.000
/ɒ/	/ʊ/	-.22*	.001
	/e/	-.10	.182
/ʊ/	/e/	.12	.162

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.3.1.2 Identification at 110 ms.

Table 62

Within-Subjects Effects for Mean Identification of Short Vowels at 110 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	5	17.287	.000	.249	1.000
Greenhouse-Geisser	3.636	17.287	.000	.249	1.000
Huynh-Feldt	3.943	12.787	.000	.249	1.000
Lower-bound	1.000	12.787	.000	.249	.983

[a]. Computed using alpha = .05

As shown in Table 62 above , one-way repeated measures ANOVA with the lower-bound correction determined that identification rates differed statistically significantly between short vowels at 110 ms duration, $F(1, 52) = 17.287, p < .000, \eta^2 = .249$. Post hoc test with the Bonferroni adjustment provided further detail. The vowel /ɪ/ was always identified at a significantly higher rate than /æ/ (*MD* = .31, *p* < .000) and /ɒ/ (*MD* = .31, *p* < .000). The vowel /ʌ/ was again identified significantly at a higher rate than /æ/ (*MD* = .21, *p* < .000) and /ɒ/ (*MD* = .20, *p* < .000), and /ʊ/ was statistically highly identified than /æ/ (*MD* = .33, *p* < .000). However, the vowels /e/ and /ʊ/ were identified at similar rates (*MD* = .07, *p* = 1).

Table 63

Pairwise Comparisons of Mean Identification of Short Vowels at 110 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* ^{b}
/ɪ/	/æ/	.31*	.000
	/ʌ/	.10	1.000
	/ɒ/	.31*	.000
	/ʊ/	-.02	1.000
	/e/	.05	1.000
/æ/	/ʌ/	-.21*	.000
	/ɒ/	-.01	1.000
	/ʊ/	-.33*	.000
	/e/	-.26*	.000
/ʌ/	/ɒ/	.20*	.000
	/ʊ/	-.13	.408
	/e/	-.05	1.000
/ɒ/	/ʊ/	-.33*	.000
	/e/	-.26*	.000
/ʊ/	/e/	.07	1.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.3.1.3 *Identification at 130 ms.*

Table 64

Within-Subjects Effects for Mean Identification of Short Vowels at 130 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power ^{a}
Sphericity Assumed	5	15.328	.000	.228	1.000
Greenhouse-Geisser	3.848	15.328	.000	.228	1.000
Huynh-Feldt	4.194	15.328	.000	.228	1.000
Lower-bound	1.000	15.328	.000	.228	.970

[a]. Computed using alpha = .05

One-way repeated measures ANOVA with the lower-bound correction determined that identification rates differed significantly between short vowels at 130 ms duration, yielding an *F* ratio of $F(1, 52) = 15.328$, $p < .000$, $\eta^2 = .228$. Post hoc test with the Bonferroni adjustment showed that the vowel /ɪ/ was identified at a higher rate than /æ/ (*MD* = .25, *p* < .000), /ɒ/ (*MD* = .29, *p* < .000) but not /e/ (*MD* = .02, *p* = 1). The vowel /ʌ/ was again identified at a significantly higher rate than /æ/ (*MD* = .16, *p* < .042) and /ɒ/ (*MD* = .20, *p* < .000), and /ʊ/ was statistically identified at a higher rate than /æ/ (*MD* = .31, *p* < .000). However, /e/ and /ʊ/ identification rates showed statistically non-significant differences (*MD* = .08, *p* = 1).

Table 65

Pairwise Comparisons of Mean Identification of Short Vowels at 130 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* ^{b}
/ɪ/	/æ/	.25*	.000
	/ʌ/	.09	1.000
	/ɒ/	.29*	.000
	/ʊ/	-.06	1.000
	/e/	.02	1.000
/æ/	/ʌ/	-.16*	.041
	/ɒ/	.04	1.000
	/ʊ/	-.31*	.000
	/e/	-.23*	.000
/ʌ/	/ɒ/	.20*	.000
	/ʊ/	-.15	.135
	/e/	-.07	1.000
/ɒ/	/ʊ/	-.35*	.000
	/e/	-.27*	.000
/ʊ/	/e/	.08	1.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
^{b}. Adjustment for multiple comparisons: Bonferroni.

6.3.3.1.4 Identification at 150 ms.

Table 66

Within-Subjects Effects for Mean Identification of Short Vowels at 150 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	5	18.416	.000	.262	1.000
Greenhouse-Geisser	3.646	18.416	.000	.262	1.000
Huynh-Feldt	3.954	18.416	.000	.262	1.000
Lower-bound	1.000	18.416	.000	.262	.988

[a]. Computed using alpha = .05

One-way repeated measures ANOVA with the lower-bound correction determined that identification rates differed statistically significantly across short vowels at 150 ms duration, yielding an F ratio of $F(1, 52) = 18.416$, $p < .000$, $\eta^2 = .262$. This time, the vowel /ɪ/ was identified at a significantly higher rate than /ɒ/ (*MD* = .30, $p < .000$) and /ʊ/ (*MD* = .22, $p < .004$) but not /æ/ (*MD* = .16, $p = .059$), /ʌ/ (*MD* = .10, $p = 1$), and /e/ (*MD* = -.07, $p = 1$). The vowel /ʌ/ was identified at a higher rate than /ɒ/ (*MD* = .20, $p < .000$) and /ʊ/ (*MD* = .12, $p < .014$), and /ʊ/ was consistently statistically identified at a lower rate than all short vowels in the test, except from /æ/ (*MD* = .06, $p = 1$).

Table 67

Pairwise Comparisons of Mean Identification of Short Vowels at 150 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/ɪ/	/æ/	.16	.059
	/ʌ/	.10	1.000
	/ɒ/	.30*	.000
	/ʊ/	.22*	.003
	/e/	-.07	1.000
/æ/	/ʌ/	-.06	1.000
	/ɒ/	.14*	.003
	/ʊ/	.06	1.000
	/e/	-.23*	.000
/ʌ/	/ɒ/	.20*	.000
	/ʊ/	.12*	.013
	/e/	-.17*	.028
/ɒ/	/ʊ/	-.08	.095
	/e/	-.37*	.000
/ʊ/	/e/	-.29*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.3.1.5 *Identification at 170 ms.*

Table 68

Within-Subjects Effects for Mean Identification of Short Vowels at 170 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	5	16.147	.000	.237	1.000
Greenhouse-Geisser	4.126	16.147	.000	.237	1.000
Huynh-Feldt	4.526	16.147	.000	.237	1.000
Lower-bound	1.00	16.147	.000	.237	.976

[a]. Computed using alpha = .05

One-way repeated measures ANOVA with the lower-bound correction determined that identification rates differed statistically significantly between short vowels at 170 ms duration, yielding an *F* ratio of $F(1, 52) = 16.147$, $p < .000$, $\eta^2 = .237$. This time, the vowel /ɪ/ was identified at a significantly higher rate than /ɒ/ only (*MD* = .29, $p < .000$) and /ɒ/ proved to be statistically significantly identified at a significantly lower rate than all other short vowels in the experiment. The vowel /ʊ/ was identified at a significantly higher rate than /æ/ (*MD* = .24, *p* < .002), /ʌ/ (*MD* = .21, *p* < .004), and /ɒ/ (*MD* = .42, *p* < .000).

Table 69

Pairwise Comparisons of Mean Identification of Short Vowels at 170 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/ɪ/	/æ/	.11	.883
	/ʌ/	.08	1.000
	/ɒ/	.29*	.000
	/ʊ/	-.14	.444
	/e/	-.05	1.000
/æ/	/ʌ/	-.03	1.000
	/ɒ/	.18*	.000
	/ʊ/	-.24*	.001
	/e/	-.15	.141
/ʌ/	/ɒ/	.21*	.000
	/ʊ/	-.21*	.003
	/e/	-.12	.332
/ɒ/	/ʊ/	-.42*	.000
	/e/	-.33*	.000
/ʊ/	/e/	.09	1.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.3.1.6 *Identification at 190 ms.*

Table 70

Within-Subjects Effects for Mean Identification of Short Vowels at 190 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	5	14.364	.000	.216	1.000
Greenhouse-Geisser	3.818	14.364	.000	.216	1.000
Huynh-Feldt	4.158	14.364	.000	.216	1.000
Lower-bound	1.000	14.364	.000	.216	.961

[a]. Computed using alpha = .05

One-way repeated measures ANOVA with the lower-bound correction determined that identification rates differed statistically significantly between short vowels at 190 ms duration, yielding an *F* ratio of $F(1, 52) = 14.364$, $p < .000$, $\eta^2 = .216$. This time, the vowel /ʊ/ was identified at a significantly higher rate than /ɪ/ (*MD* = .18, $p < .003$) and /ʌ/ (*MD* = .23, $p < .000$), while the vowel /ɒ/ was always statistically significantly identified at a lower rate compared to all short vowels.

Table 71

Pairwise Comparisons of Mean Identification of Short Vowels at 190 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* ^{b}
/ɪ/	/æ/	-.06	1.000
	/ʌ/	.04	1.000
	/ɒ/	.20*	.000
	/ʊ/	-.18*	.002
	/e/	-.06	1.000
/æ/	/ʌ/	.11	1.000
	/ɒ/	.26*	.000
	/ʊ/	-.12	.701
	/e/	.00	1.000
/ʌ/	/ɒ/	.16*	.000
	/ʊ/	-.23*	.000
	/e/	-.10	.557
/ɒ/	/ʊ/	-.38*	.000
	/e/	-.26*	.000
/ʊ/	/e/	.12	.089

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
^{b}. Adjustment for multiple comparisons: Bonferroni.

6.3.3.2 Long vowels.

Table 72

Mean Identification of Long Vowels across Temporal Manipulations

	/iː/	/ɑː/	/ɔː/	/uː/	/ɜː/
Duration	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
150 ms	.08 (.17)	.04 (.06)	.05 (.08)	.03 (.07)	.08 (.11)
170 ms	.28 (.23)	.11 (.14)	.11 (.14)	.10 (.13)	.17 (.20)
190 ms	.60 (.23)	.32 (.21)	.37 (.29)	.25 (.22)	.45 (.28)
210 ms	.85 (.19)	.64 (.23)	.63 (.26)	.40 (.25)	.65 (.23)
230 ms	.90 (.17)	.83 (.19)	.84 (.18)	.49 (.30)	.76 (.22)
250 ms	.94 (.16)	.88 (.14)	.92 (.12)	.53 (.28)	.82 (.21)
270 ms	.95 (.17)	.92 (.13)	.95 (.08)	.56 (.28)	.83 (.20)
290 ms	.96 (.17)	.93 (.10)	.96 (.07)	.55 (.29)	.85 (.21)

Note. N = 53.

6.3.3.2.1 Identification at 150 ms.

Table 73

Within-Subjects Effects for Mean Identification of Long Vowels at 150 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	4	4.083	.003	.073	.911
Greenhouse-Geisser	2.310	4.083	.015	.073	.759
Huynh-Feldt	2.424	4.083	.013	.073	.744
Lower-bound	1	4.083	.048	.073	.509

[a]. Computed using alpha = .05

One-way repeated measures ANOVA with the lower-bound correction determined that identification rates differed statistically significantly between long vowels at 150 ms duration, yielding an *F* ratio of $F(1, 52) = 4.083$, $p < .05$, $\eta^2 = .073$. However, the reported statistical significance did not show interesting results, except for the slightly higher rate of identification for the vowel /ɜː/ compared to /ɑː/ (*MD* = .05, $p < .029$) and /uː/ (*MD* = .06, $p < .004$). Differences in identification rates of other vowels were not statistically significant. Considering identification rates as displayed above in Table 70, it would be reasonable to suggest that, for the participants in this research, spectral qualities alone might not be sufficient for identifying accurately English prototypic long vowel at 150 ms.

Table 74

Pairwise Comparisons of Mean Identification of Long Vowels at 150 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/iː/	/ɑː/	.05	.544
	/ɔː/	.04	1.000
	/uː/	.06	.219
	/ɜː/	-.00	1.000
/ɑː/	/ɔː/	-.01	1.000
	/uː/	.01	1.000
	/ɜː/	-.05*	.028
/ɔː/	/uː/	.02	1.000
	/ɜː/	-.04	.139
/uː/	/ɜː/	-.06*	.003

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.3.2.2 Identification at 170 ms.

Table 75

Within-Subjects Effects for Mean Identification of Long Vowels at 170 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	4	12.282	.000	.191	1.000
Greenhouse-Geisser	3.107	12.282	.000	.191	1.000
Huynh-Feldt	3.327	12.282	.000	.191	1.000
Lower-bound	1	12.282	.001	.191	.930

[a]. Computed using alpha = .05

One-way repeated measures ANOVA with the lower-bound correction determined that identification rates differed statistically significantly between long vowels at 170 ms duration, yielding an F ratio of $F(1, 52) = 12.282$, $p < .01$, $\eta^2 = .191$. However, the reported statistical significance showed interesting results about the vowel /iː/, being identified at a statistically significantly higher rate than all other vowels, except for the vowel /ɜː/ ($MD = .11$, $p = .094$). This suggests a probable perceptual advantage for the vowel /iː/ compared to other vowels in the test.

Table 76

Pairwise Comparisons of Mean Identification of Long Vowels at 170 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/iː/	/ɑː/	.17*	.000
	/ɔː/	.17*	.000
	/uː/	.18*	.000
	/ɜː/	.11	.094
/ɑː/	/ɔː/	.00	1.000
	/uː/	.01	1.000
	/ɜː/	-.06	.198
/ɔː/	/uː/	.01	1.000
	/ɜː/	-.06	.329
/uː/	/ɜː/	-.07	.139

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.3.2.3 Identification at 190 ms.

Table 77

Within-Subjects Effects for Mean Identification of Long Vowels at 190 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	4	23.922	.000	.315	1.000
Greenhouse-Geisser	3.801	23.922	.000	.315	1.000
Huynh-Feldt	4.000	23.922	.000	.315	1.000
Lower-bound	1	23.922	.000	.315	.998

[a]. Computed using alpha = .05

One-way repeated measures ANOVA with the lower-bound correction determined that identification rates differed statistically significantly between long vowels at 190 ms duration, yielding an *F* ratio of $F(1, 52) = 23.922$, $p < .000$, $\eta^2 = .315$. This time, the reported statistical significance showed more effect size and further interesting results about the vowel /iː/ again, being identified at a statistically significantly higher rate than all other vowels. Additionally, the vowel /ɜː/ was significantly identified at a higher rate than /ɑː/ ($MD = .13$, $p < .013$) and /uː/

(*MD* = .20, *p* < .000), demonstrating a certain pattern in participants' perceptual abilities at identifying specific vowels with increasing duration.

Table 78

Pairwise Comparisons of Mean Identification of Long Vowels at 190 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/iː/	/ɑː/	.28*	.000
	/ɔː/	.23*	.000
	/uː/	.35*	.000
	/ɜː/	.15*	.007
/ɑː/	/ɔː/	-.06	1.000
	/uː/	.07	.943
	/ɜː/	-.13*	.012
/ɔː/	/uː/	.12	.065
	/ɜː/	-.08	.548
/uː/	/ɜː/	-.20*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.3.2.4 *Identification at 210 ms.*

Table 79

Within-Subjects Effects for Mean Identification of Long Vowels at 210 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	4	31.146	.000	.375	1.000
Greenhouse-Geisser	3.456	31.146	.000	.375	1.000
Huynh-Feldt	3.732	31.146	.000	.375	1.000
Lower-bound	1	31.146	.000	.375	1.000

[a]. Computed using alpha = .05

As for identification of long vowels at 210 ms, one-way repeated measures ANOVA with the lower-bound correction determined that identification rates differed statistically significantly between manipulated stimuli, yielding an *F* ratio of $F(1, 52) = 31.146$, $p < .000$, $\eta^2 = .375$. This time, the reported statistical significance showed again a larger effect size and similar results about the vowel /iː/ as previously described, being statistically identified at a significantly higher rate than all other vowels. Additionally, the vowels /ɜː/ and /ɑː/ were identified at a significantly higher rate than /uː/ (*MD* = .25, $p < .000$) for both. These results suggest again perceptual advantage of the vowel /iː/ over the other vowels. Remarkably interesting, at 210 ms, all long vowels were identified above chance level, except for /uː/.

Table 80

Pairwise Comparisons of Mean Identification of Long Vowels at 210 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/iː/	/ɑː/	.21*	.000
	/ɔː/	.22*	.000
	/uː/	.46*	.000
	/ɜː/	.21*	.000
/ɑː/	/ɔː/	.01	1.000
	/uː/	.25*	.000
	/ɜː/	-.00	1.000
/ɔː/	/uː/	.24*	.000
	/ɜː/	-.02	1.000
/uː/	/ɜː/	-.25*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.3.2.5 *Identification at 230 ms.*

Table 81

Within-Subjects Effects for Mean Identification of Long Vowels at 230 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	4	37.648	.000	.420	1.000
Greenhouse-Geisser	3.205	37.648	.000	.420	1.000
Huynh-Feldt	3.440	37.648	.000	.420	1.000
Lower-bound	1	37.648	.000	.420	1.000

[a]. Computed using alpha = .05

At 230 ms, one-way repeated measures ANOVA with the lower-bound correction determined that identification rates differed statistically significantly between vowels, yielding an *F* ratio of $F(1, 52) = 37.648$, $p < .000$, $\eta^2 = .420$. This time, the vowel /uː, ɑː, ɔː/ were identified significantly at similar rates (*MD* ranging between .07 and .08, *p* value ranging between .321 and .447), and that /uː/ was identified at a significantly lower rate than all other vowels.

Table 82

Pairwise Comparisons of Mean Identification of Long Vowels at 230 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/iː/	/ɑː/	.08	.321
	/ɔː/	.07	.447
	/uː/	.41*	.000
	/ɜː/	.15*	.005
/ɑː/	/ɔː/	-.01	1.000
	/uː/	.34*	.000
	/ɜː/	.07	.586
/ɔː/	/uː/	.35*	.000
	/ɜː/	.08	.257
/uː/	/ɜː/	-.27*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.3.2.6 *Identification at 250 ms.*

Table 83

Within-Subjects Effects for Mean Identification of Long Vowels at 250 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	4	48.286	.000	.481	1.000
Greenhouse-Geisser	3.117	48.286	.000	.481	1.000
Huynh-Feldt	3.338	48.286	.000	.481	1.000
Lower-bound	1	48.286	.000	.481	1.000

[a]. Computed using alpha = .05

One-way repeated measures ANOVA with the lower-bound correction determined that identification rates differed statistically significantly between long vowels at 250 ms duration, yielding an *F* ratio of $F(1, 52) = 48.286, p < .000, \eta^2 = .481$. As in the just described results, the three vowels /iː/, /ɑː/ and /ɔː/ were identified significantly at similar rates, and /uː/ was identified at a significantly lower rate than all other vowels (*MD* ranging between .29 and .41, $p < .000$), suggesting a perceptual advantage of all long vowels over the vowel /uː/ at 250 ms duration. Importantly however, it was not until 250 ms duration that the vowel /uː/ came to be identified above chance level.

Table 84

Pairwise Comparisons of Mean Identification of Long Vowels at 250 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/iː/	/ɑː/	.05	.885
	/ɔː/	.02	1.000
	/uː/	.41*	.000
	/ɜː/	.12*	.026
/ɑː/	/iː/	-.05	.885
	/ɔː/	-.04	.920
	/uː/	.36*	.000
	/ɜː/	.06	.479
/ɔː/	/uː/	.39*	.000
	/ɜː/	.10*	.039
/uː/	/ɜː/	-.29*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.3.2.7 Identification at 270 ms.

Table 85

Within-Subjects Effects for Mean Identification of Long Vowels at 270 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	4	44.864	.000	.463	1.000
Greenhouse-Geisser	2.883	44.864	.000	.463	.717
Huynh-Feldt	3.019	44.864	.000	.463	1.000
Lower-bound	1	44.864	.000	.463	1.000

[a]. Computed using alpha = .05

One-way repeated measures ANOVA with the lower-bound correction determined that identification rates differed statistically significantly between long vowels at 270 ms duration, yielding an *F* ratio of $F(1, 52) = 44.864$, $p < .000$, $\eta^2 = .463$. In a constant fashion, the three vowels /iː/, /ɑː/ and /ɔː/ were identified significantly at similar rates (*MD* ranging between .00 and .03 and $p = 1$), and /uː/ was identified at a significantly lower rate than all other vowels (*MD* ranging between .28 and .39 and $p < .000$), suggesting a perceptual advantage of all long vowels over the vowel /uː/ at 270 ms. With respect to identification rate of the vowel /uː/, it was constantly identified slightly above chance level.

Table 86

Pairwise Comparisons of Mean Identification of Long Vowels at 270 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/iː/	/ɑː/	.03	1.000
	/ɔː/	-.00	1.000
	/uː/	.39*	.000
	/ɜː/	.11*	.038
/ɑː/	/ɔː/	-.03	.685
	/uː/	.36*	.000
	/ɜː/	.08	.140
/ɔː/	/uː/	.39*	.000
	/ɜː/	.12*	.002
/uː/	/ɜː/	-.28*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.3.2.8 Identification at 290 ms.

Table 87

Within-Subjects Effects for Mean Identification of Long Vowels at 290 ms

Source: Vowel	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	4	53.283	.000	.506	1.000
Greenhouse-Geisser	2.812	53.283	.000	.506	1.000
Huynh-Feldt	2.989	53.283	.000	.506	1.000
Lower-bound	1	53.283	.000	.506	1.000

[a]. Computed using alpha = .05

One-way repeated measures ANOVA with the lower-bound correction determined that identification rates differed statistically significantly between long vowels at 290 ms duration, yielding an F ratio of $F(1, 52) = 53.283$, $p < .000$, $\eta^2 = .506$. The previous pattern of identification was further confirmed. The three vowels /iː, ɑː, ɔː/ were identified significantly at similar rates (*MD* ranging between .01 and .02 and $p = 1$). Once more, /uː/ was identified at a significantly lower rate than all other vowels (*MD* ranging between .38 and .41 and $p < .000$), suggesting a constant perceptual advantage of all long vowels over the vowel /uː/, even at 290 ms. Once more, participants' pooled identification rate of the vowel /uː/ did only slightly go above chance level, reaching .60.

Table 88

Pairwise Comparisons of Mean Identification of Long Vowels at 290 ms

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/iː/	/ɑː/	.02	1.000
	/ɔː/	-.01	1.000
	/uː/	.40*	.000
	/ɜː/	.11*	.031
/ɑː/	/ɔː/	-.03	1.000
	/uː/	.38*	.000
	/ɜː/	.09	.065
/ɔː/	/uː/	.41*	.000
	/ɜː/	.11*	.003
/uː/	/ɜː/	-.30*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.4 Aggregate data analysis.

In order to compare the effect of vowel features (spectral and temporal properties) and the interaction effect between spectral and temporal features on identification rate, data measurements across temporal change were combined for aggregate analysis. Vowels' mean identification hit rates were calculated based on their individual mean identification hit rates combined. A two-way ANOVA with repeated measures was conducted to compare the main effects of vowel, vowel duration, and the interaction between vowel and duration on identification rates. Obtained results are displayed in the Table 89 through Table 97.

6.3.4.1 Short vowels.

Table 89

Aggregate Mean of Identification and Reaction Time of Short Vowel

	Identification		Reaction time	
Vowel	*M*	*SE*	*M*	*SE*
/æ/	.16	.02	3.38	0.17
/ɒ/	.07	.02	3.53	0.22
/ʌ/	.26	.03	3.66	0.19
/ʊ/	.37	.03	3.74	0.25
/ɪ/	.34	.04	3.93	0.19
/e/	.33	.03	4.24	0.26

Note. N = 53.

As clearly observed in Table 91, a two-way ANOVA with repeated measures using the lower-bound correction demonstrated the effect of vowel, temporal manipulations, and the interaction between vowel and duration on mean vowel identification rate. For the effect of vowel on mean identification rate, the analysis yielded an F ratio of $F(1, 52) = 16.066, p < .000$, $\eta^2 = .236$, indicating a significant difference in identification between vowels. Post-hoc tests using the Bonferroni correction showed that some vowels differed significantly from others. For instance, the vowel /ɪ/ was identified at a significantly higher rate than /ɒ/ ($MD = .28, p < .000$). The vowel /ʌ/ was identified at a significantly higher rate than /ɒ/ ($MD = .19, p < .000$). The vowel /ʊ/ was identified at a significantly higher rate than /ɒ/ ($MD = .30, p < .000$) and /æ/ ($MD = .20, p < .000$). The vowel /ɒ/ was identified at a significantly lower rate than all other vowels. In addition to the three previous cases, the vowel /ɒ/ was identified at a significantly lower rate than /e/ ($MD = .27, p < .000$) and /æ/ ($MD = .10, p < .023$), suggesting the fact that this vowel was the least accurately identified among all short vowels. The very low identification of the vowel /ɒ/ suggests its perceptual difficulty encountered by participants, compared to other vowels in the current set.

Similarly, temporal manipulations and interaction between vowel and duration had a significant effect on mean vowel identification rate, yielding an *F* ratio of $F(1, 52) = 6.454$, $p < .05$, $\eta^2 = .110$, and an *F* ratio of $F(1, 52) = 15.073$, $p < .000$, $\eta^2 = .225$. As shown in Table 90, short vowels identification at 90 ms was significantly lower than at 110, 130 and 170 ms (*MD* ranging between -.04 and -.05, $p < .05$). However, their identification at 170 ms was significantly higher than at 90, 15 and 170 ms (*MD* ranging between .33 and .53, $p < .05$), suggesting likely a temporal preference for identification of short vowels at 170 ms.

However, using the lower-bound correction, neither vowel, duration nor interaction between vowel and duration did have a significant effect on mean reaction time, as displayed below in Table 91.

Table 90

Pairwise Comparisons of Aggregate Mean Identification of Short Vowels

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/ɪ/	/ʌ/	.09	1.000
	/ɒ/	.28*	.000
	/ʊ/	-.02	1.000
	/e/	.01	1.000
	/æ/	.18*	.003
/ʌ/	/ɒ/	.19*	.000
	/ʊ/	-.11	.142
	/e/	-.08	1.000
	/æ/	.09	.609
/ɒ/	/ʊ/	-.30*	.000
	/e/	-.27*	.000
	/æ/	-.10*	.022
/ʊ/	/e/	.03	1.000
	/æ/	.20*	.000
/e/	/æ/	.17*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

Table 91

Within-Subjects Effects of Vowel, Duration, and Interaction between Vowel and Duration on Identification and Reaction Time of Short Vowels

Source	Measure		*df*	*F*	*Sig.*	η^2	Observed power [a]
Vowel	Identification	Sphericity Assumed	5	16.066	.000	.236	1.000
		Greenhouse-Geisser	3.759	16.066	.000	.236	1.000
		Huynh-Feldt	4.089	16.066	.000	.236	1.000
		Lower-bound	1.000	16.066	.000	.236	.976
	Reaction time	Sphericity Assumed	5	2.267	.048	.042	.730
		Greenhouse-Geisser	4.3999	2.267	.057	.042	.687
		Huynh-Feldt	4.856	2.267	.050	.042	.720
		Lower-bound	1.000	2.267	.138	.042	.315
Duration	Identification	Sphericity Assumed	5	6.454	.000	.110	.997
		Greenhouse-Geisser	3.561	6.454	.000	.110	.983
		Huynh-Feldt	3.856	6.454	.000	.110	.988
		Lower-bound	1.000	6.454	.014	.110	.703
	Reaction time	Sphericity Assumed	5	3.420	.005	.062	.905
		Greenhouse-Geisser	4.348	3.420	.008	.062	.871
		Huynh-Feldt	4.794	3.420	.006	.062	.896
		Lower-bound	1.000	3.420	.070	.062	.443
Vowel*Duration	Identification	Sphericity Assumed	25	15.073	.000	.225	1.000
		Greenhouse-Geisser	10.898	15.073	.000	.225	1.000
		Huynh-Feldt	14.004	15.073	.000	.225	1.000
		Lower-bound	1.000	15.073	.000	.225	.968
	Reaction time	Sphericity Assumed	25	1.180	.246	.022	.909
		Greenhouse-Geisser	14.862	1.180	.282	.022	.757
		Huynh-Feldt	21.156	1.180	.259	.022	.867
		Lower-bound	1.000	1.180	.282	.022	.187

[a]. Computed using alpha = .05

Table 92

Pairwise Comparisons of Aggregate Mean Identification among Temporal Manipulations of Short Vowels

Temporal manipulation (I)	Temporal manipulation (J)	Mean difference (I – J)	*Sig.* [b]
90 ms	110 ms	-.04*	.002
	130 ms	-.04*	.001
	150 ms	-.01	1.000
	170 ms	-.05*	.001
	190 ms	-.02	1.000
110 ms	130 ms	.01	1.000
	150 ms	.02	.473
	170 ms	-.02	1.000
	190 ms	.02	1.000
130 ms	150 ms	.03	.086
	170 ms	-.01	1.000
	190 ms	.02	.673
150 ms	170 ms	-.04*	.005
	190 ms	-.01	1.000
170 ms	190 ms	.03*	.017

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.4.2 Long vowels.

Table 93

Aggregate Mean of Identification and Reaction Time of Long Vowel

	Identification		Reaction time	
Vowel	*M*	*SE*	*M*	*SE*
/ɔː/	.60	.02	2.53	0.12
/iː/	.70	.02	2.91	0.20
/ɑː/	.58	.01	3.11	0.13
/uː/	.36	.03	3.98	0.20
/ɜː/	.58	.02	4.21	0.19

Note. N = 53.

As clearly observed in Table 95, a two-way ANOVA with repeated measures using the lower-bound correction demonstrated the effect of vowel, temporal manipulations, and the interaction between vowel and duration on mean vowel identification rate. The effect of vowel on mean identification rate was statistically significant, *F*(1, 52) = 53.606, p < .000, η^2= .505. Post hoc tests using the Bonferroni correction showed mean identification differed significantly between vowels in the experimentation set. The vowel /iː/ was identified at a statistically higher rate than all vowels. The vowel /ɑː/ was identified at a significantly higher rate than /uː/ (*MD* = .22, *p* < .000), and the vowel /ɔː/ was identified non-significantly at a higher rate than /ɜː/ (*MD* = .03, *p* = 1) and significantly at a higher rate than /uː/ (*MD* = .24, *p* < .000). Lastly, the vowel /ɜː/ was identified at a significantly higher rate than /uː/ (*MD* = .21, *p* < .000). The latter fact demonstrated the perceptual disadvantage of the vowel /uː/ to be identified among all long vowels and to differ significantly from them. This finding stands in complete opposition to its short counterpart that was identified at significantly higher rate than other short vowels as previously demonstrated.

Table 94

Pairwise Comparisons of Aggregate Mean Identification of Long Vowels

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* ^{b}
/iː/	/ɑː/	.11*	.000
	/ɔː/	.09*	.001
	/uː/	.33*	.000
	/ɜː/	.12*	.000
/ɑː/	/ɔː/	-.02	1.000
	/uː/	.22*	.000
	/ɜː/	.01	1.000
/ɔː/	/uː/	.24*	.000
	/ɜː/	.03	1.000
/uː/	/ɜː/	-.21*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
^{b}. Adjustment for multiple comparisons: Bonferroni.

Table 95

Within-Subjects Effects of Vowel, Duration, and Interaction between Vowel and Duration on Identification and Reaction Time of Long Vowels

Source	Measure		*df*	*F*	*Sig.*	η^2	Observed power [a]
Vowel	Identification	Sphericity Assumed	4	53.606	.000	.508	1.000
		Greenhouse-Geisser	3.142	53.606	.000	.508	1.000
		Huynh-Feldt	3.367	53.606	.000	.508	1.000
		Lower-bound	1.000	53.606	.000	.508	1.000
	Reaction time	Sphericity Assumed	4	22.152	.000	.299	1.000
		Greenhouse-Geisser	3.566	22.152	.000	.299	1.000
		Huynh-Feldt	3.860	22.152	.000	.299	1.000
		Lower-bound	1.000	22.152	.000	.299	.996
Duration	Identification	Sphericity Assumed	7	722.474	.000	.933	1.000
		Greenhouse-Geisser	3.259	722.474	.000	.933	1.000
		Huynh-Feldt	3.503	722.474	.000	.933	1.000
		Lower-bound	1.000	722.474	.000	.933	1.000
	Reaction time	Sphericity Assumed	7	1.932	.064	.036	.762
		Greenhouse-Geisser	5.990	1.932	.076	.036	.709
		Huynh-Feldt	6.856	1.932	.065	.036	.755
		Lower-bound	1.000	1.932	.171	.036	.276
Vowel*Duration	Identification	Sphericity Assumed	28	17.336	.000	.250	1.000
		Greenhouse-Geisser	12.337	17.336	.000	.250	1.000
		Huynh-Feldt	16.434	17.336	.000	.250	1.000
		Lower-bound	1.000	17.336	.000	.250	.983
	Reaction time	Sphericity Assumed	28	.466	.995	.009	.429
		Greenhouse-Geisser	16.446	.446	.972	.009	.311
		Huynh-Feldt	24.459	.466	.991	.009	.395
		Lower-bound	1.000	.446	.507	.009	.101

[a]. Computed using alpha = .05

Similarly, the vowel type demonstrated a significant effect on mean reaction time, yielding an *F* ratio of $F(1, 52) = 22.152$, $p < .000$, $\eta^2 = .299$. As displayed in Table 96, findings revealed that the vowel /ɔː/ was identified at significantly shorter reaction times than /ɑː/, /uː/ and /ɜː/. Indeed, the vowel /ɔː/ was also identified at a shorter reaction time than /iː/, though not significantly (*MD* = -.38, *p* = .869). The vowel /iː/ was identified at a significantly shorter reaction time than /uː/ (*MD* = -1.07 ms, *p* < .01) and /ɜː/ (*MD* = -1.30 ms, *p* < .000). The substantially longer reaction time to identify /uː/ and /ɜː/ likely suggests the difficulty of the latter vowels to be identified compared to others.

Table 96

Pairwise Comparisons of Reaction Time among Long Vowels

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/iː/	/ɑː/	-.20	1.000
	/ɔː/	.38	.896
	/uː/	-1.07*	.001
	/ɜː/	-1.30*	.000
/ɑː/	/ɔː/	.58*	.023
	/uː/	-.87*	.006
	/ɜː/	-.11*	.000
/ɔː/	/uː/	-1.45*	.000
	/ɜː/	-1.68*	.000
/uː/	/ɜː/	-.23	1.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

Similarly, temporal manipulations and interaction between vowel and duration had a significant effect on mean vowel identification rate, yielding an *F* ratio of $F(1, 52) = 722.474$, $p < .000$, $\eta^2 = .933$, and an *F* ratio of $F(1, 52) = 17.336$, $p < .000$, $\eta^2 = .250$. Pairwise comparisons with Bonferroni adjustments of effect of temporal manipulation on vowel identification showed statistically significant results across all durations except at 270 and 290 ms, as identification increased as a function of vowel duration as displayed in Table 97 below.

However, as displayed above in Table 95, with sphericity assumed or corrected, neither duration nor interaction between vowel and duration did have a significant effect on mean reaction time.

Table 97

Pairwise Comparisons of Aggregate Mean Identification among Temporal Manipulations of Long Vowels

Temporal manipulation (I)	Temporal manipulation (J)	Mean difference (I – J)	*Sig.* [b]
150 ms	170 ms	-.10*	.000
	190 ms	-.34*	.000
	210 ms	-.34*	.000
	230 ms	-.71*	.000
	250 ms	-.76*	.000
	270 ms	-.79*	.000
	290 ms	-.79*	.000
170 ms	190 ms	-.25*	.000
	210 ms	-.48*	.000
	230 ms	-.61*	.000
	250 ms	-.67*	.000
	270 ms	-.69*	.000
	290 ms	-.70*	.000
190 ms	210 ms	-.24*	.000
	230 ms	-.36*	.000
	250 ms	-.42*	.000
	270 ms	-.44*	.000
	290 ms	-.45*	.000
210 ms	230 ms	-.13*	.000
	250 ms	-.19*	.000
	270 ms	-.21*	.000
	290 ms	-.22*	.000
230 ms	250 ms	-.06*	.000
	270 ms	-.08*	.000
	290 ms	-.09*	.000
250 ms	270 ms	-.02	.288
	290 ms	-.03*	.039
270 ms	290 ms	-.01	1.000

Note. Pairwise comparisons are based on estimated marginal means.

*. The mean difference is significant at the .05 level.

[b]. Adjustment for multiple comparisons: Bonferroni.

6.3.4.3 All vowels combined.

As displayed in Table 98 below, one-way repeated measures ANOVA with the lower-bound correction showed that mean identification differed significantly between all vowels, $F(1, 52) = 64.969$, $p < .000$, $\eta^2 = .555$. Post hoc tests using the Bonferroni correction revealed that several pairwise comparisons differed significantly among compared vowels. The most important observations to make is the significant perceptual advantage of the vowel /iː/ compared to all investigated vowels, being identified at a significantly higher rate than all vowels. Pairwise comparisons of mean identifications showed the followed pattern in an ascending order: /ɔː/ (*MD* = .09, $p < .01$), /ɑː/ (*MD* = .11, $p < .01$), /ɜː/ (*MD* = .12, $p < .01$), /uː/ (*MD* = .33, $p < .000$), /ʊ/ (*MD* = .33, $p < .000$), /ɪ/ (*MD* = .35, $p < .000$), /e/ (*MD* = .36, $p < .000$), /ʌ/ (*MD* = .44, $p < .01$) and /æ/ (*MD* = .53, $p < .01$). The largest difference in mean identification was found between /iː/ and /ɒ/ (*MD* = .09, $p < .01$). Further pairwise comparisons demonstrated participants' substantial increased sensitivity to identification of long vowels against short ones.

Two other remarkable cases were observed with the vowel /ɒ/ and /æ/. The vowel /ɒ/ was identified at a significantly lower rate than all vowels, except for /æ/ (*MD* = 10, $p = .08$). Significant differences were found with vowels /ɪ/ (*MD* = -.28, $p < .000$), /ʌ/ (*MD* = -.19, $p < .000$), /ɑː/ (*MD* = -.52, $p < .000$), /ɔː/ (*MD* = -.54, $p < .000$), /ʊ/ (*MD* = -.30, $p < .000$), /uː/ (*MD* = -.30, $p < .000$), /ɜː/ (*MD* = -.51, $p < .000$) and /e/ (*MD* = -.27, $p < .000$). The vowel /æ/ was similarly identified at a significantly lower rate than /iː/ (*MD* = -.35, $p < .000$), /ɪ/ (*MD* = -.18, $p < .05$), /ɑː/ (*MD* = -.42, $p < .000$), /ɔː/ (*MD* = -.44, $p < .000$), /ʊ/ (*MD* = -.20, $p < .000$), /uː/ (*MD* = -.20, $p < .000$), /ɜː/ (*MD* = -.41, $p < .000$) and /e/ (*MD* = -.17, $p < .01$). However, the vowel /æ/ was not significantly identified at different rates compared to /ʌ/ (*MD* = -09, $p = 1$) and /ɒ/ (*MD* = .10, $p = .08$).

Table 98

Within-Subjects Effects for Aggregate Mean Identification and Reaction Time of All Vowels Combined

Source	Measure		*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Vowel	Identification	Sphericity Assumed	10	64.969	.000	.555	1.000
		Greenhouse-Geisser	5.617	64.969	.000	.555	1.000
		Huynh-Feldt	6.375	64.969	.000	.555	1.000
		Lower-bound	1	64.969	.000	.555	1.000
	Reaction time	Sphericity Assumed	10	9.051	.000	.148	1.000
		Greenhouse-Geisser	7.362	9.051	.000	.148	1.000
		Huynh-Feldt	8.696	9.051	.000	.148	1.000
		Lower-bound	1.000	9.051	.004	.148	.840

[a]. Computed using alpha = .05

Table 99

Pairwise Comparisons of Aggregate Mean Identification of All Vowels Combined

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/iː/	/ɪ/	.35*	.000
	/æ/	.53*	.000
	/ʌ/	.44*	.000
	/ɑː/	.11*	.001
	/ɒ/	.63*	.000
	/ɔː/	.09*	.007
	/ʊ/	.33*	.000
	/uː/	.33*	.000
	/ɜː/	.12*	.003
	/e/	.36*	.000
/ɪ/	/æ/	.18*	.011
	/ʌ/	.09	1.000
	/ɑː/	-.24*	.000
	/ɒ/	.28*	.000
	/ɔː/	-.26*	.000
	/ʊ/	-.02	1.000
	/uː/	-.02	1.000
	/ɜː/	-.23*	.001
	/e/	.01	1.000
/æ/	/ʌ/	-.09	1.000
	/ɑː/	-.42*	.000
	/ɒ/	.10	.081
	/ɔː/	-.44*	.000
	/ʊ/	-.20*	.000
	/uː/	-.20*	.000
	/ɜː/	-.41*	.000
	/e/	-.17*	.001
/ʌ/	/ɑː/	-.33*	.000
	/ɒ/	.19*	.000
	/ɔː/	-.35*	.000
	/ʊ/	-.11	.520
	/uː/	-.11	.595
	/ɜː/	-.32*	.000
	/e/	-.08	1.000

Note. Pairwise comparisons are based on estimated marginal means.

*. The mean difference is significant at the .05 level.

[b]. Adjustment for multiple comparisons: Bonferroni.

Table 99 *(continued)*

Pairwise Comparisons of Aggregate Mean Identification of All Vowels Combined

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/ɑː/	/ɒ/	.52*	.000
	/ɔː/	-.02	1.000
	/ʊ/	.22*	.000
	/uː/	.22*	.000
	/ɜː/	.01	1.000
	/e/	.25*	.000
/ɒ/	/ɔː/	-.54*	.000
	/ʊ/	-.30*	.000
	/uː/	-.30*	.000
	/ɜː/	-.51*	.000
	/e/	-.27*	.000
/ɔː/	/ʊ/	.24*	.000
	/uː/	.24*	.000
	/ɜː/	.03	1.000
	/e/	.27*	.000
/ʊ/	/uː/	.00	1.000
	/ɜː/	-.21*	.000
	/e/	.03	1.000
/uː/	/ɜː/	-.21*	.000
	/e/	.03	1.000
/ɜː/	/e/	.24*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

Analogously with the previous analysis, a one-way repeated measures ANOVA with the lower-bound correction showed a significant effect of vowel on reaction time, yielding an *F* ratio of $F(1, 52) = 9.051$, $p < .005$, $\eta^2 = .148$ (see Table 98 above). Results from Table 100 below demonstrated interesting results regarding the vowels /ɔː/, /iː/ and /ɑː/. The former vowel was identified at a significantly the shortest reaction time compared to all vowels, followed by /iː/, which was identified at a longer reaction time than /ɔː/ (*MD* = .28, *p* = 1), and /ɑː/ that was identified at a longer reaction time than /ɔː/ and /iː/ (*MD* = .58, *p* = .126 and *MD* = .20, *p* = 1, respectively). These findings demonstrated a double perceptual advantage for these vowels in terms of identification rate and reaction time.

Table 100

Pairwise Comparisons of Aggregate Mean Reaction Time of All Vowels Combined

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* ^{b}
/iː/	/ɪ/	-1.20*	.009
	/æ/	-.47	1.000
	/ʌ/	-.75	.295
	/ɑː/	-.20	1.000
	/ɒ/	-.62	1.000
	/ɔː/	.38	1.000
	/ʊ/	-.83	.798
	/uː/	-1.07*	.003
	/ɜː/	-1.30*	.000
	/e/	-1.33*	.005
/ɪ/	/æ/	.55	1.000
	/ʌ/	.27	1.000
	/ɑː/	.82*	.011
	/ɒ/	.40	1.000
	/ɔː/	1.40*	.000
	/ʊ/	.89	1.000
	/uː/	-.05	1.000
	/ɜː/	-.28	1.000
	/e/	-.31	1.000
/æ/	/ʌ/	-.28	1.000
	/ɑː/	.28	1.000
	/ɒ/	-.15	1.000
	/ɔː/	.85*	.001
	/ʊ/	-.36	1.000
	/uː/	.60	.532
	/ɜː/	-.83*	.003
	/e/	-.86	.122
/ʌ/	/ɑː/	.55	.683
	/ɒ/	.13	1.000
	/ɔː/	1.13*	.001
	/ʊ/	-.08	1.000
	/uː/	-.32	1.000
	/ɜː/	-.55	1.000
	/e/	-.58	1.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
^{b}. Adjustment for multiple comparisons: Bonferroni.

Table 100 *(continued)*

Pairwise Comparisons of Aggregate Mean Reaction Time of All Vowels Combined

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/ɑː/	/ɒ/	-.42	1.000
	/ɔː/	.58	.126
	/ʊ/	-.64	1.000
	/uː/	-.87*	.031
	/ɜː/	-1.10*	.001
	/e/	-1.14*	.011
/ɒ/	/ɔː/	1.00*	.007
	/ʊ/	-.21	1.000
	/uː/	-.45	1.000
	/ɜː/	-.68	.207
	/e/	-.71	1.000
/ɔː/	/ʊ/	-1.22*	.000
	/uː/	-1.45*	.000
	/ɜː/	-1.68*	.000
	/e/	-1.71*	.000
/ʊ/	/uː/	-.24	1.000
	/ɜː/	-.47	1.000
	/e/	-.50	1.000
/uː/	/ɜː/	-.23	1.000
	/e/	-.26	1.000
/ɜː/	/e/	-.03	1.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

Conversely, the vowels /uː/ and /ɜː/ showed a different pattern of reaction time compared to other vowels. The former vowel was identified at a shorter reaction time than /e/ (*MD* = -.03, *p* = 1), /ɜː/ (MD = -23, *p* = 1) and /æ/ (*MD* = .60, *p* = .532), while the latter vowel was identified at a shorter reaction time than only /e/ (*MD* = -.03, *p* = 1). The perceptual advantage for these vowels at identification did not manifest in reaction time in a similar pattern to the other long vowels, suggesting a likely difficulty in identifying the vowels at hand. Interestingly however, the vowel /æ/, one of the least accurately identified vowels, proved to be identified at shorter reaction times compared to several vowels such as /ʌ/ (*MD* = -.28, *p* = 1), /ʊ/ (*MD* = -.36, *p* = 1), /ɒ/ (*MD* = -.15, *p* = 1), /ɜː/ (*MD* = -.83, *p* < *.01*), /ɪ/ (*MD* = -.55, *p* = 1) and /e/ (*MD* = -.86, *p* = 12).

6.3.4.4 Variation in identification among proficiency groups.

Isolated vowel tokens identification shows a random variation among proficiency groups and among vowels, with poor-proficiency participants identifying the vowels /iː/, /ɪ/, /ɒ/ and /uː/at a higher rate than high-proficiency participants and vice versa for /æ/, /ʊ/ and /ɜː/.

Table 101

Descriptive Statistics of Isolated Vowel Tokens Identification between Proficiency Groups

					95% CI	
Vowel	Proficiency group	*M*	*SD*	*SE*	Lower bound	Upper bound
/iː/	Poor	.72	.10	.02	.67	.77
	Mediocre	.69	.11	.02	.64	.75
	High	.66	.21	.06	.52	.79
/ɪ/	Poor	.38	.29	.07	.24	.52
	Mediocre	.35	.30	.06	.22	.49
	High	.27	.34	.10	.05	.48
/æ/	Poor	.12	.12	.03	.06	.18
	Mediocre	.19	.18	.04	.11	.27
	High	.19	.16	.05	.09	.30
/ʌ/	Poor	.23	.20	.05	.14	.33
	Mediocre	.29	.26	.06	.18	.41
	High	.23	.16	.05	.13	.33
/ɑː/	Poor	.57	.09	.02	.53	.61
	Mediocre	.60	.10	.02	.55	.64
	High	.58	.10	.03	.51	.64
/ɒ/	Poor	.08	.12	.03	.02	.13
	Mediocre	.06	.13	.03	.00	.12
	High	.07	.16	.05	-.03	.17
/ɔː/	Poor	.60	.10	.02	.55	.65
	Mediocre	.61	.11	.02	.56	.66
	High	.59	.11	.03	.52	.67
/ʊ/	Poor	.40	.20	.04	.31	.49
	Mediocre	.31	.23	.05	.21	.41
	High	.41	.20	.06	.28	.54

Note. N = 53

Table 101 (*continued*)

Descriptive Statistics of Isolated Vowel Tokens Identification between Proficiency Groups

					95% CI	
Vowel	Proficiency group	*M*	*SD*	*SE*	Lower bound	Upper bound
/uː/	Poor	.40	.18	.04	.32	.49
	Mediocre	.36	.19	.04	.27	.45
	High	.31	.15	.04	.21	.40
/ɜː/	Poor	.55	.16	.04	.48	.63
	Mediocre	.58	.17	.04	.50	.65
	High	.62	.09	.03	.56	.68
/e/	Poor	.38	.20	.05	.28	.47
	Mediocre	.28	.17	.04	.21	.36
	High	.35	.22	.06	.21	.49

Note. N = 53

As displayed in Table 102 below, there were statistically non-significant differences between proficiency groups as determined by one-way between subjects ANOVA across all vowels, with $F(2,50)$ ranging from 0.10 to 1.33 and p from .90 to .27, respectively. Based on those findings, there is much reason to suggest that speaking proficiency has no significant effect on isolated vowel tokens identification.

Table 102

Variation in Identification of Isolated Vowel Tokens between Proficiency Groups

Vowel	*F*	*Sig.*	ES η^2
/iː/	0.79	.46	.03
/ɪ/	0.53	.59	.02
/æ/	1.33	.27	.05
/ʌ/	0.54	.59	.02
/ɑː/	0.33	.72	.01
/ɒ/	0.10	.90	.00
/ɔː/	0.16	.85	.01
/ʊ/	1.28	.29	.05
/uː/	1.11	.34	.04
/ɜː/	0.78	.46	.03
/e/	1.31	.28	.05

Note. N= 53
df (between groups) = 2
df (within groups) = 50

6.3.4.5 Isolated vowel tokens identification and proficiency.

We performed Pearson product-moment correlation coefficient to assess relationship between isolated vowel tokens identification and speaking proficiency. As displayed in Table 58 below, there were no significant correlations between the two variables. Overall findings suggest that identification of isolated vowel tokens, as manipulated in various temporal conditions, does not seem to increase or decrease systematically as a function of participants' speaking proficiency. It is to note that non-significant correlations between isolated vowel tokens identifications and participants' proficiency come in line with previous findings regarding variation in identification across speaking proficiency groups, suggesting that perceptual and productive abilities may be unrelated.

Although secondary to our analysis, it is to note that some inter-synthetic vowel tokens identifications were significantly correlated, positively and negatively as well. There were significant, positive correlations between identification of (a) /ɪ/ and /iː/, r = .274, 2-tailed $p < .05$; (b) /ɔː/ and /ɑː/, r = .524, 2-tailed $p < .01$; (c) /uː/ and /iː/, r = .288, 2-tailed $p < .05$; (d) /uː/ and /ʊ/, r = .565, 2-tailed $p < .01$ and (e) /ɜː/ and /uː/, r = .459, 2-tailed $p < .01$. However, there were two significant, negative correlations between identification of /ʌ/ and /æ/ (r = -.424, 2-tailed $p < .01$) and /ɔː/ and /ʌ/ (r = -.287, $p < .05$). Taken together, findings suggest that long vowel identification correlate positively significantly with each other.

Table 103

Matrix of Correlations among Isolated Vowel Tokens Identification and Speaking Proficiency

	/iː/	/ɪ/	/æ/	/ʌ/	/ɑː/	/ɒ/	/ɔː/	/ʊ/	/uː/	/ɜː/	/e/	Proficiency
/iː/	1	**.274***	-.038	-.152	-.014	.053	.153	.178	**.288***	.087	.011	-.134
/ɪ/		1	.099	-.222	-.129	.138	.061	-.137	-.041	-.123	.113	-.246
/æ/			1	**-.424****	.206	-.002	.190	-.070	-.017	.004	-.105	.171
/ʌ/				1	-.050	.018	**-.287***	.001	-.141	-.243	-.163	.078
/ɑː/					1	-.018	**.524****	-.155	.084	.248	-.095	.097
/ɒ/						1	.183	-.178	-.089	-.099	.087	.040
/ɔː/							1	-.061	.090	.267	-.039	.066
/ʊ/								1	**.565****	.267	.172	.010
/uː/									1	**.459****	.189	-.235
/ɜː/										1	.228	.181
/e/											1	-.033
Proficiency												1

Note. N = 53
**. Correlation significant at the 0.01 level (2-tailed)
*. Correlation significant at the 0.05 level (2-tailed)

Chapter 7

Test of Vowel Identification in Context

Introduction

Vowel identification in context is an m-alternative forced choice test (*m-AFC*). It was guided by the principle that monophthongs' identification would significantly vary when the monophthongs occur within a contextual constraints or simply a meaningful context. Phonetic contexts and constraints were controlled by specifying the number of syllables of the selected words, one- and two-syllable words, and the occurrence of the targeted vowel in various phonetic contexts, that is, followed by a voiceless, voiced, or nasal consonant.

We designed test of vowel identification in context to

a) Investigate whether identification of vowels would improve as a function of interaction between vowel and consonantal context, and

b) Find out phonetic contexts for optimal identification

Some comparisons between certain monophthongs were included more than once in order to make sure the presence of more irrelevant choices would not affect the participants' responses, and that sensitivity would remain significantly the same (Luce, 1977). Moreover, this was designed to make participants keep focus on the task and not develop any response strategy.

7.1 Stimuli Materials

We used PRAAT Speech Synthesizer, set at RP English and a normal male voice variant (m7) to synthesise words. Stimuli in test of vowel identification in context consisted of 89 word tokens (see Flowchart to test of in-context vowel identification for conception of detailed phonetic consonantal contexts and Appendix 10 for the list of word tokens). Word tokens were:

a) Sixty eight one-syllable words, and

b) Twenty three two-syllable words

Word tokens contained various RP English vowels occurring in different phonetic contexts. All word tokens were saved in 32-bit WAV files for test administration.

7.2 Procedure

The consisted of 21 block of trials, with a changing number of trials per each as a function of word number in the series. Trials per block ranged from 20 to 50. By the time participants took this test, we did not see the need to insert a familiarisation block of trials, for they were sufficiently acquainted with test conditions and environment previously. Each trial began with the presentation of an alert message with task heading and instructions on the computer screen. Participants listened to one word token in each trial and had to identify the heard word among suggested choices. Every trial was preceded by a 150-ms initial silence and was followed by no final silence.

There were 21 series of suggested choices, corresponding to presented words in every block of trials. Participants had to respond simply by clicking on one of the buttons on the PC screen, with the phonemic transcription of the word on it. There was no logical order of the presented words on the PC screen and word tokens were computer randomised during administration. Upon completion of every block of trials, the *TP* software provided participants with immediate feedback and alerted them to their performance level for accuracy and reaction time. Participants had a systematic break after completion of every block of trials, during which we tried to alleviate the test pressure that would had built up, by encouraging them to report on any issue relating to the test. Though participants were encouraged to break enough before resuming the test of their own volition and not to respond quickly, they reported the identification test to have been a lot much easier than the two previous ones. Participants took about 2 hours to complete the test. Upon completion of the test, we collected test auto-saved result files of participants' data from their respective workstations.

Number of Stimuli	Phonetic constraints								
	1-Syllable words			2-Syllable words			1 & 2-Syllable words		
	Vowels compared	Series	Phonetic context (syllable)	Vowels compared	Series	Phonetic context (syllable)	Vowels compared	Series	Phonetic context (syllable)
2				/ʌ, ɒ/	Series 20	final			
3	/ʌ, æ, ɑː/	Series 6	+ voiceless						
4	/æ, ɑː/	Series 2	+ voiced + voiceless	/æ, ɔː, ʌ, ɑː/	Series 3	final	/ɪ, iː/	Series 8	+ voiceless final
	/æ, ɑː/	Series 4	+ voiced	/æ, ɒ/	Series 21	final	/e, ɜː/	Series 11	+ nasal, final
	/ɪ, iː/	Series 7	+ nasal + voiceless				/ʊ, uː/	Series 13	+ voiceless + voiced
	/æ, ɑː, ʌ, ɒ/	Series 14	+ voiced				/æ, ɑː, ɒ/	Series 16	+ nasal, final
	/ɒ, ʌ/	Series 19	+ voiceless + nasal				/æ, ʌ, ɒ/	Series 18	+ voiced, final
5	/æ, ɔː, ʌ, ɑː, ɒ/	Series 1	+ nasal	/æ, ʌ, ɑː/	Series 5	final			
	/æ, ʌ, ɑː, ɒ, ɔː/	Series 9	+ voiced						
	/æ, ɔː, ʌ, ɑː, ɒ/	Series 10	+ voiceless						
	/ɒ, ɑː, ɔː/	Series 15	+ voiced + nasal						
	/æ, ʌ, ɔː, ɑː/	Series 17	+ voiceless						
6							/ʊ, ɪ/	Series 12	+ voiceless + voiced

Figure 29. Flowchart to Experiment 3. Series numbers are reported as administered during testing sessions. Some vowels were experimented on simultaneously in the same series (*block of trials*), to ensure optimal time use while maintaining unbiased response accuracy. Some vowels were experimented on more than once on the same variable for practical reasons related to adopted methodology. Some vowel pairs were experimented on only in same different vowel discrimination test and not in test of vowel identification in context due to lack of adequate vocabulary items that meet experimentation objectives. Vocabulary Series 12 was used as a *break* bloc. The symbol + indicates vowel followed by the described sound.

7.3 Results

This section presents detailed sensitivity measurements for tests of vowel identification in context (*m-AFC)* for all word token series and their respective confusion matrices. Reaction time analysis was not included, for all tests' results showed no statistical significance.

7.3.1 Sensitivity measurements.

7.3.1.1 Vocabulary Series 1.

As displayed in *Figure 30*, participants demonstrated a considerable sensitivity towards the long vowels /ɔː/ and /ɑː/, identified with a pooled sensitivity index *d'* of 2.19 and 1.66, respectively. The vowel /æ/ was also perceived with a higher sensitivity index *d'* of 1.39 than the two remaining vowels /ɒ/ and /ʌ/. These initial results of vowel identification in context may suggest higher sensitivity to long vowels compared to short ones. It is possible to suggest that full vowel lengthening of short vowels before a voiced consonant might have caused participants to identify them as long ones, regardless of differences in their acoustic features.

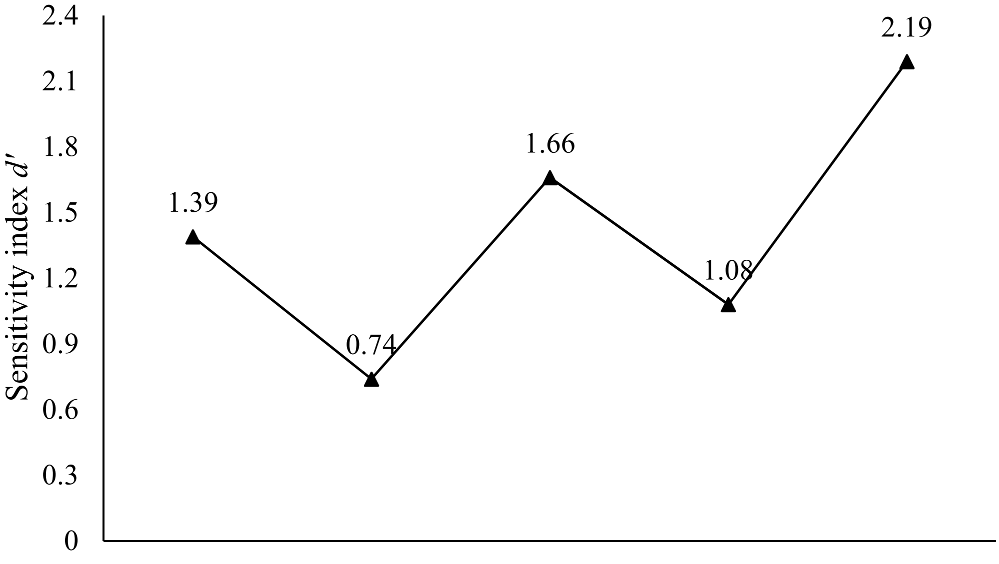


Figure 30. Identification of vowels in synthetic word tokens
Series 1. N = 53

7.3.1.2 Vocabulary Series 2.

Figure 31 reveals interesting findings with the respect to identification of the vowels /ɑː/ and /æ/ when followed by a voiced or voiceless consonant. With few stimuli presented and varying phonetic context, participants showed identical perceptual sensitivity to vowels /ɑː/ and /æ/ when followed by a voiced consonant. However, when followed by voiceless consonant, the vowel /æ/ was identified with a higher sensitivity *d'* of 2.45 compared to the vowel /ɑː/ that was identified with a sensitivity index *d'* of 1.6. Vowel shortening or simply pre-fortis clipping might have caused participants to identify the vowel /ɑː/ with less perceptual sensitivity, irrespective of its inherent acoustic qualities.

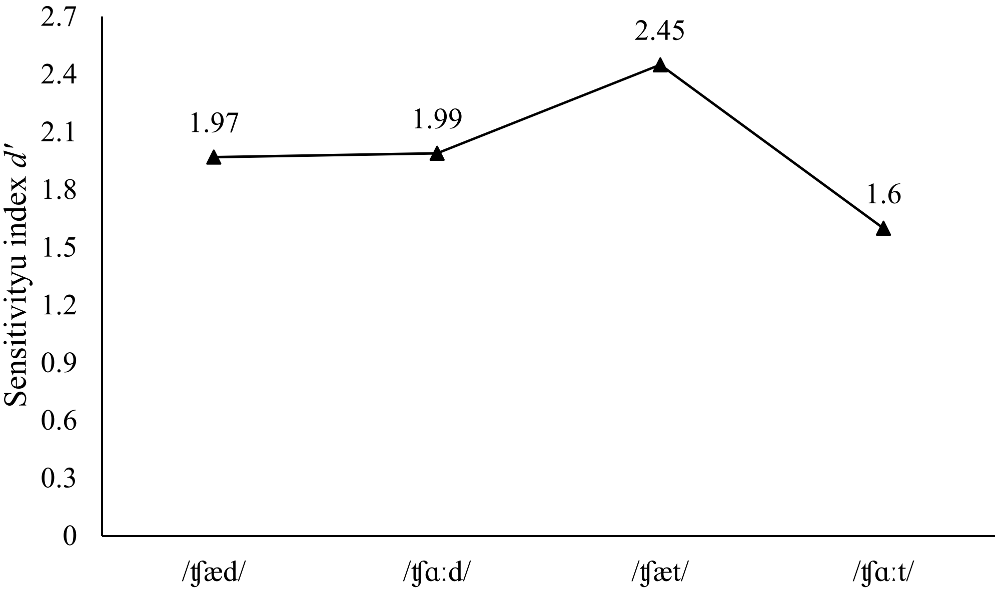


Figure 31. Identification of vowels in synthetic word tokens Series 2. N = 53

7.3.1.3 Vocabulary Series 3.

Figure 32 below confirms the initially observed phenomenon of higher perceptual sensitivity in identifying long vowel compared to short ones. The vowel /ɔː/ was identified with a high perceptual sensitivity index *d'* of 3.22, followed by 1.57 for /ɑː/, 1.15 for /æ/, and 0.68 for /ʌ/. It seems that in the absence of several stimuli presentation and regardless of phonetic context, perceptual sensitivity in identifying long vowels is higher than for short ones.

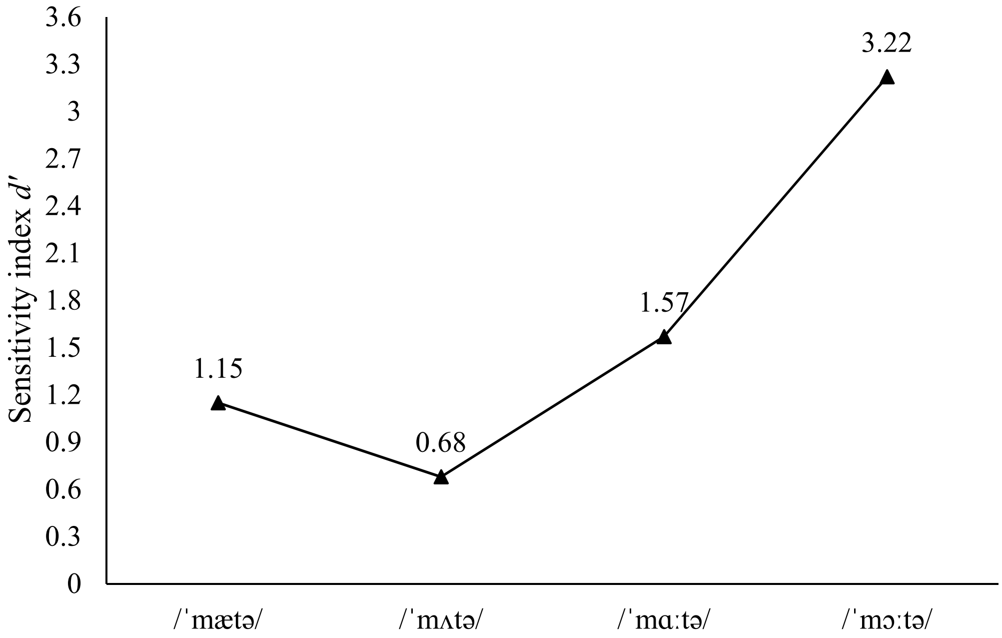


Figure 32. Identification of vowels in synthetic word tokens
Series 3. N = 53

7.3.1.4 Vocabulary Series 4.

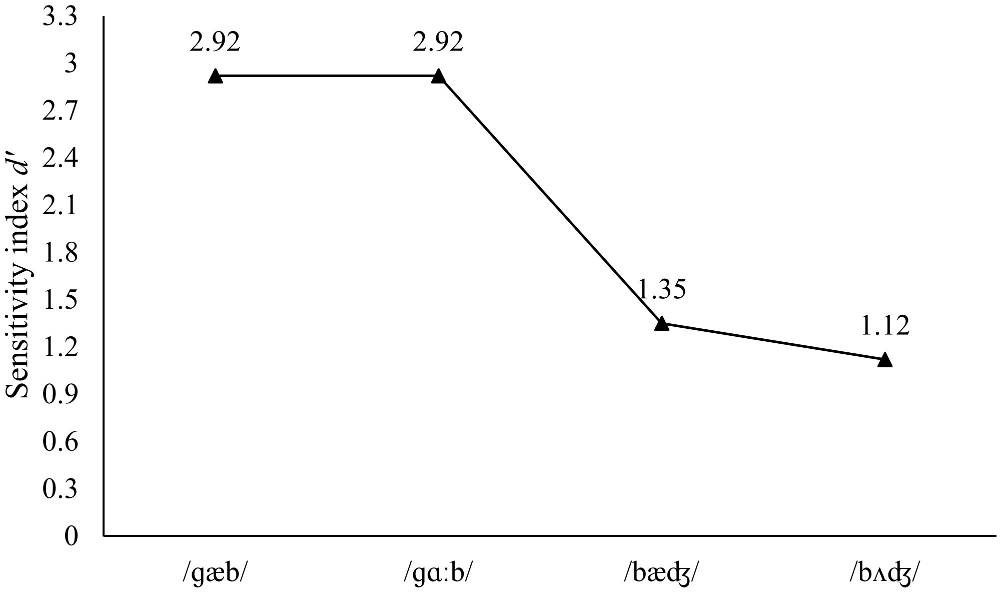


Figure 33. Identification of vowels in synthetic word tokens
Series 4. N = 53

Again, the absence of several similar stimuli presentation and the specific context of vowel occurrence likely caused learners considerable perceptual sensitivity to both vowels /æ/

and /ɑː/ that were identified with an equal sensitivity index *d'* of 2.92. This result is very similar to the one obtained in word token Series 2. A similar case of relatively equal perceptual sensitivity was observed with the vowels /æ/ and /ʌ/ that were identified with slightly close sensitivity indices, 1.35 and 1.12, respectively.

7.3.1.5 Vocabulary Series 5.

The same pattern of perceptual sensitivity found above was observed here. The presence of several stimuli caused the participants to rely more on the length of the vowel to identify it. Though occurring before a voiceless consonant, the long vowel /ɑː/ was perceived with a high sensibility index *d'* of 3.19, followed by 1.33 for /ʌ/, 1.27 for /æ/ in the context /b.tə/. Additionally, participants showed a consistently higher perceptual sensitivity to /æ/ than /ʌ/ in the context of /fl.tə/.

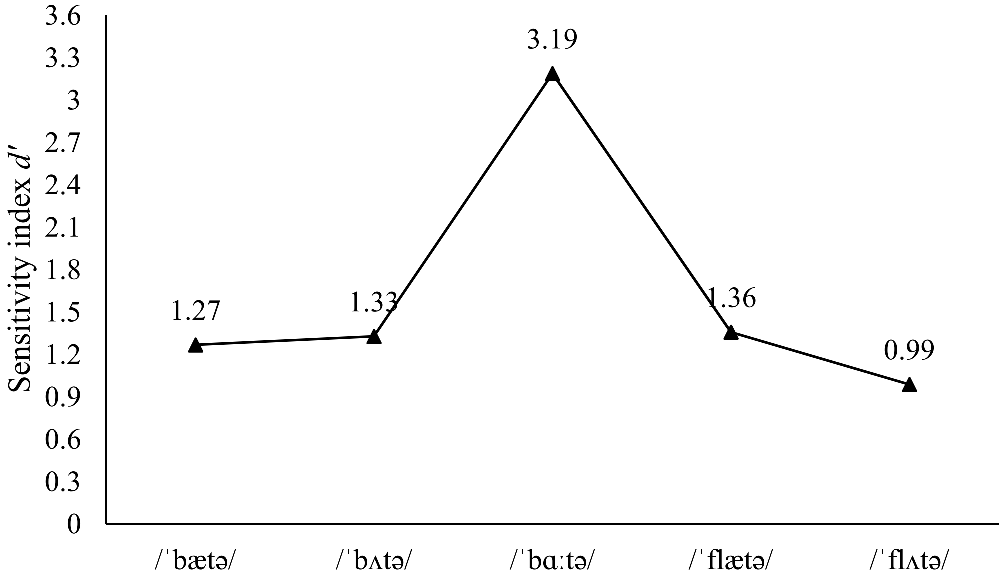


Figure 34. Identification of vowels in synthetic word tokens Series 5. N = 53

7.3.1.6 Vocabulary Series 6.

Figure 35 displays a further confirmation to the previously established pattern of perceptual sensitivity of participants in identifying the three vowels /æ/, /ʌ/, and /ɑː/. A higher sensitivity was reported for /ɑː/ that was identified with a sensitivity index *d'* of 2.16, followed by /æ/ with *d'* of 1.28, and finally /ʌ/ with *d'* of 0.82. Pre-fortis clipping of the tested vowels did not seem to have affected participants' sensitivity to the long vowel.

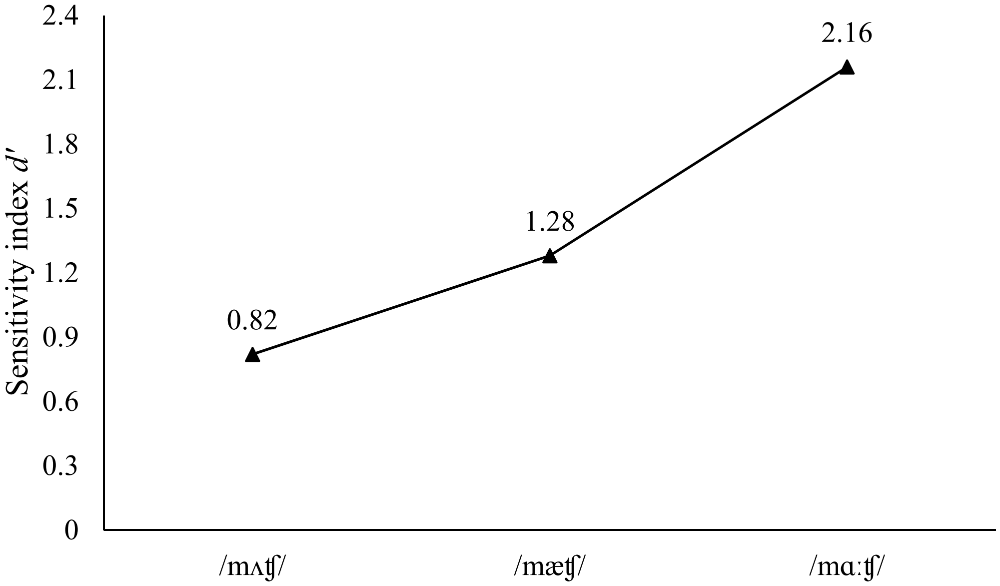


Figure 35. Identification of vowels in synthetic word tokens Series 6. N = 53

7.3.1.7 Vocabulary Series 7.

Figure 36 demonstrates quite a different story with the vowels /ɪ/ and /iː/, the former being identified with a higher sensitivity than the latter in the context of /b.n/ and a relatively equal sensitivity in the context of /.ʧ/. Participants demonstrated a perceptual advantage in identifying the vowel /ɪ/. This finding stands in opposition to the previous findings established with the vowels /æ/, /ɔː/, /ʌ/, /ɑː/, and /ɒ/. However, this finding may not be that surprising if we consider the perceptual advantage in identifying certain vowels along their temporal features to be a preference exhibited within a specific area on the vowel perceptual map and not the entire map. Alternatively, the findings may be simply accidental, and further findings herein will confirm or disconfirm either claim.

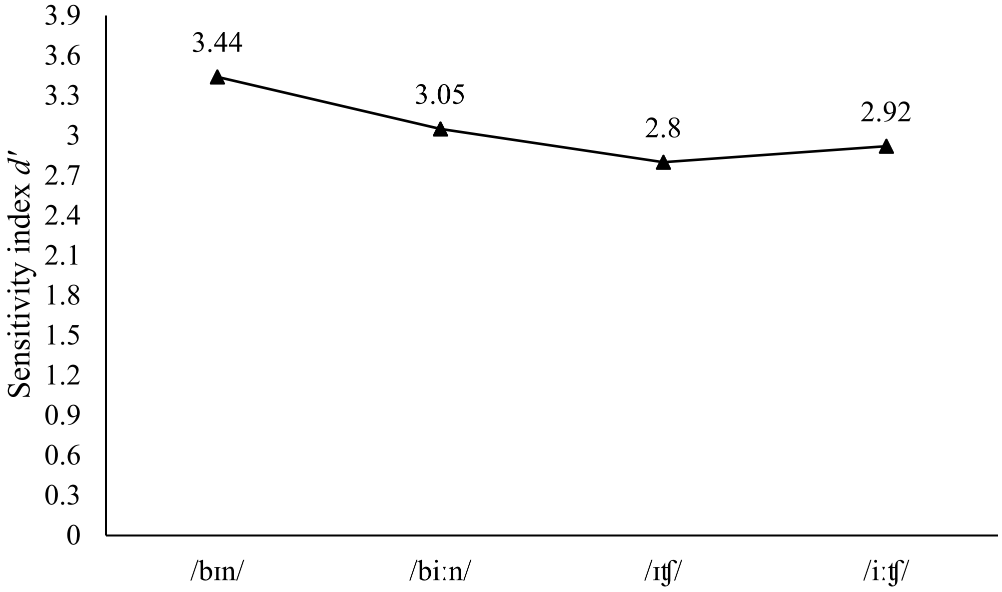


Figure 36. Identification of vowels in synthetic word tokens
Series 7. N = 53

7.3.1.8 Vocabulary Series 8.

As displayed below, *Figure 37* shows a similar pattern of previous findings regarding the relative perceptual advantage in identifying the vowel /ɪ/ compared to /iː/ in both contexts. Unfortunately, given the very few examples involving the present vowel pair in the vowel identification test in context, we could neither deny the first claim nor confirm the second. There were only 2 series involving the contrast /ɪ/ and /iː/, for our present was focused on vowels with which participants had major difficulties in discriminating and identifying in isolation across temporal manipulations in previous experiments.

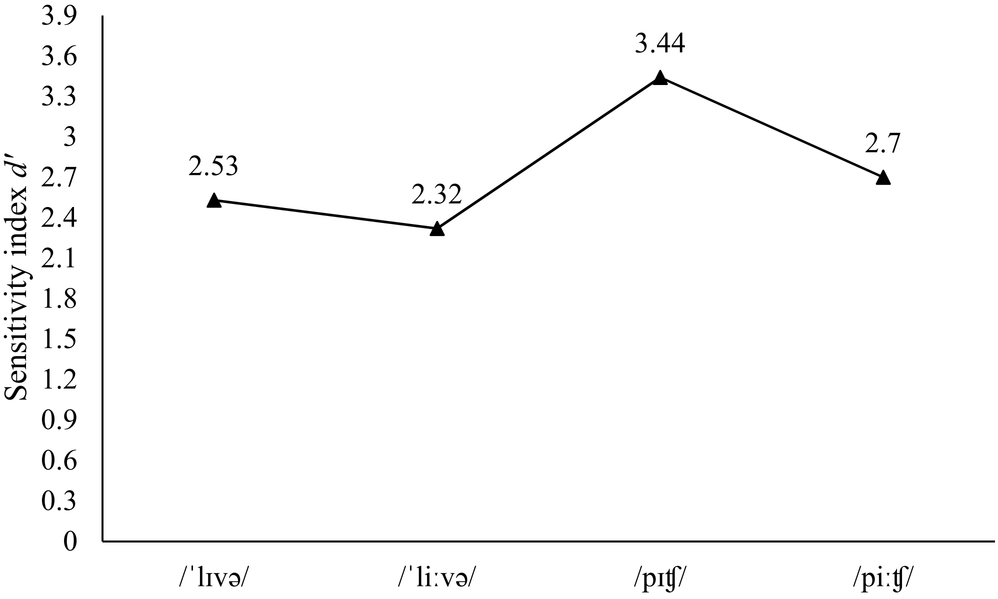


Figure 37. Identification of vowels in synthetic word tokens
Series 8. N = 53

7.3.1.9 Vocabulary Series 9.

Figure 38 below shows consistently the perceptual advantage in identification of long vowels /ɔː/ and /ɑː/, with a sensitivity index *d'* of 2.47 for the former and 2.53 for the latter. Among the short vowels, we find /æ/ identified with considerably a higher sensitivity (*d'* = 2.14) than /ɔ/and /ʌ/ identified with almost equal sensitivity indices, 0.96 and 0.93, respectively.

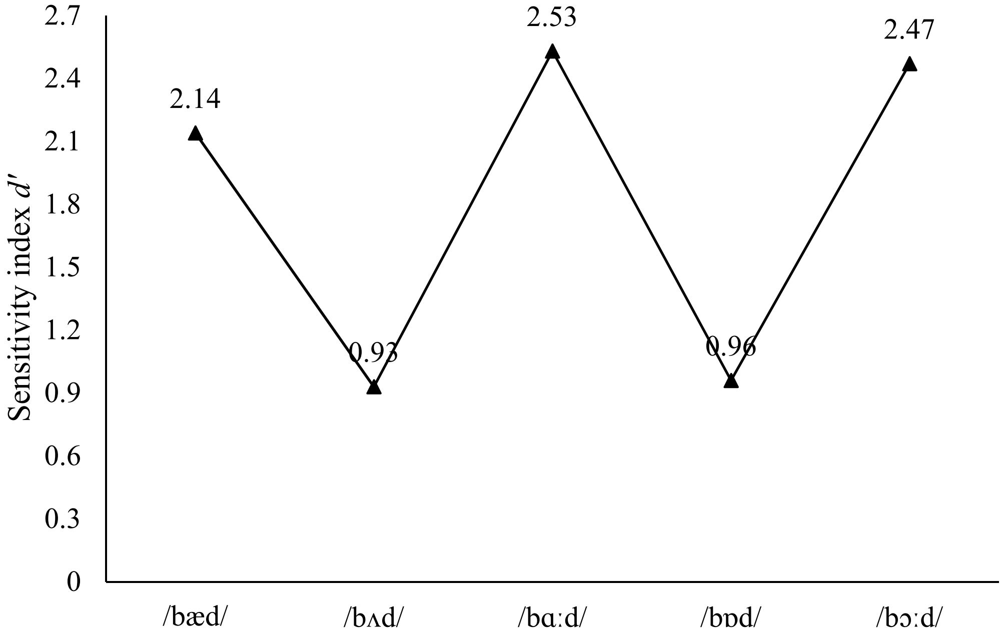


Figure 38. Identification of vowels in synthetic word tokens Series 9. N = 53

7.3.1.10 Vocabulary Series 10.

There was no surprise with this series with respect to long vowel identification with a considerable sensitivity index than the other vowels and following exactly the same pattern, /ɔː/ being the most accurately identified with *d'* of 2.47, followed first by /ɑː/ with *d'* of 2.05, and second by /æ/ with *d'* of 1.7. However, there was a slight change in the pattern of identification of /ʌ/ and /ɒ/, with the former being identified with fairly a higher sensitivity *d'* of 1.11 than the latter (*d'* = 0.99).

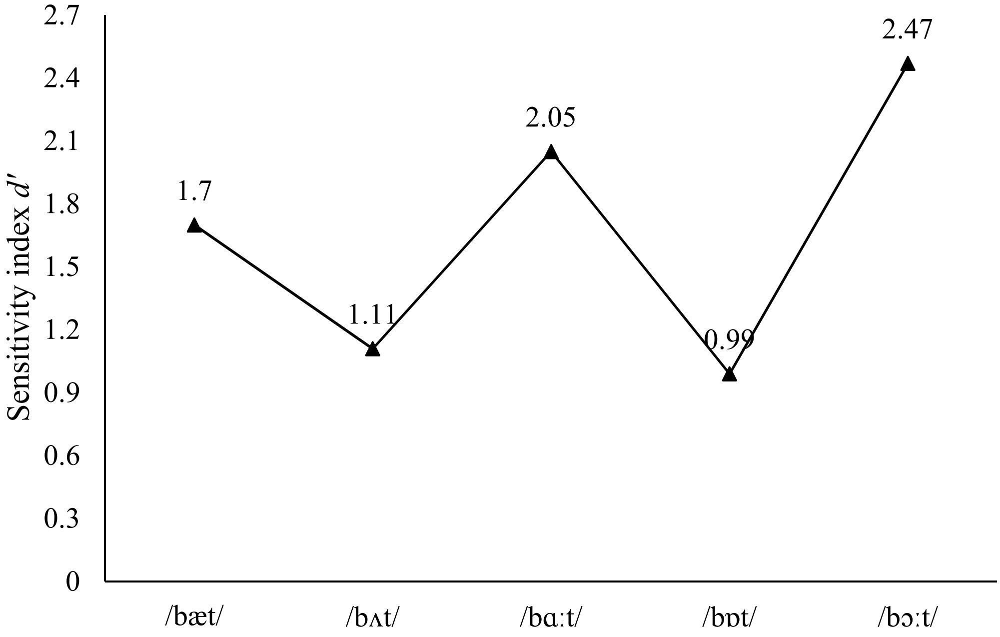


Figure 39. Identification of vowels in synthetic word tokens
Series 10. N = 53

7.3.1.11 Vocabulary Series 11.

Figure 46 demonstrates no specific pattern of identification of the vowels /e/ and /ɜː/ in context, with participants shifting in sensitivity across both vowels in the experimented on context. Unfortunately, there were no other instances of these vowel contrasts in context for further comparison to investigate any probable pattern of sensitivity.

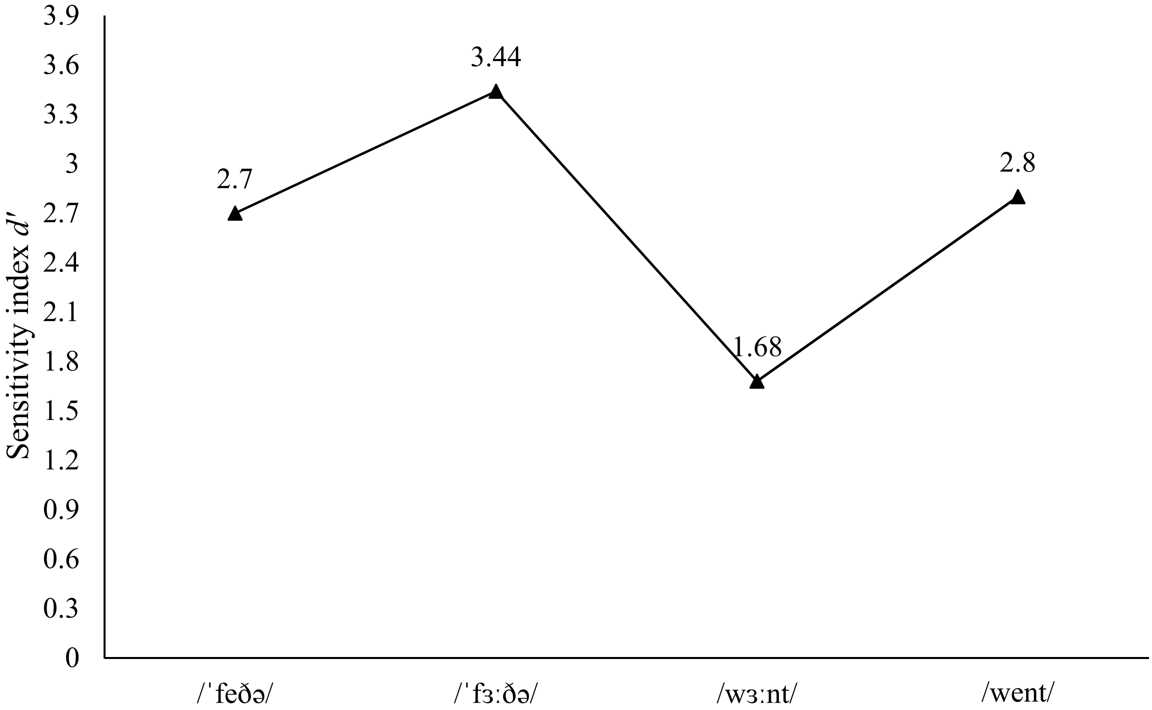


Figure 40. Identification of vowels in synthetic word tokens Series 11. N = 53

7.3.1.12 Vocabulary Series 12.

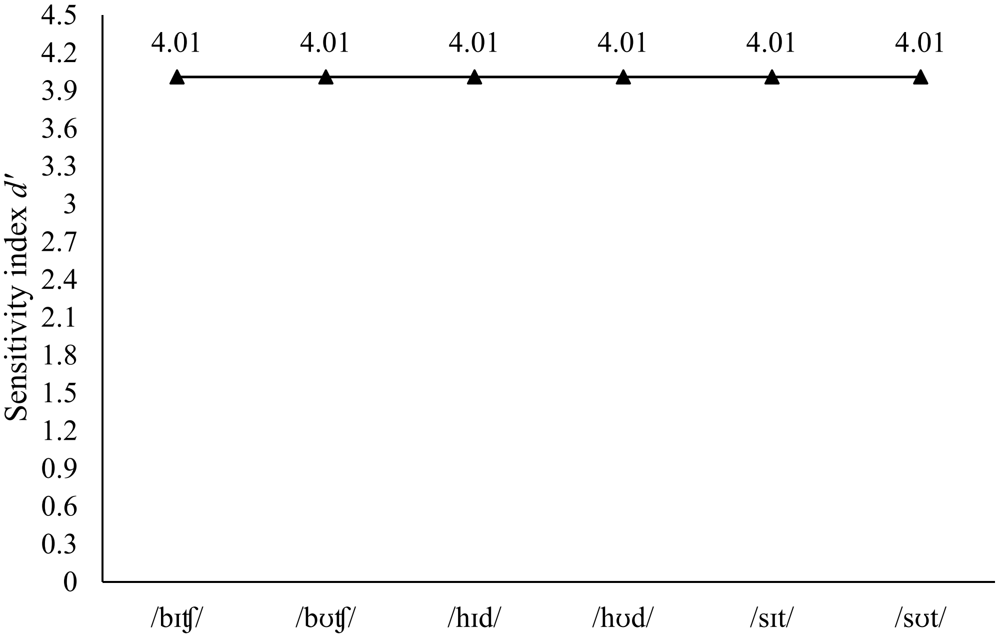


Figure 41. Identification of vowels in synthetic word tokens
Series 12. N = 53

The present Vocabulary Series 12 in *Figure 41* above was designed to make participants relax during the experiment and make sure that some misidentifications of the vowel /ɪ/ with /ʊ/ in previously described confusion matrices was just accidental. The obtained findings showed participants' high sensitivity to these contrasted vowels as reflected in one equally high sensitivity index *d'* of 4.01.

7.3.1.13 Vocabulary Series 13.

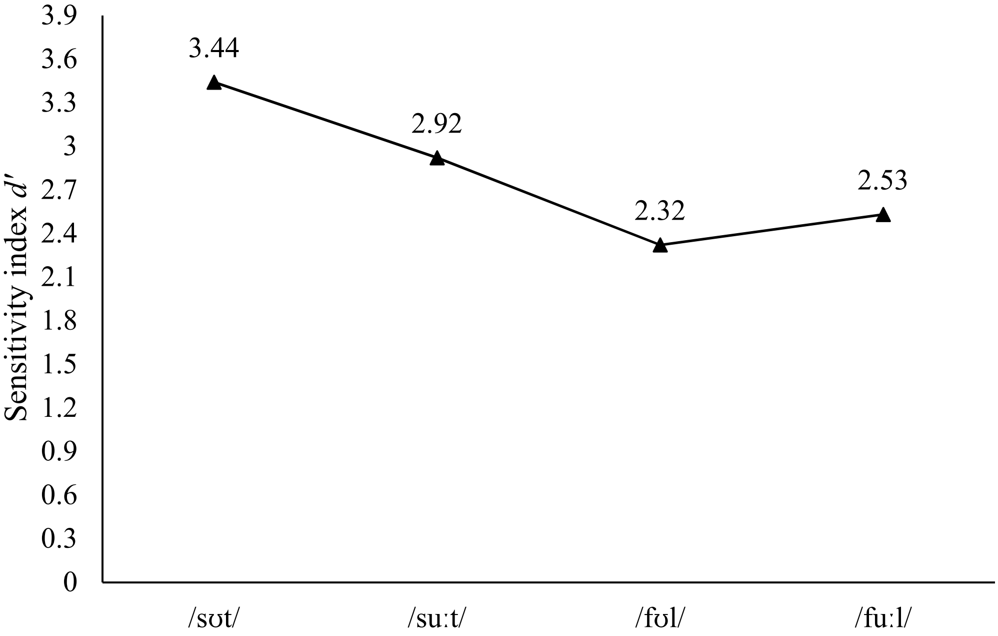


Figure 42. Identification of vowels in synthetic word tokens
Series 13. N = 53

Findings in *Figure 42* demonstrated further confirmation for participants' considerable sensitivity to the vowels /ʊ/ and /uː/ in context, with *d'* of 3.44 for /ʊ/ and 2.92 for /uː/ in the context of /s.t/, showing a higher sensitivity towards /ʊ/. However, in the context of /f.l/, participants showed higher sensitivity towards /uː/, with *d'* of 2.53 compared to 2.32 for /ʊ/.

7.3.1.14 Vocabulary Series 14.

Figure 43 below highlights again the decreasing sensitivity pattern participants showed towards the vowels /ɑː/, /æ/, /ʌ/, and /ɒ/, respectively. Participants showed a high sensitivity to the vowel /ɑː/, with a sensitivity index *d'* of 2.61, and least sensitivity to /ɒ/, with a sensitivity index *d'* of 0.74. So far, participants demonstrated systematic high sensitivity to both vowels /ɑː/ and /æ/, respectively.

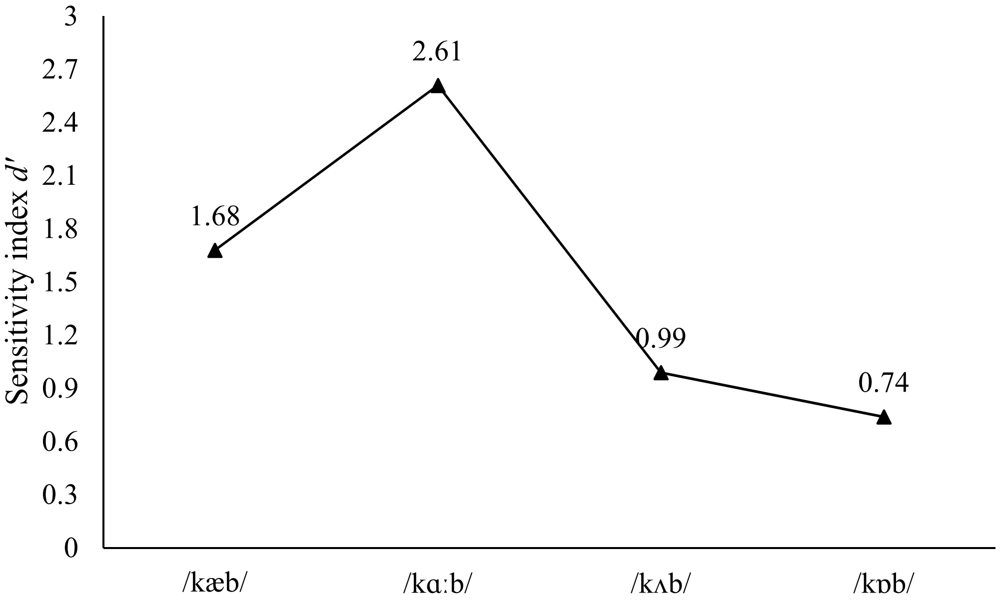


Figure 43. Identification of vowels in synthetic word tokens Series 14. N = 53

7.3.1.15 Vocabulary Series 15.

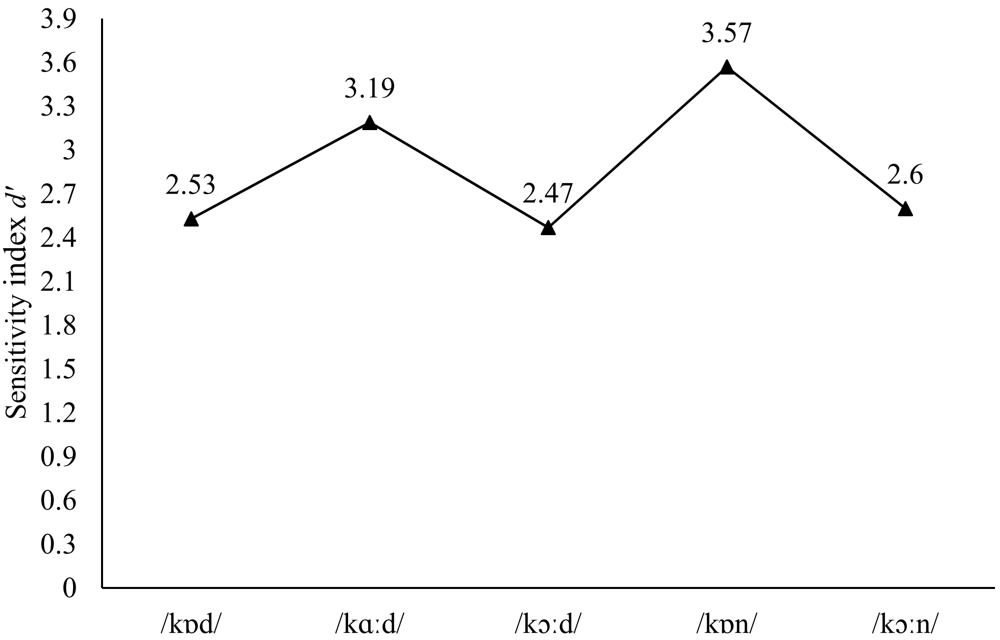


Figure 44. Identification of vowels in synthetic word tokens
Series 15. N = 53

In a rather surprising way to what was established before, *Figure 44* above demonstrates for the first time participants' high sensitivity to the vowel /ɒ/, with a sensitivity index *d'* of 3.57 compared to /ɔː/ (*d'* = 2.6) in the context of /k.n/. A similar result was observed in the context /k.d), except for the fact that participants were very sensitive to the vowel /ɑː/ in the context /k.d/. Based on previously confirmed pattern of sensitivity towards this set of vowels, the present finding is rather exceptional and may not be of a considerable significance.

7.3.1.16 Vocabulary Series 16.

As displayed in *Figure 45*, participants demonstrated a considerable sensitivity to all vowels in this series with indices ranging between 2.45 for /ɒ/ in the context /k.n/ and 3.05 for the same vowel in the /ˈh.tɪ/, with a more observed sensitivity of participants towards /æ/ in /k.n/. Demonstrating a bit of a different pattern of sensitivity from previous series, participants maintained the same high sensitivity to /æ/ than /ɒ/.

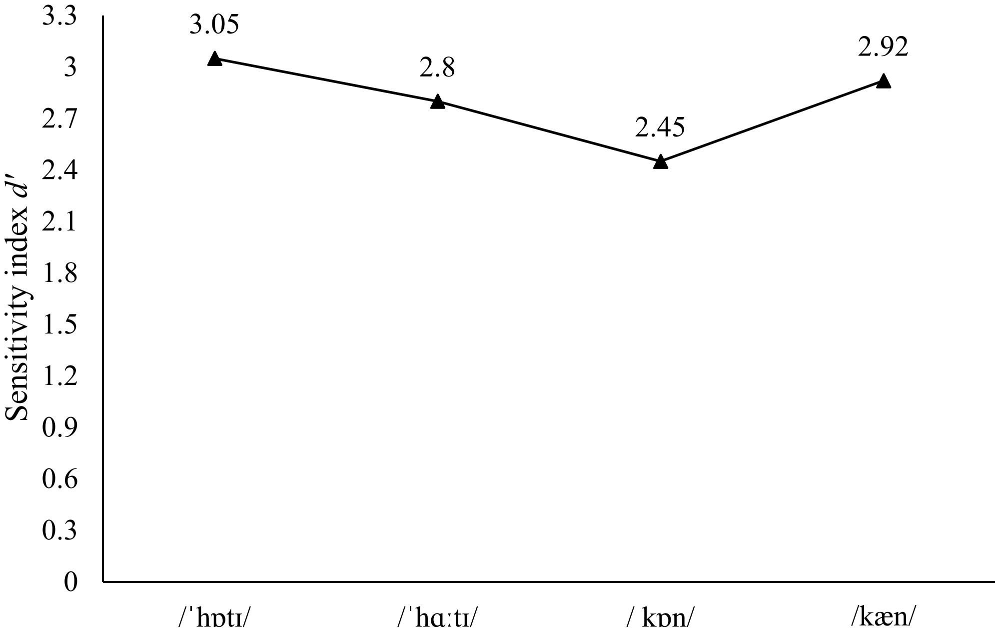


Figure 45. Identification of vowels in synthetic word tokens
Series 16. N = 53

7.3.1.17 Vocabulary Series 17.

Once more, participants showed their high sensitivity towards the vowels /ɑː/, /ɔː/, and /æ/ over /ʌ/, with a very high sensitivity index *d'* of 3.35 for /ɔː/ in the context /ʧ.k/ and 2.94 for /ɑː/ in the context /b.k/.

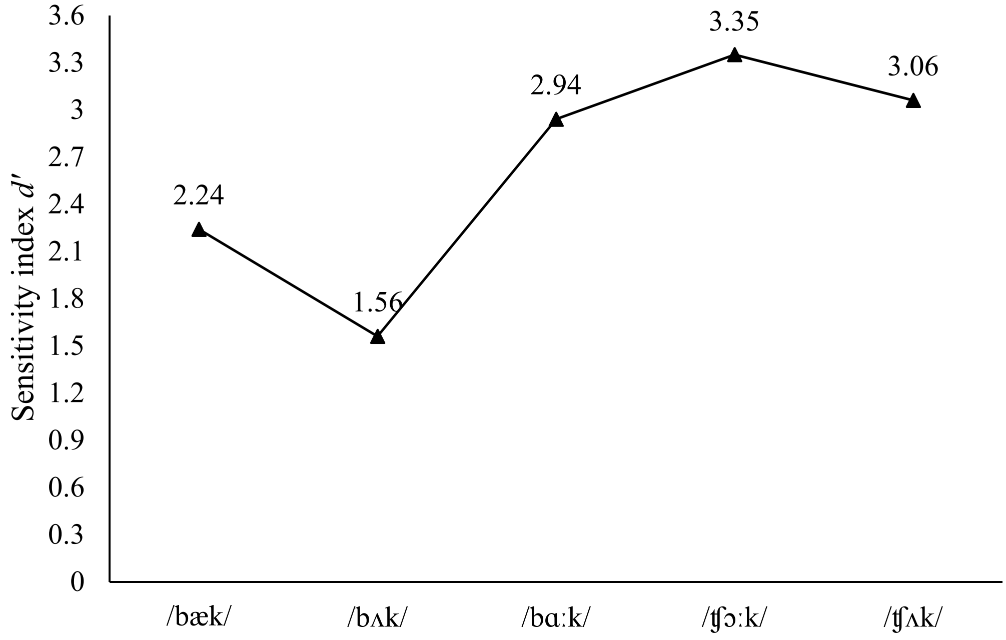


Figure 46. Identification of vowels in synthetic word tokens
Series 17. N = 53

7.3.1.18 Vocabulary Series 18.

Among the three vowels experimented on in this series, participants showed almost the same sensitivity to the vowels /æ/ and /ʌ/ in the context of /b.bl/, with a sensitivity index *d'* of 1.6. However, participants demonstrated fairly a relative higher sensitivity to the vowel /ʌ/ in comparison with /ɒ/ in the context of /f.g/.

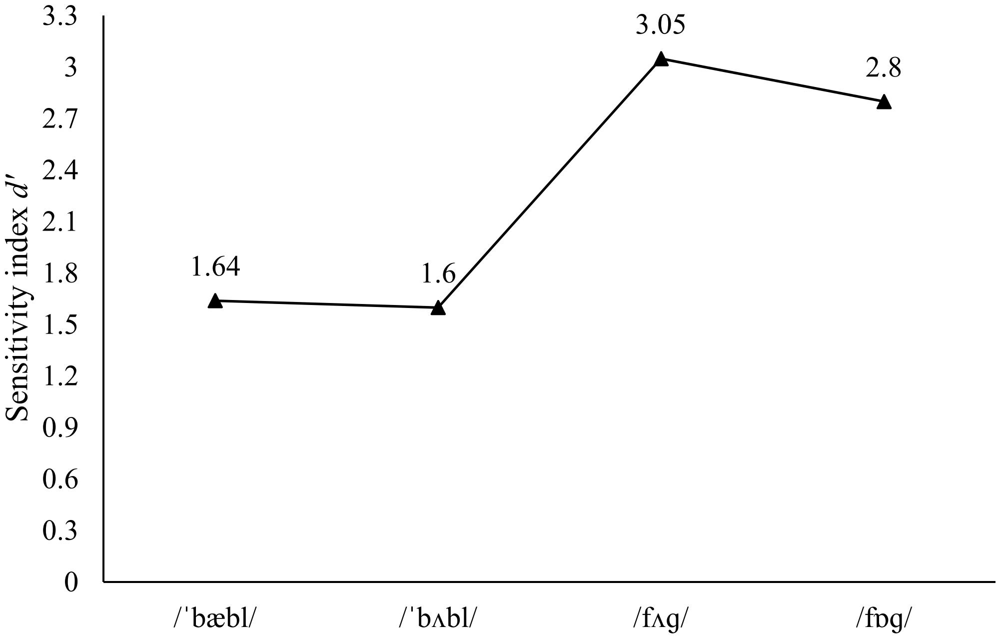


Figure 47. Identification of vowels in synthetic word tokens
Series 18. N = 53

7.3.1.19 Vocabulary Series 19.

Among the vowels /æ/ and /ʌ/, participants demonstrated a higher sensitivity towards the latter, with a sensitivity index *d'* ranging between 3.22 in the context /f.nd/ and 3.44 in the context /ʃ.t/. This finding does but confirm previous pattern of sensitivity with relation to these two vowels.

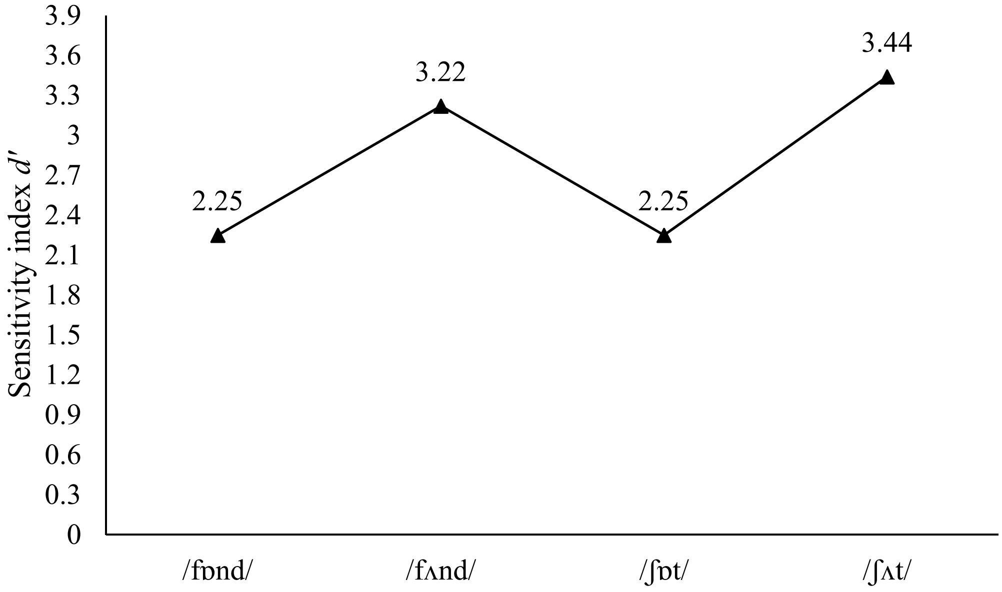


Figure 48. Identification of vowels in synthetic word tokens
Series 19. N = 53

7.3.1.20 Vocabulary Series 20.

In this two stimuli identification series, participants confirmed their considerable and high sensitivity towards the vowel /ʌ/ over /ɒ/. This fact holds true regardless of whether these vowels occur in 1- or 2-syllable word.

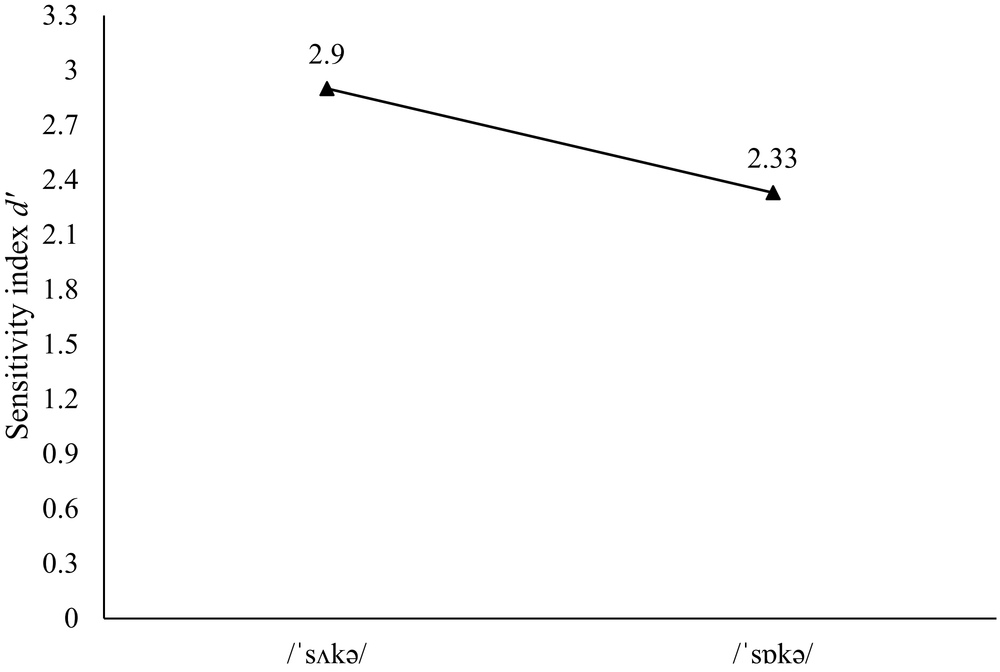


Figure 49. Identification of vowels in synthetic word tokens Series 20. N = 53

7.3.1.21 Vocabulary Series 21.

With respect to the last vowels compared in the last vocabulary series as displayed in *Figure 50*, participants continued to show their high sensitivity to the vowel /æ/ with an index *d'* ranging between 2.08 and 3.05 compared to the vowel /ɒ/ that was identified with an index *d'* ranging between 1.77 and 1.99 in the experimented on contexts.

In sum, findings of vowel identification in context revealed fairly participants' increased sensitivity towards several vowels and demonstrated probably a specific pattern of sensitivity. The vowels /ɔː/, /ɑː/, and /æ/ proved to be accurately identified in comparison with other vowels that themselves showed some variability across phonetic context. Away from vowel duration, participants revealed a sensitivity preference towards the vowel /æ/ in opposition to /ɒ/.

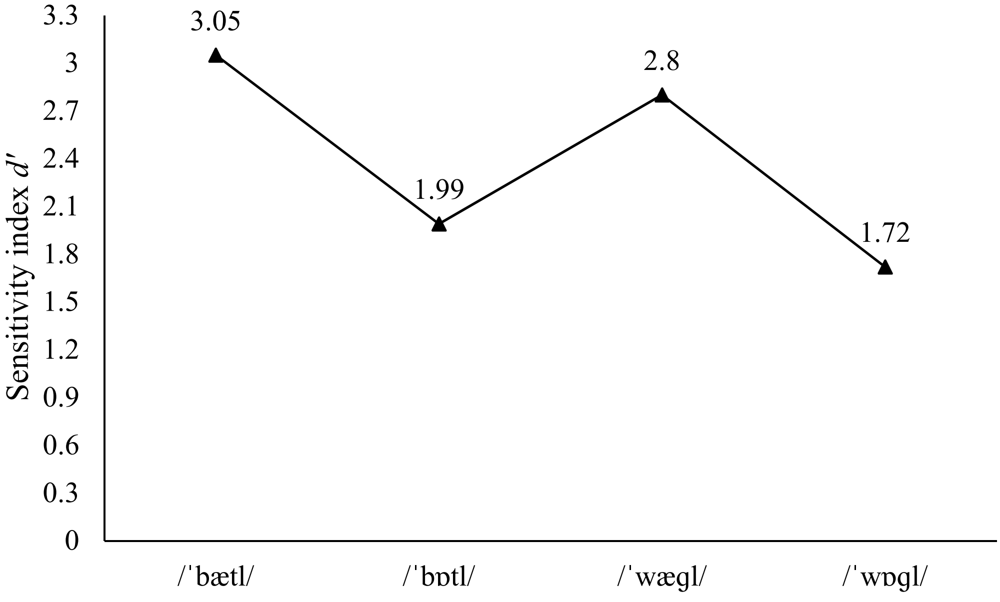


Figure 50. Identification of vowels in synthetic word tokens
Series 21. N = 53

In the following section, we present the confusion matrices to allow locating cases of misidentification and source of confusion.

7.3.2 Confusion matrices.

Table 104

Mean Identification of Vowels in 1-Syllable Words (Confusion Matrices)

Number of stimuli				
NS = 3	Vocabulary series *(m-Alternative Forced Choice)*			
Word token	Series 6			
	/mæʧ/	/mʌʧ/	/mɑːʧ/	
/mæʧ/	**.71**	.27	.02	
/mʌʧ/	.41	**.58**	.01	
/mɑːʧ/	.05	.06	**.89**	
NS = 4	Vocabulary series *(m-Alternative Forced Choice)*			
Word token	Series 2			
	/ʧæd/	/ʧɑːd/	/ʧæt/	/ʧɑːt/
/ʧæd/	**.78**	.11	.09	.02
/ʧɑːd/	.09	**.82**	.01	.00
/ʧæt/	.06	.00	**.90**	.03
/ʧɑːt/	.05	.11	.12	**.73**
Word token	Series 4			
	/gæb/	/gɑːb/	/bæʤ/	/bʌʤ/
/gæb/	**.95**	.05	.00	.00
/gɑːb/	.05	**.95**	.00	.00
/bæʤ/	.00	.00	**.66**	.33
/bʌʤ/	.00	.00	.41	**.59**
Word token	Series 7			
	/bɪn/	/biːn/	/ɪʧ/	/iːʧ/
/bɪn/	**.98**	.02	.00	.00
/biːn/	.04	**.96**	.00	.00
/ɪʧ/	.00	.01	**.95**	.05
/iːʧ/	.00	.00	.05	**.94**
Word token	Series 14			
	/kæb/	/kɑːb/	/kʌb/	/kɒb/
/kæb/	**.75**	.08	.16	.01
/kɑːb/	.02	**.92**	.03	.02
/kʌb/	.40	.01	**.55**	.04
/kɒb/	.13	.07	.13	**.47**
Word token	Series 19			
	/fɒnd/	/fʌnd/	/ʃɒt/	/ʃʌt/
/fɒnd/	**.87**	.13	.00	0.00
/fʌnd/	.02	**.97**	.00	.00
/ʃɒt/	.00	.00	**.87**	.13
/ʃʌt/	.00	.00	.02	**.98**

Note. N = 53. Values in boldface are mean hit rates for vowels. Values are rounded to second decimal. $\sum$ Trials per word = 10. Sum of rates per vowel identification may not equal 1 due to rounding. *NS:* Number of Stimuli.

Table 104 (*continued*)

Mean Identification of Vowels in 1-Syllable Words (Confusion Matrices)

Number of stimuli ***NS = 5***	Vocabulary series *(m-Alternative Forced Choice)*				
Word token			Series 1		
	/bæn/	/bʌn/	/bɑːn/	/bɒn/	/bɔːn/
/bæn/	**.62**	.26	.11	.01	.00
/bʌn/	.50	**.41**	.04	.04	.01
/bɑːn/	.09	.05	**.70**	.06	.10
/bɒn/	.13	.24	.08	**.52**	.03
/bɔːn/	.00	.00	.00	.17	**.83**
Word token			Series 9		
	/bæd/	/bʌd/	/bɑːd/	/bɒd/	/bɔːd/
/bæd/	**.82**	.12	.06	.00	.00
/bʌd/	.47	**.47**	.00	.03	.02
/bɑːd/	.03	.05	**.89**	.01	.02
/bɒd/	.05	.35	.09	**.48**	.02
/bɔːd/	.00	.00	.02	.09	**.88**
Word token			Series 10		
	/bæt/	/bʌt/	/bɑːt/	/bɒt/	/bɔːt/
/bæt/	**.71**	.21	.08	.00	.00
/bʌt/	.43	**.53**	.02	.02	.00
/bɑːt/	.04	.10	**.80**	.03	.04
/bɒt/	.08	.32	.09	**.49**	.03
/bɔːt/	.00	.01	.01	.10	**.88**
Word token			Series 15		
	/kɒd/	/kɑːd/	/kɔːd/	/kɒn/	/kɔːn/
/kɒd/	**.89**	.09	.02	.00	.00
/kɑːd/	.03	**.96**	.01	.00	.00
/kɔːd/	.05	.06	**.88**	.00	.00
/kɒn/	.00	.00	.00	**.98**	.02
/kɔːn/	.00	.00	.01	.09	**.90**
Word token			Series 17		
	/bæk/	/bʌk/	/bɑːk/	/ʧɔːk/	/ʧʌk/
/bæk/	**.84**	.12	.03	.00	.00
/bʌk/	.32	**.67**	.02	.00	.00
/bɑːk/	.02	.03	**.94**	.00	.00
/ʧɔːk/	.00	.00	.00	**.97**	.03
/ʧʌk/	.00	.00	.02	.02	**.95**

Note. N = 53. Values in boldface are mean hit rates for vowel. Values are rounded to second decimal. ∑ Trials per word = 10. Rates per vowel identification may not equal 1 due to rounding. *NS:* Number of Stimuli.

Table 104 (*continued*)

Mean Identification of Vowels in 1-Syllable Words (Confusion Matrices)

Number of stimuli					
***NS* = 2**	Vocabulary series *(m-Alternative Forced Choice)*				
Word token	Series 20				
	/ˈsʌkə/	/ˈsɒkə/			
/ˈsʌkə/	**.98**	.02			
/ˈsɒkə/	.05	**.95**			
***NS* = 4**	Vocabulary series				
Word token	Series 3				
	/ˈmætə/	/ˈmʌtə/	/ˈmɑːtə/	/ˈmɔːtə/	
/ˈmætə/	**.60**	.36	.03	.01	
/ˈmʌtə/	.53	**.45**	.01	.01	
/ˈmɑːtə/	.09	.09	**.72**	.09	
/ˈmɔːtə/	.01	.02	.01	**.97**	
Word token	Series 21				
	/ˈbætl/	/ˈbɒtl/	/ˈwægl/	/ˈwɒgl/	
/ˈbætl/	**.96**	.04	.00	.00	
/ˈbɒtl/	.18	**.82**	.00	.00	
/ˈwægl/	.00	.00	**.94**	.06	
/ˈwɒgl/	.00	.01	.23	**.76**	
Word token	Series 5				
	/ˈbætə/	/ˈbʌtə/	/ˈbɑːtə/	/ˈflætə/	/ˈflʌtə/
/ˈbætə/	**.58**	.30	.11	.00	.00
/ˈbʌtə/	.36	**.60**	.04	.00	.00
/ˈbɑːtə/	.02	.02	**.96**	.00	.00
/ˈflætə/	.00	.00	.00	**.61**	.39
/ˈflʌtə/	.00	.00	.00	.51	**.49**

Note. N = 53. Values in boldface are mean hit rates for vowels. Values are rounded to second decimal. $\sum$ Trials per word = 10. Rates per vowel identification may not equal 1 due to rounding. *NS:* Number of Stimuli.

Table 104 demonstrates notable findings about vowel misidentification in context, with participants showing also some perceptual difficulties at the level of short vowel identification. Participants failed at perceptually identify some vowels in a systematic way. For instance, participants misidentified and confused between the vowels /æ/ and /ʌ/ in several contexts such as /m.ʧ/, /b.ʤ/, /k.b/, /b.d/, /b.t/, /b.k/, /m.tə/, /b.tə/, /fl.tə/, and /b.n/. In an analogous way, participants misidentified and confused between /æ/ and /ɑː/ in contexts such as /ʧ.t/, /ʧ.d/, /k.b/, /b.n/, /b.d/, /b.t/, and /b.k/. Further misidentification and confusion was observed among participants for the pair /ʌ/ and /ɒ/ in contexts such as /k.b/, /f.nd/, and /ʃ.t/.

Table 105

Mean Identification of Vowels in 1- & 2-Syllable Words (Confusion Matrices)

Number of stimuli				
NS = 4	Vocabulary series *(m-Alternative Forced Choice)*			
Word token	Series 8			
	/ˈlɪvə/	/ˈliːvə/	/pɪʧ/	/piːʧ/
/ˈlɪvə/	**.91**	.09	.00	.00
/ˈliːvə/	.12	**.88**	.00	.00
/pɪʧ/	.00	.00	**.98**	.02
/piːʧ/	.00	.00	.07	**.93**
Word token	Series 11			
	/ˈfeðə /	/ˈfɜːðə/	/wɜːnt/	/went/
/ˈfeðə/	**.93**	.07	.00	.00
/ˈfɜːðə/	.02	**.98**	.00	.00
/wɜːnt/	.00	.00	**.75**	.25
/went/	.00	.00	.06	**.94**
Word token	Series 13			
	/sʊt/	/suːt/	/fʊl/	/fuːl/
/sʊt/	**.98**	.02	.00	.00
/suːt/	.05	**.95**	.00	.00
/fʊl/	.00	.00	**.88**	.12
/fuːl/	.00	.00	.09	**.91**
Word token	Series 16			
	/ˈhɒtɪ/	/ˈhɑːtɪ/	/ kɒn/	/kæn/
/ˈhɒtɪ/	**.96**	.03	.00	.00
/ˈhɑːtɪ/	.06	**.94**	.00	.00
/ kɒn/	.00	.00	**.90**	.10
/kæn/	.00	.00	.05	**.95**
Word token	Series 18			
	/kæb/	/kɑːb/	/kʌb/	/kɒb/
/ˈbæbl/	**.74**	.26	.00	.00
/ˈbʌbl/	.27	**.73**	.00	.00
/fʌg/	.00	.00	**.96**	.03
/fɒg/	.00	.00	.06	**.94**

Note. N = 53. Values in boldface are mean hit rates for vowels. Values are rounded to second decimal. ∑ Trials per word = 10. Rates per vowel identification may not equal 1 due to rounding. *NS:* Number of Stimuli.

However, few cases of misidentification and confusion were observed among participants for long vowels. Participants misidentified and confused the vowels /ɔː/ and /ɑː/ in contexts such as /b.n/ and /b.d/, and the vowels /ʊ/ and /uː/ in the context of /f.l/.

Table 105 (*continued*)

Mean Identification of Vowels in 1 & 2-Syllable Words (Confusion Matrices)

Number of stimuli (*break block*)						
NS = 6	Vocabulary series *(m-Alternative Forced Choice)*					
Word token	Series 12					
	/bɪʧ/	/bʊʧ/	/hɪd/	/hʊd/	/sɪt/	/sʊt/
/bɪʧ/	**1.00**	.00	.00	.00	.00	.00
/bʊʧ/	.00	**1.00**	.00	.00	.00	.00
/hɪd/	.00	.00	**1.00**	.00	.00	.00
/hʊd/	.00	.00	.00	**1.00**	.00	.00
/sɪt/	.00	.00	.00	.00	**1.00**	.00
/sʊt/	.00	.00	.00	.00	.00	**1.00**

Note. N = 53. Values in boldface are mean of hit rate for vowel. Values are rounded to second decimal. ∑ Trials per word = 10. Rates per vowel identification may not equal 1 due to rounding. *NS:* Number of Stimuli.

7.3.3 Variation in identification of vowel in context.

To investigate further any statistical significance regarding the observed cases of misidentification and confusion between compared vowels, we ran a series of *t*-tests and one-way ANOVA on identification measures. It is to note that the analyses were performed on data corresponding to the vowels occurring in the same context only. That is, when the phonetic context was experimented on with two vowels only, a paired-samples *t*-test was used. However, when the phonetic context was experimented on using more than two vowels, a one-way ANOVA with repeated measures was performed. We present the findings in the following sections.

7.3.3.1 Identification across /æ/, /ʌ/, /a:/, /ɒ/ and /ɔ:/.

Table 106 shows that paired samples *t*-test on identification rate of vowel among /æ/, /ɔ:/, /ʌ/, /ɑ:/ and /ɒ/ demonstrated few significant results. The vowel /ɑ:/ in the context /ʧ d/ was identified at a significantly higher rate than /æ/, $t(51) = -6.26$, 2-tailed $p = .000$, Cohen's $d = .86$. Participants identified /æ/ at a statistically significantly higher rate than /ʌ/ in the context /fl.tə/, ($t(51) = 2.78$, 2-tailed $p = .008$, Cohen's $d = .38$. The vowel /ʌ/ was identified at a significantly higher rate than the vowel /ɒ/ in the contexts /ʃ.t/, $t(51) = -6.26$, 2-tailed $p = .000$, Cohen's *d* = .58, and /f.nd/ (*t*(51) = -4.78, 2-tailed *p* = .000, Cohen's *d* = .66). Other significant differences in identification rate of vowel included /æ/ and /ɒ/ in the context of /k.n/, with the former being identified at a higher rate than the latter, *t*(51) =2.65, 2-tailed *p* = .001, Cohen's *d* = .36. The only unexpected result was observed with /ɒ/ being identified at a significantly higher rate than /ɔː/ in the same previous context /k.n/, *t*(51) = -3.20, 2-tailed *p* = .002, Cohen's *d* = .44.

Table 106

Paired Samples t-Test on Mean Identification across Vowels /æ/, /ʌ/, /ɑː/, /ɒ/, /ɔː/

	Vowels					Paired differences				
	/æ/	/ʌ/	/ɑː/	/ɒ/	/ɔː/	Hit Rates				
Word token	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	Mean difference	*SD*	*t*	*Sig.* (*2-tailed*)	ES Cohen's *d*
/ʧ d/	.78 (.19)		.82 (.17)			.04	.20	1.50	.139	
/ʧ t/	.90 (.13)		.73 (.22)			-.18*	.21	-6.26	.000	.86
/g b/	.95 (.07)		.94 (.11)			.01	.12	0.46	.646	.06
/b ʤ/	.66 (.27)	.60 (.27)				.07	.29	1.80	.077	.25
/f nd/		.97 (.07)		.87 (.14)		.10*	.15	-4.78	.000	.66
/ʃ t/		.98 (.08)		.87 (.17)		.18*	.18	-6.26	.000	.58
/ʧ k/		.95 (.08)			.97 (.07)	.02	.08	1.81	.077	
/k n/				.98 (.05)	.90 (.17)	-.08*	.19	-3.20	.002	.44
/k n/	.95 (.10)			.90 (.16)		.05*	.14	2.65	.011	.36
/ˈs .kə/		.98 (.05)		.94 (.13)		.04	.14	-2.01	.049	
/ˈb tl/	.95 (.08)			.82 (.17)		.14*	.18	5.50	.000	.76
/ˈw gl/	.94 (.14)			.77 (.22)		.17*	.21	5.60	.000	.89
/ˈfl .tə/	.61 (.22)	.49 (.21)				.11*	.30	2.78	.008	.38
/ˈh .ti/			.94 (.10)	.96 (.07)		-.03	.11	-1.72	.090	
/ˈb bl/	.74 (.28)	.73 (.33)				.01	.20	.28	.782	
/f g/		.96 (.07)		.94 (.08)		.02	.08	-1.75	.086	

Note. N = 53. *df* = 52. Values are mean hit rates. Values are rounded to second decimal, except for probability. Confidence interval (CI) = 95.

7.3.3.2 Identification across /ɪ/, /iː/, /ʊ/, /uː/, /e/ and /ɜː/.

As for variation in vowel identification among the above-cited vowels, paired-samples t-test in Table 107 demonstrated few significant findings. Participants identified the vowel /ɪ/ at a statistically significantly higher rate than the vowel /iː/ in the context of /p.ʧ/, $t(51) = -2.67$, 2-tailed $p = .010$, Cohen's $d = .37$. For the vowels /e/ and /ɜː/, findings demonstrated their different patterns of identification in tests contexts. The vowel /e/ was identified at a significantly higher rate than /ɜː/ in the /w.nt/ context, $t(51) = 6.87$, 2-tailed $p = .000$, Cohen's $d = .94$. However, the vowel /ɜː/ was identified at a significantly higher rate than /e/ in the context of /ˈf .ðə/, $t(51) = -2.50$, 2-tailed $p = .016$, Cohen's $d = .34$. Lastly, the vowel /ʊ/ was identified at a significantly higher rate than /uː/ in the context of /s.t/, $t(51) = 3.05$, 2-tailed $p = .004$, Cohen's $d = .42$. These findings suggest that pattern of identification of these vowels may not be easily predicted from context. Nonetheless, overall identification rates of these vowels in context are fairly higher than their identification in isolation.

Table 107

Paired Samples t-Test on Mean Identification across Vowels

	Vowels						Paired differences				
	/ɪ/	/iː/	/ʊ/	/uː/	/e/	/ɜː/	Hit Rates				
Word token	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	Mean difference	*SD*	*t*	*Sig.* (*2-tailed*)	ES Cohen's *d*
/ˈl .və/	.91 (.13)	.88 (18)					.02	.18	-0.99	.328	
/p ʧ/	.98 (.06)	.93 (.13)					.05*	.13	-2.67	.010	.37
/s t/			.98 (.04)	.95 (.08)			.04*	.09	3.05	.004	.42
/f l/			.88 (.20)	.91 (.18)			-.03	.11	1.79	.080	
/ˈf .ðə/					.93 (.17)	.98 (.05)	-.06*	.16	-2.50	.016	.34
/w nt/					.94 (.11)	.74 (.22)	.19*	.21	6.87	.000	.94
/h d/	.99 (.02)		.99 (.01)				.00	.02	-0.57	.569	
/b ʧ /	.99 (.02)		.99 (.00)				.00	.02	-1.43	.159	
/ ʧ/	.95 (.10)	.94 (.12)					.01	.11	-0.63	.534	
/b n/	.98 (.06)	.96 (.11)					.02	.09	-1.32	.192	
/s t/	.99 (.01)		99 (.01)	.			.00	.01	-1.00	.322	

Note. N = 53. *df* = 52. Values are mean hit rates. Values are rounded to second decimal, except for probability. Confidence interval (CI) = 95%.

7.3.3.3 Identification in vocabulary Series 1.

Table 108

Within-Subjects Effects for Mean Identification of Vowel (Vocabulary Series 1)

Source: /b n/	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	4	25.826	.000	.332	1.000
Greenhouse-Geisser	3.245	25.826	.000	.332	1.000
Huynh-Feldt	3.487	25.826	.000	.332	1.000
Lower-bound	1	25.826	.000	.332	.999

[a]. Computed using alpha = .05

As displayed in Table 108, a one-way ANOVA with repeated measures was conducted to compare the effects of vowel on mean identification rate in the phonetic context /b.n/. The one-way repeated measures ANOVA with the lower-bound correction showed that mean identification rate differed significantly between vowels /æ/, /ɑː/, /ɒ/, /ɔː/ and /ʌ/, $F(1, 52) = 25.826$, $p < .000$, $\eta^2 = .332$.

As displayed in Table 109 below, post hoc tests using the Bonferroni correction revealed that /æ/ was identified at significantly a lower than the vowel /ɔː/ (*MD* = -.21, $p < .000$) and at significantly higher rate than /ʌ/ (*MD* = .21, $p < .000$). The vowel /ɑː/ was identified at a significantly higher rate than /ɒ/ and /ʌ/, with *MD* = .18 ($p < .000$) and .29 ($p < .000$), respectively. However, participants identified /ɔː/ at a significantly higher rate than /ɑː/ (*MD* = .13, $p < .000$), /ʌ/ (*MD* = .42, $p < .000$), and /ɒ/ (*MD* = .31, $p < .000$). These findings demonstrated a perceptual advantage for the long vowels over the short ones, with /ɔː/ as the most significantly highly identified. Yet, within the short ones, the findings demonstrated this specific perceptual decreasing pattern of identification, /æ/, /ɒ/, and /ʌ/.

Table 109

Pairwise Comparisons of Mean Identification of Vowel (Vocabulary Series 1)

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/æ/	/ɑː/	-.08	.804
	/ɒ/	.10	.602
	/ɔː/	-.21*	.000
	/ʌ/	.21*	.001
/ɑː/	/ɒ/	.18*	.019
	/ɔː/	-.13*	.004
	/ʌ/	.29*	.000
/ɒ/	/ɔː/	-.31*	.000
	/ʌ/	.11*	.329
/ɔː/	/ʌ/	.42*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

7.3.3.4 Identification in vocabulary Series 3.

Table 110

Within-Subjects Effects for Mean Identification of Vowel (Vocabulary Series 3)

Source: /ˈm .tə/	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	3	69.446	.000	.572	1.000
Greenhouse-Geisser	2.475	69.446	.000	.572	1.000
Huynh-Feldt	2.608	69.446	.000	.572	1.000
Lower-bound	1	69.446	.000	.572	1.000

[a]. Computed using alpha = .05

In an analogous manner to what is shown in Table 110, a repeated measure ANOVA with the lower-bound correction showed that mean identification rate of vowel differed significantly between vowels, $F(1, 52) = 69.446$, $p < .000$, $\eta^2 = .572$. Post hoc tests with Bonferroni correction demonstrated that the vowel /ɔː/ was identified at a significantly higher rate than all other vowels, with a mean difference ranging between .24 to .52, followed by the vowel /ɑː/ that was identified at a significantly higher rate than /æ/ (*MD* = .13, $p < .05$) and /ʌ/

(*MD* = .27, $p < .000$). Once more, the findings suggest higher identification rate to be strongly correlated with long vowels.

Table 111

Pairwise Comparisons of Mean Identification of Vowel (Vocabulary Series 3)

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/ɑː/	/æ/	.13*	.020
	/ɔː/	-.24*	.000
	/ʌ/	.27*	.000
/æ/	/ɔː/	-.37*	.000
	/ʌ/	.15*	.011
/ɔː/	/ʌ/	.52*	.031

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

7.3.3.5 Identification in vocabulary Series 5.

Table 112

Within-Subjects Effects for Mean Identification of Vowel (Vocabulary Series 5)

Source: /ˈb .tə/	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	2	77.199	.000	.598	1.000
Greenhouse-Geisser	1.930	77.199	.000	.598	1.000
Huynh-Feldt	2	77.199	.000	.598	1.000
Lower-bound	1	77.199	.000	.598	1.000

[a]. Computed using alpha = .05

Table 112 above displays significant findings of a repeated measures ANOVA (with the lower-bound correction) on effect of vowel on mean identification rate, $F(1, 52) = 77.199$, $p < .000$, $\eta^2 = .598$. Post hoc tests using the Bonferroni correction showed that the vowel /ɑː/ was always identified at a significantly higher rate than /æ/ (*MD* = .38, $p < .000$) and /ʌ/ (*MD* = .37, $p < .000$). However, participants identified /æ/ and /ʌ/ at almost similar rates.

Table 113

Pairwise Comparisons of Mean Identification of Vowel (Vocabulary Series 5)

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/ɑː/	/æ/	.38*	.000
	/ʌ/	.37*	.000
/æ/	/ʌ/	-.02	1.000

Note. Pairwise comparisons are based on estimated marginal means.

*. The mean difference is significant at the .05 level.

[b]. Adjustment for multiple comparisons: Bonferroni.

7.3.3.6 Identification in vocabulary Series 6.

Table 114

Within-Subjects Effects for Mean Identification of Vowel (Vocabulary Series 6)

Source: /m ʧ/	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	2	50.249	.000	.491	1.000
Greenhouse-Geisser	1.801	50.249	.000	.491	1.000
Huynh-Feldt	1.801	50.249	.000	.491	1.000
Lower-bound	1	50.249	.000	.491	1.000

[a]. Computed using alpha = .05

The previous findings of the perceptual advantage of long vowels as being highly identified than others in context is demonstrated again as revealed in Table 114. A one-way ANOVA with repeated measures using the lower-bound correction showed that identification rate differed as a function of vowel, $F(1, 52) = 50.249$, $p < .000$, $\eta^2 = .491$. Post hoc test using the Bonferroni correction revealed that /ɑː/ was identified at a significantly higher rate than /æ/ (*MD* = .18, $p < .000$) and /ʌ/ (*MD* = .31, $p < .000$). For the latter pair of vowels, /æ/ was identified at a significantly higher rate than /ʌ/ with a mean difference equal to .13.

Table 115

Pairwise Comparisons of Mean Identification of Vowel (Vocabulary Series 6)

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/ɑː/	/æ/	.18*	.000
	/ʌ/	.31*	.000
/æ/	/ʌ/	.13*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

7.3.3.7 Identification in vocabulary Series 9.

Table 116

Within-Subjects Effects for Mean Identification of Vowel (Vocabulary Series 9)

Source: /b d/	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	4	70.194	.000	.574	1.000
Greenhouse-Geisser	3.117	70.194	.000	.574	1.000
Huynh-Feldt	3.338	70.194	.000	.574	1.000
Lower-bound	1	70.194	.000	.574	1.000

[a]. Computed using alpha = .05

Table 116 above shows the results of a one-way ANOVA with repeated measures using the lower-bound correction of vowel effect on identification rate. The obtained findings show statistical significance ($F(1, 52) = 70.194$, $p < .000$, $\eta^2 = .574$), with long vowels identified at significantly higher rate than short ones. While /ɑː/ and /ɔː/ were identified with almost identical rates, the latter was identified at a significantly higher rate than /ɒ/ (*MD* = .40) and /ʌ/ (*MD* = .42, $p < .000$). Moreover, difference in mean identification of /ɒ/ and /ʌ/ was not statistically significant.

Table 117

Pairwise Comparisons of Mean Identification of Vowel (Vocabulary Series 9)

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* ^{b}
/æ/	/ɑː/	-.07	.227
	/ɔː/	-.06	.615
	/ɒ/	.34*	.000
	/ʌ/	.35*	.000
/ɑː/	/ɔː/	.01	1.000
	/ɒ/	.41*	.000
	/ʌ/	.43*	.000
/ɔː/	/ɒ/	.40*	.000
	/ʌ/	.42*	.000
/ɒ/	/ʌ/	.01	1.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
^{b}. Adjustment for multiple comparisons: Bonferroni.

7.3.3.8 Identification in vocabulary Series 10.

As shown in Table 118, a one-way repeated measures ANOVA with the lower-bound correction showed that mean identification rate differed significantly between vowels, *F*(1, 52) = 33.624, $p < .000$, $\eta^2 = .393$. Post hoc test using the Bonferroni correction revealed that the vowels /ɔː/ was always identified with statistically a higher rate than the vowels /ʌ/ (*MD* = .35, $p < .000$), /ɒ/ (*MD* = .39), and /æ/ (*MD* = .17, $p < .05$). While the vowel /ɑː/ was identified with a statistically higher rate than /ɒ/ (*MD* = .31, $p < .000$) and /ʌ/ (*MD* = .27, $p < .000$), the latter vowels did not differ significantly in their identification. The reported findings suggest consistently the perceptual identification advantage of long vowels over the short ones in context.

Table 118

Within-Subjects Effects for Mean Identification of Vowel (Vocabulary Series 10)

Source: /b t/	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	4	33.624	.000	.393	1.000
Greenhouse-Geisser	3.331	33.624	.000	.393	1.000
Huynh-Feldt	3.587	33.624	.000	.393	1.000
Lower-bound	1	33.624	.000	.393	1.000

[a]. Computed using alpha = .05

Table 119

Pairwise Comparisons of Mean Identification of Vowel (Vocabulary Series 10)

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/ɑː/	/æ/	.09	.262
	/ɒ/	.31*	.000
	/ɔː/	-.08	.129
	/ʌ/	.27*	.000
/æ/	/ɒ/	.22*	.000
	/ɔː/	-.17*	.000
	/ʌ/	.17*	.016
/ɒ/	/ɔː/	-.39*	.000
	/ʌ/	-.05	1.000
/ɔː/	/ʌ/	.35*	.000

Note. Pairwise comparisons are based on estimated marginal means.

*. The mean difference is significant at the .05 level.

[b]. Adjustment for multiple comparisons: Bonferroni.

7.3.3.9 Identification in vocabulary Series 14.

Table 120

Within-Subjects Effects for Mean Identification of Vowel (Vocabulary Series 14)

Source: /k b/	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	3	41.283	.000	.443	1.000
Greenhouse-Geisser	2.794	41.283	.000	.443	1.000
Huynh-Feldt	2.969	41.283	.000	.443	1.000
Lower-bound	1	41.283	.000	.443	1.000

[a]. Computed using alpha = .05

Table 120 shows significant results of a one-way ANOVA with repeated measures conducted on Vocabulary Series 14 data. Using the lower-bound correction, mean identification rate different significantly between vowels, *F*(1, 52) = 41.283, $p < .000$, $\eta^2 = .443$. Post hoc tests using Bonferroni correction further revealed that the long vowel /ɑː/ was identified at a significantly higher rate than /æ/ (*MD* = .17, $p < .01$), /ɒ/ (*MD* = .45, $p < .000$), and /ʌ/ (*MD* = .38, $p < .000$). Additionally, the vowel /æ/ was identified at a significantly higher rate than /ɒ/ (*MD* = .28, $p < .000$) and /ʌ/ (*MD* = .20, $p < .01$). However, the latter vowels did not differ significantly in identification.

Table 121

Pairwise Comparisons of Mean Identification of Vowel (Vocabulary Series 14)

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/æ/	/ɑː/	-.17*	.001
	/ɒ/	.28*	.000
	/ʌ/	.20*	.001
/ɑː/	/ɒ/	.45*	.000
	/ʌ/	.38*	.000
/ɒ/	/ʌ/	-.08	.603

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

7.3.3.10 ***Identification in vocabulary Series 15.***

Table 122

Within-Subjects Effects for Mean Identification of Vowel (Vocabulary Series 15)

Source: /k d/	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	2	4.745	.011	.084	.781
Greenhouse-Geisser	1.842	4.745	.013	.084	.781
Huynh-Feldt	1.907	4.745	.012	.084	.781
Lower-bound	1	4.745	.034	.084	.781

[a]. Computed using alpha = .05

A repeated measure ANOVA with the lower-bound correction, as displayed in Table 122, showed that mean identification rate differed significantly between vowels, $F(1, 52) = 4.745$, $p < .000$, $\eta^2 = .084$. Post hoc tests using the Bonferroni correction revealed this time that the vowel /ɑː/ was identified at a significantly higher rate than /ɒ/ ($MD = .06$, $p < .05$) and /ɔː/ ($MD = .08$, $p < .05$), and that the latter did not differ significantly in their identification ($MD = .01$). However, these findings did not replicate the previous pattern of identification, and the order of the perceptual pattern was not the same. Indeed, the very low effect size reported in Table 122 ($\eta^2 = .084$) undermines the power of the statistical test and thus the possibility of considering these results to try to generalise these findings.

Table 123

Pairwise Comparisons of Mean Identification of Vowel (Vocabulary Series 15)

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/ɑː/	/ɒ/	.06*	.017
	/ɔː/	.08*	.029
/ɒ/	/ɔː/	.01	1.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

7.3.3.11 Identification in vocabulary Series 17.

Table 124

Within-Subjects Effects for Mean Identification of Vowel (Vocabulary Series 17)

Source: /b k/	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	2	41.113	.000	.442	1.000
Greenhouse-Geisser	1.725	41.113	.000	.442	1.000
Huynh-Feldt	1.778	41.113	.000	.442	1.000
Lower-bound	1	41.113	.000	.442	1.000

[a]. Computed using alpha = .05

Table 124 reports the findings of a one-way ANOVA with repeated measures conducted on effect of vowel on mean identification rate. Using the lower-bound correction, the findings revealed that mean identification rate differed significantly between vowels, *F*(1, 52) = 41.113, $p < .000$, $\eta^2 = .442$. Post hoc tests using the Bonferroni correction revealed that participants identified the vowel /ɑː/ at a significantly higher rate than /æ/ (*MD* = .10, $p < .000$) and /ʌ/ (*MD* = .28, $p < .000$), and identified too the vowel /æ/ at a significantly higher rate than /ʌ/ (*MD* = .28, $p < .000$). These findings replicated the previous order of perceptual patterns.

Table 125

Pairwise Comparisons of Mean Identification of Vowel (Vocabulary Series 17)

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/æ/	/ɑː/	-.10*	.000
	/ʌ/	.18*	.000
/ɑː/	/ʌ/	.28*	.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

7.3.4 Variation in identification and reaction time across vowel and context.

We present here:

(a) Aggregate calculated means of vowel identification and reaction time to help visualise the pattern of the latter in all experimental manipulations, in isolation and in context (Table 126); and

(b) Mean for vowel identification and reaction time to help visualise the factors averaged means for every mean identification and reaction time for vowel and context (Table 127).

As showed in Table 126, participants identified long vowels at substantially higher mean identification rates than short vowels, indicating further the easiness and precision with which participants could identify them.

Table 126

Aggregate Mean Identification of Vowel across all Manipulations and Reaction Time

	Mean identification (*SD*)			Mean reaction time (*SD*)		
Vowel	*Isolated*	*Context 1*	*Context 2*	*Isolated*	*Context 1*	*Context 2*
/iː/	0.70 (0.14)	0.94 (0.10)	0.88 (0.18)	2.91 (1.46)	2.55 (1.33)	2.51 (2.23)
/ɔː/	0.34 (0.11)	0.98 (0.10)	0.91 (0.09)	2.53 (0.88)	3.16 (1.21)	3.08 (2.26)
/ɑː/	0.58 (0.09)	0.86 (0.09)	0.88 (0.09)	3.11 (0.95)	3.06 (0.92)	3.08 (1.79)
/ɜː/	0.58 (0.15)	0.75 (0.22)	0.98 (0.05)	4.21 (1.38)	2.65 (2.23)	2.99 (2.80)
/ʊ/	-	-	-	-	-	-
/uː/	-	-	-	-	-	-
/ɪ/	0.34 (0.30)	0.98 (0.03)	0.91 (0.13)	3.93 (1.38)	2.49 (0.85)	2.59 (2.26)
/e/	0.33 (0.20)	0.94 (0.11)	0.93 (0.17)	4.24 (1.87)	2.10 (1.86)	2.99 (2.63)
/ʌ/	0.25 (0.21)	0.98 (0.03)	0.91 (0.13)	3.66 (1.37)	2.95 (0.90)	3.11 (1.29)
/æ/	0.16 (0.16)	0.99 (0.23)	0.74 (0.13)	3.38 (1.26)	3.34 (0.95)	3.12 (1.20)
/ɒ/	0.07 (0.13)	0.74 (0.10)	0.87 (0.10)	3.53 (1.62)	2.82 (0.81)	2.27 (1.11)

Note. N = 53
Context 1: 1-syllable word
Context 2: 2-syllable word
For the vowels /ʊ/ and /uː/ were not experimented on in 2-syllable words, we excluded them from the analysis.

Table 127 (*Part A*)

Estimated Marginal Means of Vowel Identification and Reaction Time

	Identification				Reaction time			
			95% CI				95% CI	
Vowel	*M*	*SE*	Lower bound	Upper bound	*M*	*SE*	Lower bound	Upper bound
/iː/	0.84	0.01	0.81	0.87	2.66	0.14	2.38	2.94
/ɔː/	0.81	0.01	0.79	0.83	3.11	0.15	2.80	3.41
/ɑː/	0.77	0.01	0.75	0.79	3.29	0.15	2.99	3.58
/ɜː/	0.77	0.01	0.74	0.80	2.92	0.14	2.65	3.19
/ɪ/	0.74	0.01	0.72	0.77	2.87	0.11	2.65	3.10
/e/	0.73	0.01	0.70	0.76	3.08	0.12	2.85	3.32
/ʌ/	0.71	0.01	0.69	0.74	3.24	0.12	3.00	3.47
/æ/	0.63	0.02	0.59	0.67	3.28	0.10	3.07	3.49
/ɒ/	0.56	0.01	0.54	0.58	3.00	0.13	2.75	3.26

Note. N = 53.

Table 127 (*Part B*)

Estimated Marginal Means of Identification and Reaction Time for the Interaction of Vowel and Context

	Identification				Reaction time			
			95% CI				95% CI	
Measure	*M*	*SE*	Lower bound	Upper bound	*M*	*SE*	Lower bound	Upper bound
Isolated	0.40	0.01	0.39	0.42	3.50	0.09	3.32	3.68
Context 1	0.90	0.01	0.88	0.92	2.79	0.08	2.63	2.95
Context 2	0.89	0.01	0.87	0.91	2.86	0.10	2.67	3.05

Note. N = 53. Context 1: 1-syllable word. Context 2: 2-syllable word

Table 128

Within-Subjects Effects for Aggregate Mean Vowel Identification across All Manipulations

Source	Measure		*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Vowel	Mean identification	Sphericity Assumed	8	53.304	.000	.506	1.000
		Greenhouse-Geisser	5.537	53.304	.000	.506	1.000
		Huynh-Feldt	6.272	53.304	.000	.506	1.000
		Lower-bound	1	53.304	.000	.506	1.000
	Mean reaction time	Sphericity Assumed	8	3.229	.001	.058	.971
		Greenhouse-Geisser	6.266	3.229	.004	.058	.935
		Huynh-Feldt	7.218	3.229	.002	.058	.958
		Lower-bound	1	3.229	.078	.058	.422
Context	Mean identification	Sphericity Assumed	2	1874.631	.000	.973	1.000
		Greenhouse-Geisser	1.767	1874.631	.000	.973	1.000
		Huynh-Feldt	1.824	1874.631	.000	.973	1.000
		Lower-bound	1	1874.631	.000	.973	1.000
	Mean reaction time	Sphericity Assumed	2	31.001	.000	.373	1.000
		Greenhouse-Geisser	1.929	31.001	.000	.373	1.000
		Huynh-Feldt	2.000	31.001	.000	.373	1.000
		Lower-bound	1	31.001	.000	.373	1.000
Vowel*Context	Mean identification	Sphericity Assumed	16	60.095	.000	.536	1.000
		Greenhouse-Geisser	7.604	60.095	.000	.536	1.000
		Huynh-Feldt	9.033	60.095	.000	.536	1.000
		Lower-bound	1	60.095	.000	.536	1.000
	Mean reaction time	Sphericity Assumed	16	4.642	.000	.082	1.000
		Greenhouse-Geisser	8.754	4.642	.000	.082	.999
		Huynh-Feldt	10.682	4.642	.000	.082	1.000
		Lower-bound	1	4.642	.036	.082	.561

[a]. Computed using alpha = .05

Table 129

Pairwise Comparisons of Aggregate Mean Vowel Identification Test 2

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/iː/	/ɪ/	.10*	.000
	/æ/	.21*	.000
	/ʌ/	.13*	.000
	/ɑː/	.07*	.000
	/ɒ/	.28*	.000
	/ɔː/	.03	.655
	/ɜː/	.07*	.001
	/e/	.11*	.000
/ɪ/	/æ/	.12*	.000
	/ʌ/	.03	1.000
	/ɑː/	-.03	1.000
	/ɒ/	.18*	.000
	/ɔː/	-.06*	.013
	/ɜː/	-.02	1.000
	/e/	.01	1.000
/æ/	/ʌ/	-.09*	.049
	/ɑː/	-.14*	.000
	/ɒ/	.09*	.010
	/ɔː/	-.18*	.000
	/ɜː/	-.14*	.000
	/e/	-.11	.000
/ʌ/	/ɑː/	-.06*	.008
	/ɒ/	.15*	.000
	/ɔː/	-.09*	.000
	/ɜː/	-.05	.244
	/e/	-.02	1.000
/ɑː/	/ɒ/	-.21*	.000
	/ɔː/	-.03*	.003
	/ɜː/	.00	1.000
	/e/	.04	.812
/ɒ/	/ɔː/	-.25*	.000
	/ɜː/	-.21*	.000
	/e/	-.17*	.000

Note. Pairwise comparisons are based on estimated marginal means.

*. The mean difference is significant at the .05 level.

[b]. Adjustment for multiple comparisons: Bonferroni.

Table 129 (continued)

Pairwise Comparisons of Aggregate Mean Identification of Vowel

Vowel (I)	Vowel (J)	Mean difference (I – J)	*Sig.* [b]
/ɔː/	/ɜː/	.04	.433
	/e/	.07*	.002
/ɜː/	/e/	.04	.444

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

As shown in Table 128 above, a two-way repeated measures ANOVA with the lower-bound correction was conducted to compare the main effects of vowel, context and the interaction of vowel and context on identification and reaction time. All effects were statistically significant at the .01 level, except for vowel on reaction time that yielded an F ratio of $F(1, 416) = 3.229$, $p = .078$. The main effect of vowel on identification yielded an F ratio of $F(1, 416) = 53.304$, $p < .001$, indicating a significant difference between vowels. As demonstrated by pairwise multiple comparisons with post hoc Bonferroni adjustments in Table 129, higher mean identification rates were found with long vowels compared to their short counterparts. The main effect of content on identification yielded an F ratio of $F(1, 52) = 1874.631$, $p < .001$, indicating a significant difference between phonetic contexts, as demonstrated below in Table 130. In a similar fashion to previous findings, the interaction between vowel and phonetic content yielded a significant effect on identification and reaction time, yielding an F ratio of $F(1, 52) = 60.095$, $p < .001$ and $F(1, 52) = 4.642$, $p < .05$, respectively.

As shown above in Tables 128 through 130, the occurrence of isolated vowel tokens or within context affected significantly identification rate and reaction time as well. When the vowel occurred in isolation, it was identified at a significantly lower rate than when it occurred in context. Similarly, presentation of isolated vowel tokens or within context significantly affected the reaction time of participants' responses. The pattern was that vowels were identified at a significantly higher rate when they occurred in context than when they occurred in isolation (*MD* = .49, *p* < .000), with no further significant differences as to whether they occurred in 1- or 2-syllable words (*MD* = .01, *p* = 1). In an analogous way, the findings demonstrated that reaction time in participants' response differed significantly between contexts, mainly when the vowels occurred in isolation or in context. When the vowel occurred in 1-syllable context, it was identified in a significantly shorter reaction time (*MD* = .71 sec, *p* < .000). When the vowel occurred in 2-syllable words, it was also identified in a significantly shorter reaction time (*MD* = 0.64 sec, *p* < .000). However, there were no significant in reaction time in vowel identification between 1- and 2-syllable words (*MD* = -.07 sec, p = 1). The findings suggest the likely easiness of vowel identification in context.

Table 130

Pairwise Comparisons of Vowel Identification and Reaction Time across Manipulations

Measure	Context (I)	Context (J)	Mean difference (I – J)	*Sig.* ^{b}
Identification	Context 1	Context 2	-.49*	.000
		Context 3	-.49*	.000
	Context 2	Context 3	.01	1.000
Reaction time	Context 1	Context 2	.71*	.000
		Context 3	.64*	.000
	Context 2	Context 3	-.07	1.000

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
^{b}. Adjustment for multiple comparisons: Bonferroni.
Context 1: Isolated vowel token
Context 2: Vowel in 1-syllable word token
Context 3: Vowel in 2-syllable word token

7.3.5 Variation in vowel in-context identification among proficiency groups.

A one-way between subjects ANOVA was conducted to compare the effect of speaking proficiency on vowel identification in 1- and 2-syllable word tokens conditions. There was one statistically significant difference for the vowel /ɒ/ in 2-syllable word tokens as determined by *F*(2,50) = 3.54, *p* < .05. A Bonferroni post hoc test revealed that mean identification of the vowel /ɒ/ among the high proficiency group was statistically significantly higher than among the mediocre proficiency participants, *MD* = .09, $p < .05$, with a small effect size of ES $\eta^2 = .12$, as shown in Table 131.

Table 131

Variation in Identification of Vowel in-Context between Proficiency Groups

Vowel	Context	*F*	*Sig.*	ES η^2
/iː/	1-syllable	1.38	.26	.05
	2-syllable	2.32	.11	.08
/e/	1-syllable	0.30	.75	.01
	2-syllable	0.44	.65	.02
/ɒ/	1-syllable	0.27	.76	.01
	2-syllable	**3.54***	**.04***	**.12**
/ʌ/	1-syllable	0.91	.41	.04
	2-syllable	0.28	.75	.01
/ɪ/	1-syllable	1.83	.17	.07
	2-syllable	0.33	.72	.01
/æ/	1-syllable	0.10	.91	.00
	2-syllable	0.85	.43	.03
/ɔː/	1-syllable	0.45	.64	.02
	2-syllable	0.53	.59	.02
/ɑː/	1-syllable	0.84	.44	.03
	2-syllable	2.05	.14	.08

Note. N= 53
df (between groups) = 2
df (within groups) = 50

Table 132

Pairwise Comparisons of Mean Vowel in-Context Identification for Vowel /ɒ/

Proficiency group (I)	Proficiency group (J)	Mean difference (I – J)	*Sig.* [b]
Poor	Mediocre	-.02	1.000
	High	-.09*	.035
Mediocre	High	-.07	.144

Note. Pairwise comparisons are based on estimated marginal means.
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

7.3.6 Vowel in-context identification and speaking proficiency.

Pearson product-moment correlation coefficient was performed to assess relationship between in-context vowel identification and speaking proficiency. As displayed in Table 133 below, there were several significant, positive correlations between the investigated variables. Following an ascending order, significant correlations were: (a) identification of /iː/ in 2-syllable words and speaking proficiency, r = .29, 2-tailed $p < .05$ (b) identification of /ʌ/ in 1-syllable words and speaking proficiency, r = .29, 2-tailed $p < .05$; (c) identification of /ɪ/ in 1-syllable words and speaking proficiency, r = .34, 2-tailed $p < .05$; and (d) identification of /ɒ/ in 2-syllable words, r = .41, 2-tailed $p < .01$. Taken together, there were few significant correlations between vowel in-context identification and speaking proficiency, suggesting further absence of substantial relationship between perceptual identification abilities and speaking proficiency. That is, an increase or decrease in participants' speaking proficiency may not correlate with their perceptual identification abilities.

Interestingly however, some interrelationships between vowel in-context identifications showed statistical significance, mainly between the vowels /ɑː/, /iː/ and /ɔː/, /ɪ/ and /iː/, and /æ/ and /ʌ/ in 1- and 2-syllable words. The more participants identified /ɑː/ in 1-syllable words, the more participants identified /iː/ in 1-syllable words (r = .55, 2-tailed $p < .01$) and 2-syllable words (r = .42, 2-tailed $p < .01$). Equally, the more participants identified /ɑː/ in 2-syllable words, the more participants identified /iː/ in 1-syllable words (r = .48, 2-tailed $p < .01$) and 2-syllable words (r = .41, 2-tailed $p < .01$). A similar pattern of relationships was observed with identification of /ɑː/ and /ɔː/ in words, with correlations ranging from .36 and .52, 2-tailed $p < .01$, for both 1- and 2-syllable words.

Taken together, significant correlations between vowel in-context identification in 1- and 2-syllable words were all positive, suggesting an overall increase in vowel identification in context.

Table 133

Matrix of Correlations among Vowel in-Context Identification and Speaking Proficiency

	/iː/		/e/		/ɒ/		/ʌ/		/ɪ/		/æ/		/ɔː/		/ɑː/		Pro.
/iː/	1	**.51****	.24	.10	.09	.25	.21	.16	**.56****	**.46****	.24	**.29***	**.41****	**.28***	**.55****	**.48****	.25
		1	**.49****	-.06	**.28***	**.60****	**.35****	.22	**.67****	**.34***	**.29***	.25	**.50****	.06	**.42****	**.41****	**.29***
/e/			1	.26	.10	**.35***	.14	.22	**.52****	-.13	-.03	**.33***	**.48****	.00	.09	.26	.06
				1	.08	-.10	.20	.12	.05	-.07	.09	.18	.01	-.06	-.12	.03	-.05
/ɒ/					1	**.40****	**.49****	.10	**.31***	.07	.19	.19	**.40****	.26	.22	**.34***	.11
						1	**.52****	**.38****	**.44****	.16	.23	**.41****	**.53****	.03	**.33***	**.47****	**.41****
/ʌ/							1	**.45****	.27	.07	**.31***	**.33***	**.39****	.07	**.31***	.19	**.29***
								1	.15	-.03	**.41****	**.70****	.27	.00	.11	.06	.13
/ɪ/									1	**.31***	**.33***	.23	**.51****	.20	**.37****	**.45****	**.34***
										1	.05	.13	.04	-.02	**.36****	.14	.19
/æ/											1	**.44****	.24	.02	.08	.06	.08
												1	**.39****	.10	.26	.13	.13
/ɔː/													1	**.45****	**.52****	**.47****	.12
														1	**.50****	**.36****	-.12
/ɑː/															1	**.54****	.21
																1	.25
Pro.																	1

Note. N = 53

**. Correlation significant at the 0.01 level (2-tailed)

*. Correlation significant at the 0.05 level (2-tailed)

Pro.: Proficiency shortened for word processing reason.

Rows and columns in grey correspond to vowel in 2-syllable word.

Chapter 8

Survey of Attitudes towards English Pronunciation

Introduction

We designed the survey to screen participants and explore their attitudes towards skilled pronunciation habits and their potential relationship to common learning practices to improve their pronunciation as an integrated element within speaking learning activities.

8.1 Description of the Survey

The survey consisted of four parts. We designed Part I essentially to collect data on participants' age and hearing medical history, including dyslexia, for research has demonstrated its potential effects on speech perception abilities (Messaoud-Galusi, Hazan, & Rosen, 2001). We designed Part II to collect data on participants' linguistic experience, from native language to foreign languages instruction received in middle and secondary school, specific extracurricular instruction in languages, travel history, and living experience in an English-speaking country. Additionally, we designed Part III to check participants for any significant extra-linguistic experience as music instruction, which may significantly relate to any probable improved sensitivity to sound perception. Finally, we designed Part IV to measure participants' attitudes towards several issues on importance of pronunciation, use of RP English, pronunciation feedback and active learning of pronunciation.

8.2 Procedure

Prior to survey administration, we informed participants about the nature of the task they had in hand and gave them uniform instructions on how to respond to questions and rate statements. We made participants aware of honesty required on their part to provide and report true information of themselves as a sine qua non of providing reliable data for analysis. For that reason, we informed participants about our research objectives and the utility of the survey in deciding on participants' eligibility for further research experimentation. We insisted on informing participants about confidentiality issue that their information would not be disclosed to a third party without their prior consent, and that reported information would not relate to or affect their course credits.

We administered the survey to participants in their free time to allow them sufficient time to fill it under no time constraints and provided diligent guidance in case of relevant questioning. Finally, we informed participants that there were no right or wrong answers to statements and all they had to do was to self-report what was true of them regarding any attitude object on a 5-point Likert scale, with each describing a degree of intensity of agreement, assertion, importance, or satisfaction. We accompanied participants in filling the survey that lasted for about 45 minutes and appreciated highly their efforts and collaboration.

8.3 Results

8.3.1 Descriptive statistics.

Overall survey results show participants favourable attitude towards most attitude objects at the affective, cognitive and behavioural levels. Participants showed a favourable attitude towards correct use of RP English, by reporting themselves to feel highly satisfied when accurately using RP English. More than 90% of the participants responded '*They feel either satisfied or very satisfied when they use correct RP English*,' Consistently, more than 90% of the participants responded '*They feel dissatisfied when classmates mispronounce RP English,*' and 93% of them responded '*They feel dissatisfied when they listen to distorted English.*' However, only 45% of participants responded '*They are appreciative when their teachers comment negatively on their pronunciation*,' and 38% expressed their dissatisfaction with their teachers commenting on their pronunciation. Participants' dissatisfaction with negative feedback likely reveals rejection of the latter in the form of a self-defence mechanism to maintain positive self-esteem and a high level of motivation. Less consistent with their affect towards correct use of RP English, 26% of participants reported skilled pronunciation as indispensable for a future of English, while 50% of them contended that skilled pronunciation is important.

Table 134

Mean Score of Participants per Item

Survey statements	*M*	*SD*
1. How do you feel when you use correct RP English?	4.38	0.79
2. *How do you feel when your classmates mispronounce RP English?* (-)	1.32	0.58
3. How do you feel when a teacher comments negatively on your pronunciation?	2.36	1.21
4. *How do you feel when you listen to distorted English input?* (-)	1.19	0.60
5. What do you think of possessing skilled RP pronunciation habits for a future teacher of English?	3.98	0.82
6. What do you think of possessing skilled RP pronunciation habits in comparison to accurate grammar of English?	2.98	0.77
7. What do you think of possessing skilled RP pronunciation habits in comparison to rich vocabulary?	3.51	0.95
8. What do you think of possessing skilled RP pronunciation habits in comparison to fluency?	3.15	0.74
9. What do you think of English material used in teaching phonetics and phonology of English?	4.11	0.89
10. What do you think of achieving a native-like accent?	4.04	0.85
11. What do you think of the need for native speakers around to improve one’s pronunciation?	4.70	0.61
12. What do you think of skilled English pronunciation if you want to integrate into an English speaking community?	4.06	0.82
13. Perception of English sounds is responsible for pronunciation problems.	3.66	0.83
14. What do you think of the need for effective pronunciation feedback?	4.62	0.53
15. How intense is your motivation to have skilled pronunciation habits?	4.38	0.84
16. *A teacher in middle and secondary schools does not need skilled English pronunciation habits.* (-)	1.66	0.85
17. How often do you train yourself at correct pronunciation?	3.45	0.97
18. How often do you listen to authentic English to improve your listening?	3.30	0.80
19. How often do you self-monitor your pronunciation?	2.83	0.99
20. How often do you interact within the group to check each other’s pronunciation?	2.30	0.97

Note. N = 53. *SD*: Standard Deviation. Values are rounded to second decimal. For statement 17, N = 52, due to one missing score.

Compared to grammar of English, 64% of participants reported skilled pronunciation habits as important as grammar, 19% as less important, and 15% as relatively more important. Whether the overall mean score of 3.82 (SD = 0.99) relates to participants’ actual perception of how important pronunciation is, participants' perception of pronunciation importance is not independent of a cumulative effect of the latter's negligence in early learning steps. As for richness of vocabulary, the case is relatively similar, with 58% of participants reporting pronunciation as important or less, suggesting a persistent perception of pronunciation. Further consolidation of this belief is currently revealed by 83% of participants reporting pronunciation as important as fluency or less, indicating a pattern of beliefs congruent with language elements. However, participants demonstrated a different pattern of positive beliefs regarding the possibility of achieving a native-like accent (*M* = 4.04, *SD* = 0.85) and the need for skilled pronunciation to integrate socially in an English-speaking community (*M* = 4.06, *SD* = 0.82).

8.3.2 Attitudinal factors in English pronunciation learning.

A better use of attitudinal items for further statistical analysis suggests the grouping of items in factors instead of considering them individually. For this reason, we considered 7 factors for a reliability test as follows:

(a) *Factor 1*: pronunciation importance (items: 1, 5, 6, 7, 8, 15, 16),

(b) *Factor 2*: pronunciation instruction and feedback (items: 3, 9, 14),

(c) *Factor 3*: mispronunciation (items: 2, 4),

(d) *Factor 4*: achievement of a native-like accent (item: 10),

(e) *Factor 5*: pronunciation importance in social integration (item: 12),

(f) *Factor 6*: importance of perceptual abilities in pronunciation mastery (item: 13), and

(g) *Factor 7*: attitude towards active learning of pronunciation (items: 17, 18, 19, 20).

We carried out a reliability test on the seven factor loadings. Cronbach's alphas for factors 1, 2, 3, and 7 were .69, .68, .76, and .81, respectively, demonstrating an acceptable reliability. Furthermore, the internal consistency of all items in the questionnaire was similarly satisfactory, with Cronbach's α = .73.

Table 135

Mean Attitudes of Participants

Factor	*M*	*SD*
Attitude towards importance of pronunciation (AIP)	3.82	0.99
Attitude towards pronunciation instruction and feedback (APIF)	3.70	1.33
Attitude towards mispronunciation (AM)	1.49	0.59
Attitude towards achievement of native-like accent (AANLA)	4.04	0.85
Attitude towards importance of pronunciation in social integration (APSI)	4.06	0.82
Attitude towards importance of perceptual abilities in pronunciation mastery (AIPAPM)	3.66	0.83
Attitude towards active learning of pronunciation (AALP)	2.97	1.03

Note. N = 53. *M*: Mean. *SD*: Scores for statements in italics followed by a minus sign in Table 134 above were reversed before summing factor score.

8.3.3 Attitudinal factors and English speaking proficiency.

To determine the relationship between participants' reported attitudes and their speaking proficiency, understood as their achievement in the *Speaking and Phonetics* course, we computed Pearson product-moment correlations. As displayed in Table 136 below, few correlations showed statistical significance. Among all attitudinal factors, there were only two positive correlations between attitudes and level of speaking proficiency. Participants' level of proficiency was found to correlate statistically significantly with attitude towards achievement of native-like accent, $r = .41$, 2-tailed $p < .01$, and attitude towards active learning of pronunciation, $r = .36$, 2-tailed $p < .05$. The matrix of correlations showed further internal, statistically significant relationships between attitudinal factors. There was a moderate, positive relationship between participants' reported attitude towards importance of pronunciation and mispronunciation, $r = .48$, 2-tailed $p < .01$, indicating substantial congruency among attitude elements. That is, the more the participant believes in the importance of pronunciation as an essential component of English language learning, the more likely they despise mispronunciation and avoid, therefore, ambivalence. Equally, attitude towards importance of English language pronunciation and attitude towards importance of English language pronunciation in integrating an English-speaking community were significantly, positively correlated, $r = .33$, 2-tailed $p < .05$, suggesting further congruency among investigated attitudinal elements and the probable instrumental learning and use of English pronunciation outside of the academic context. As regards pronunciation context, participants' reported attitude towards active learning of English pronunciation correlated significantly positively with their attitude towards pronunciation instruction and feedback, $r = .28$, 2-tailed $p < .05$, suggesting participants' recognition of the role of feedback and instruction in active learning of English pronunciation. Furthermore, there was a significant, positive correlation between participants' reported attitude towards achievement of native-like accent and the conception of the role of the self as an active agent in learning English pronunciation.

Table 136

Matrix of Correlations among Attitudinal Factors and Level of Speaking Proficiency

	AIP	APIF	AM	AANLA	AIPSI	AIPAPM	AALP	Proficiency
AIP	1	.199	**.477****	.215	**.325***	-.009	.129	-.111
APIF		1	.180	.256	.152	.024	**.278***	.247
AM			1	.205	.226	-.054	.066	.098
AANLA				1	.217	-.117	**.343***	**.414****
AIPSI					1	-.028	.186	-.019
AIPAPM						1	.246	-.061
AALP							1	**.335***
Proficiency								1

Note. N = 53
**. Correlation significant at the 0.01 level (2-tailed)
*. Correlation significant at the 0.05 level (2-tailed)

AIP: Attitude towards importance of pronunciation
APIF: Attitude towards pronunciation instruction and feedback
AM: Attitude towards mispronunciation
AANLA: Attitude towards achievement of native-like accent
AIPSI: Attitude towards importance of pronunciation in social integration
AIPAPM: Attitude towards importance of perceptual abilities in pronunciation mastery
AALP: Attitude towards active learning of pronunciation
Proficiency: Level of achievement in *Speaking and Phonetics* course

8.3.4 Variation in attitude across speaking proficiency groups.

To investigate further attitudinal factors in our sample, we divided participants into three groups as a function of their achievement level in *Speaking and Phonetics* course: (a) 20 poor achievers (10 ≤ course score < 12), (b) 21 mediocre achievers (12 ≤ course score < 14), and (c) 12 exceptional achievers (course score ≥ 14). Prior to grouping of participants, we equally considered methodological issues relating to robustness of the ANOVA test and scoring issues relating to variability and significance. We avoided (a) having groups of largely unequal size and (b) trivialising existing course scoring differentials, as the range of course scores was substantially narrow. Levene's test for homogeneity of variance between groups showed that the variances for attitudinal factors among the three proficiency groups were equal, as displayed in Table 137 below.

Table 137

Test of Homogeneity of Variances in Attitudinal Factors among Proficiency Groups

Attitudinal factor	Levene's statistic	*Sig.*
AIP	0.739	.483
APIF	0.542	.585
AM	0.278	.759
AANLA	1.765	.182
AIPSI	1.466	.241
AIPAPM	1.502	.233
AALP	0.965	.388

Note. Results are based on groups' arithmetic means.
*df*1 = 2
*df*2 = 50

A one-way between subjects ANOVA was conducted to compare means of attitudinal factors between the three proficiency groups. Tables 137 through 139 display descriptive and inferential statistics of attitudinal data among proficiency groups.

Table 138

Descriptive Statistics of Attitudinal Data between Proficiency Groups

					95% CI	
Attitudinal factor	Proficiency group	*M*	*SD*	*SE*	Lower bound	Upper bound
AIP	Poor	3.83	0.34	0.08	3.67	3.99
	Mediocre	3.88	0.42	0.09	3.69	4.08
	High	3.68	0.50	0.15	3.36	4.00
APIF	Poor	3.53	0.59	0.13	3.26	3.81
	Mediocre	3.78	0.51	0.11	3.55	4.01
	High	3.83	0.54	0.16	3.49	4.18
AM	Poor	4.60	0.68	0.15	4.28	4.60
	Mediocre	4.74	0.66	0.14	4.44	4.74
	High	4.79	0.40	0.11	4.54	4.79
AANLA	Poor	3.65	0.81	0.18	3.27	3.65
	Mediocre	4.19	0.93	0.20	3.77	4.19
	High	4.42	0.51	0.15	4.09	4.42
AIPSI	Poor	4.10	0.91	0.20	3.67	4.53
	Mediocre	3.95	0.74	0.16	3.62	4.29
	High	4.17	0.83	0.24	3.64	4.70
AIPAMP	Poor	3.70	0.66	0.15	3.39	4.01
	Mediocre	3.62	0.80	0.18	3.25	3.99
	High	3.67	1.15	0.33	2.93	4.40
AALP	Poor	2.76	0.52	0.12	2.52	3.00
	Mediocre	2.94	0.59	0.13	2.67	3.21
	High	3.38	0.65	0.19	2.96	3.79

Note. N = 53. CI: Confidence interval

Table 139

Variation in Attitudinal Factors between Proficiency Groups

Attitudinal factor	*F*	*Sig.*	ES η^2
Attitude towards importance of pronunciation (AIP)	0.964	.388	.04
Attitude towards pronunciation instruction and feedback (APIF)	1.499	.233	.06
Attitude towards mispronunciation (AM)	0.428	.654	.02
Attitude towards achievement of native-like accent (AANLA)	**3.990***	.025	.14
Attitude towards importance of pronunciation in social integration (AIPSI)	0.299	.743	.01
Attitude towards importance of perceptual abilities in pronunciation mastery (AIPAPM)	0.047	.954	.00
Attitude towards active learning of pronunciation (AALP)	**4.286***	.019	.15

Note. N = 53
df (between groups) = 2
df (within groups) = 50

There were statistically two significant differences between proficiency groups as determined by one-way ANOVA, $F(2, 50) = 3.990$, $p < .05$, $\eta^2 = .14$, for attitude towards achievement of native-like accent, and $F(2, 50) = 4.286$, $p < .05$, $\eta^2 = .15$, for attitude towards active learning of pronunciation. A Bonferroni post hoc test revealed that attitude towards achievement of native-like accent and active learning of pronunciation were statistically significantly higher in high achievers compared to poor achievers, with $MD = 0.776$, $p < .05$ and $MD = 0.612$, $p < .05$, respectively. However, there were no statistically significant differences between mediocre and high achievers nor between mediocre and poor achievers among all attitudinal factors.

Table 140

Pairwise Comparisons of Attitude Means between Proficiency Groups

						95% CI	
Attitudinal factor	Group (I)	Group (J)	Mean difference (I-J)	*SE*	*Sig.*	Lower bound	Upper bound
AIP	Poor	Mediocre	-0.056	0.129	1.000	-0.375	0.263
		High	0.150	0.151	.972	-0.223	0.523
	Mediocre	High	0.206	0.149	.522	-0.164	0.575
APIF	Poor	Mediocre	-0.244	0.171	.476	-0.668	0.179
		High	-0.300	0.200	.417	-0.794	0.194
	Mediocre	High	-0.056	0.198	1.000	-0.546	0.434
AM	Poor	Mediocre	-0.138	0.194	1.000	-0.619	0.343
		High	-0.192	0.227	1.000	-0.754	0.371
	Mediocre	High	-0.054	0.225	1.000	-0.611	0.504
AANLA	Poor	Mediocre	-0.540	0.253	.112	-1.166	0.085
		High	**-0.766**[*]	0.295	.037	-1.498	-0.035
	Mediocre	High	-0.226	0.293	1.000	-0.951	0.499
APSI	Poor	Mediocre	-0.148	0.259	1.000	-0.495	0.790
		High	-0.067	0.303	1.000	-0.817	0.684
	Mediocre	High	-0.214	0.300	1.000	-0.958	0.529
AIPAMP	Poor	Mediocre	0.081	0.264	1.000	-0.574	0.736
		High	0.033	0.309	1.000	-0.732	0.799
	Mediocre	High	-0.048	0.306	1.000	-0.806	0.711
AALP	Poor	Mediocre	-0.178	0.180	.983	-0.624	0.268
		High	**-0.612**[*]	0.210	.016	-1.134	-0.091
	Mediocre	High	-0.435	0.209	.127	-0.951	0.082

Note. Pairwise comparisons are based on calculated arithmetic means.

*. The mean difference is significant at the .05 level.

[b]. Adjustment for multiple comparisons: Bonferroni.

8.3.5 Power of attitudinal factors in predicting speaking proficiency.

We performed a multiple linear regression to predict participants' speaking proficiency, as understood as their level of achievement in the *Speaking and Phonetics* course, based on the seven attitudinal factors. Results of the multiple linear regression indicated there was a significant effect between attitudinal factors and level of proficiency, $F(7, 45) = 2.862, p < .016$, with an R^2 of .308. Attitudinal factors combined accounted for 30.8% of the variance in the participants' speaking proficiency. Individual predictors were further examined and indicated that attitude towards achievement of a native-like accent statistically significantly predicted the level of proficiency ($t = 2.404, p < .021$). Except for the latter, other attitudinal factors did not significantly predict participants' level of proficiency. Nonetheless, attitude towards importance of pronunciation and attitude towards active learning of pronunciation predicted participants' level of speaking proficiency at the .07 and .10 level ($t = -1.876, p < .07$; $t = 1.727$, $p < .10$), respectively. Regardless of their attested explanatory power, attitudinal factors towards pronunciation were poor predictors of participants' speaking proficiency.

Table 141

Summary of the Regression Model

Model	R	R^2	Adjusted R^2	*SE*
1	.555^a	.308	.200	.820

Note. N = 53
^{a}. Predictors: Constant, AALP, AM, AIPAMP, APSI, APIF, AANLA, AIP
Dependent variable: Speaking proficiency

Table 142

Goodness of Fit of the Regression Model

Model		*SS*	*df*	*MS*	*F*	*Sig.*
1	Regression	13.482	7	1.926	2.862	.015
	Residual	30.278	45	.673		
	Total	43.760	52			

Note. N =53
Dependent variable: Speaking proficiency
Predictors: Constant, AALP, AM, AIPAMP, APSI, APIF, AANLA, AIP

Table 143

Predictive Power of Attitudinal Factors in Forecasting Level of Speaking Proficiency

		Unstandardized coefficients		Standardized coefficients			95% CI	
Model		*B*	*SE*	*Beta*	*t*	*Sig.*	Lower bound	Upper bound
1	Constant	11.956	1.385		8.630	.000	9.165	14.746
	AIP	-.0614	.328	-.276	-1.876	.067	-1.274	.045
	APIF	.234	.2213	.141	1.056	.297	-.212	.679
	AM	.205	.213	.138	.963	.341	-.224	.635
	AANLA	**.362***	.151	.337	2.404	.020	.059	.666
	APSI	-.116	.150	-.103	-.772	.444	-.419	.187
	AIPAMP	-.091	.145	-.038	-.630	.532	-.384	.201
	AALP	.369	.214	.246	1.727	.091	-.061	.800

Note. N = 53
Dependent variable: Speaking proficiency
Predictors: Constant, AALP, AM, AIPAMP, APSI, APIF, AANLA, AIP

Chapter 9

Discussions

Introduction

This chapter discusses the findings reported in the previous chapters (5, 6, 7, and 8) in an effort to provide answers to the four main research questions of the current work as cited in Section 4.1. We will discuss those findings relating to the perception of RP English vowel contrasts in various experimental conditions, making use of arguments in connection with psychology and linguistics. Along our interpretation of perception findings, we will refer to the probable role of the participants' native language and their second language learned at school, French, in shaping some of perceptual patterns exhibited by participants, bearing essentially on phonological features of Arabic, being their L1, and French language that they have been learning since the age of 9[39]. Then, we will proceed with a discussion and interpretation of attitudinal findings typical to the Algerian context and compare them with available findings in other learning contexts.

As stated previously in the opening pages of the research, a considerable body of speech research has continuously emphasised the precedence of speech perception as a sine qua non of developing proficient and less accented speech in an L2. In order to determine participants' abilities in perceiving RP English vowels, we relied exclusively on speech perception common research tests, namely discrimination and identification. In the present section, we discuss and interpret our results in light of theoretical research frameworks outlined above and some empirical findings in the field. The discussion and analysis follow the same pattern of previously presented experiments.

---

[39] Figures 51 and 52 show the plot of the two first formants for Arabic and French language, as adapted from the work of S. H. Al-Ani (1980) and J-P. Tubach (1988). They are used as reference models for our present discussion of results.

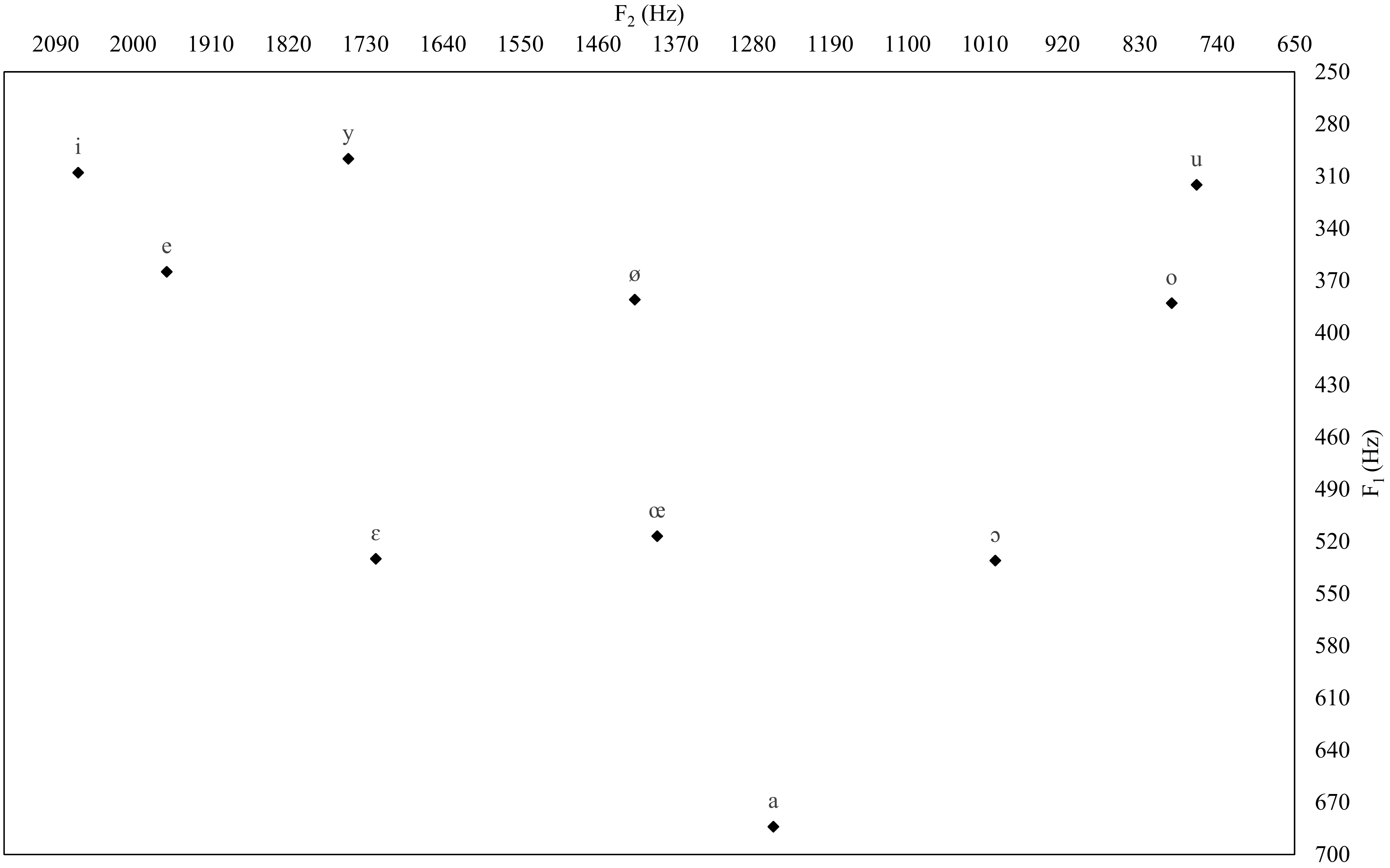


Figure 51. Plot of the two first formants for French vowels. From “La parole et son traitement automatique” by J-P. Tubach. Copyright 1988 by Mason. Adapted with permission.

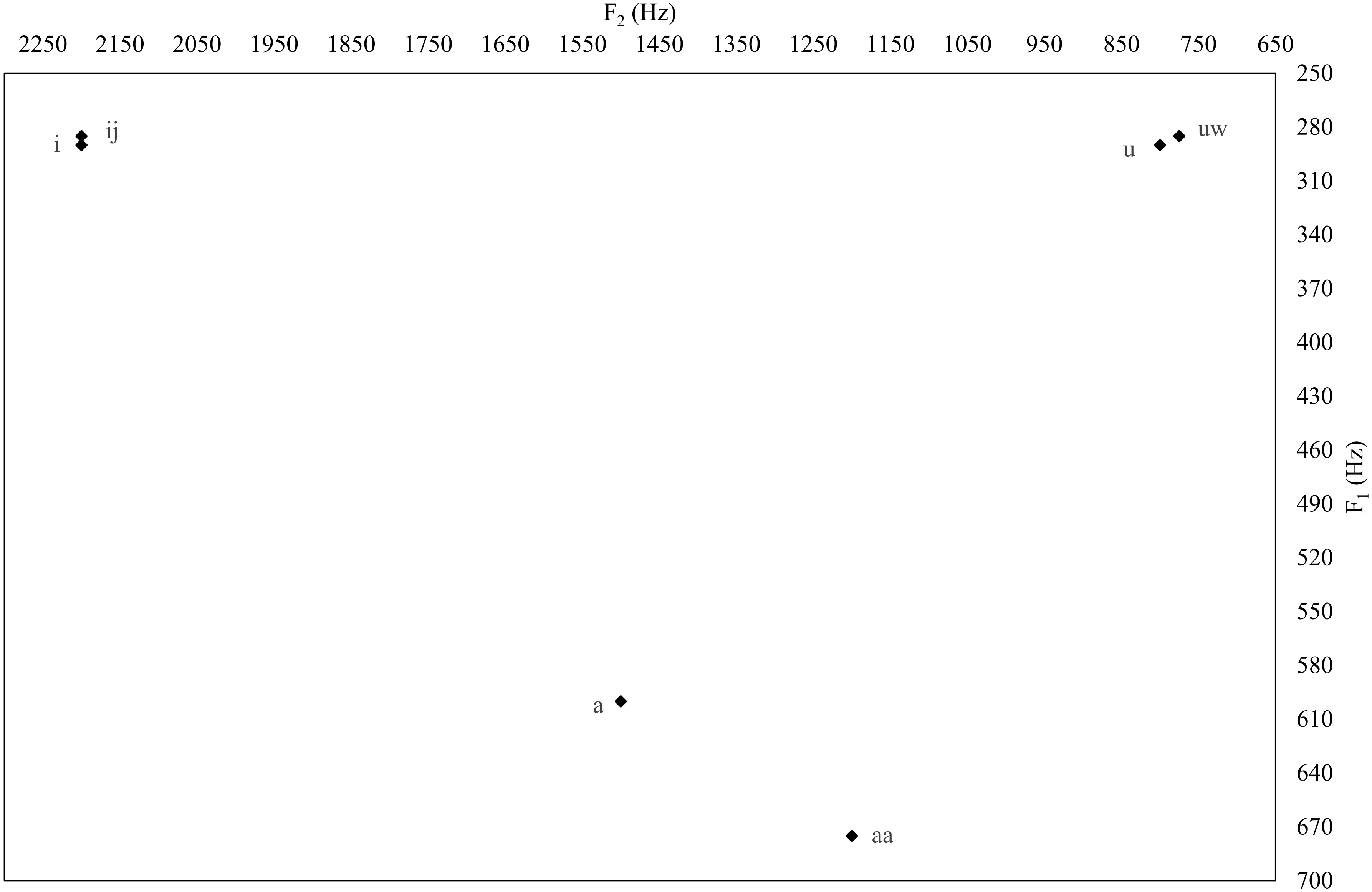


Figure 52. Plot of the two first formants for the Arabic prototypic vowels. From S. H. Al-Ani . Copyright 1980. Adapted with permission

9.1 Discussion of Discrimination Results

Discrimination findings obtained in the A*X* task revealed that Algerian learners of English, with a considerable experience with English language input and instruction, demonstrated high sensitivity indices for almost all RP English vowel contrasts. Starting with the vowel contrast in *Series 1* (/ɪ/ vs. /iː/), participants demonstrated a high perceptual sensitivity to their spectral differences. Although both prototypic vowels had an equal duration of 180 ms, participants could focus their attention on the spectral differences in the stimuli to assess their relative similarity or difference. A sensitivity index *d'* of 5.08 is very high and indicates a strong ability of discrimination of the contrasted vowels that are located in the high front area on the perceptual map. In terms of the Perceptual Assimilation Model for Best (1995), this pattern of excellent discrimination is indicative of a *Two-Category Assimilation (TC Type)*, where non-native contrasts are assimilated to different L1 categories. However, this explanation may not apply adequately here, as the learners' L1 similar vowel categories do not exhibit similar spectral differences, as the Arabic language exploits very economically the high front space on the perceptual map. That is, the Arabic vowels that are located in the high front area correspond to /i/ and /ij/ and have nearly identical spectral features to RP English prototypic vowel /iː/ (F_1 290 Hz, F_2 2200 Hz), as measured by Al-Ani (1970), and their discrimination depends largely on a temporal effect only. In other words, the Arabic vowels /i/ and /ij/ differ in the temporal ratio or the long-to-short ratio that is almost equal to 2.0, with /i/ being produced at 300 ms and /ij/ at 600 ms, respectively (Al-Ani, 1970). It should be emphasised that the high sensitivity Algerian learners of English demonstrated to this vowel contrast may not be exceptional. In a study of the interrelationship between the perception and production of English vowels, Rauber, Escudero, Bion, and Baptista (2005) reported that, amongst all tested English vowel contrasts, native speakers of Brazilian Portuguese discriminated most accurately between /ɪ/ and /iː/ with a discrimination rate of .94, though Brazilian Portuguese contains only the high front vowel /i/

(lower in both F_1 and F_2 than its English counterpart). These discrimination rates of the contrast, /ɪ/ vs. /iː/, are almost identical to an excellent discrimination rate of .99 for the same contrast among Polish advanced learners of English, whose mother tongue includes a qualitative vowel contrast, /i/ vs. /ɨ/, with the latter vowel located in the high and central area on the perceptual map, respectively (Balas, 2018).

Learners' experience with French as an L2 may help explain this high sensitivity to the spectral difference alone in the RP English vowel contrast, but remains very speculative and empirically unfounded. Given the fact that the French vowel /i/ is transcribed phonetically in a similar way to the Arabic one, it actually has different spectral properties from its counterparts in both Arabic and English, as measured by Tubach (1989) (F_1 320 Hz, F_2 3200 Hz). Exposed to French at an early age, it is likely that Algerian learners of English could distinguish the variation in the phonetic realisation of /i/ in French and Arabic by relying on spectral features alone. The effect of some perceptual advantage in relying on spectral differences alone deriving from probably Berber varieties is less likely expected as nearly all participants, including those with contact with Berber verities, scored highly on this vowel contrast (*M* = .98, *SD* = .02).

This finding supports Flege's (1987a, b) claim that new vowels in an L2 that do not exist in the L1 would be easy to learn, and this likely involves both perception and production. The statistically significant differences in mean identification of the ambiguous stimuli located between the prototypic /ɪ/ and /iː/ indicate the phenomenon of categorical perception. However, the present empirical finding may not be enough to tell which of the RP prototypic vowels may occupy a larger perceptual spectral space, as minute manipulations of the ambiguous stimuli are required to locate precisely the boundary category.

The prototypic vowel contrast in *Series 2* (/ʊ/ vs. /uː/) indicates a different finding. The Algerian learners of English demonstrated less sensitivity to the pair (/ʊ/ vs. /uː/). Pooled sensitivity *d'* of 1.87 for (/ʊ/ vs. /uː/) is far below that for (/ɪ/ vs. /iː/), *d'* of 5.08, indicating a qualitative change in use of spectral difference in discriminating the former RP English prototypic vowels as a function of location on the perceptual map. The little effect demonstrated by Algerian learners to discern spectral differences in the pair (/ʊ/ vs. /uː/) may be due to developmental processes in their L1 and their early experience with French as an L2. Arabic language has a similar vowel contrast to the English (/ʊ/ vs. /uː/), which demonstrates equally qualitative and quantitative differences. Transcribed phonetically as /u/ and /uw/, this Arabic vowel contrast differs simply on a quantitative dimension, that is, /uw/ takes twice the duration of /u/ and both share nearly identical spectral qualities (F_1 290, F_2 800 Hz) as measured by Al-Ani (1970). The similarity at the spectral level between the Arabic vowel pair /u/ and /uw/ and the French vowel /u/ (F_1 315 Hz, F_2 764 Hz), as measured by Tubach (1989), might have caused the learners lose their sensitivity to spectral differences in this area and keep perceptual effect for temporal characteristics only, as French does not exploit temporal characteristics as a major cue to distinguish vowels (Gottfried & Beddor, 1988). The distinction between (/ʊ/ vs. /uː/) falls better within the *Single-Category Assimilation (SC Type)*, where two L2 sounds are assimilated to the same native category but are discrepant from the prototypic form of it. Nonetheless, as suggested by Best (1995), sensitivity is poor and discrimination may go above chance level. Equally, this finding provides empirical support to Flege’s (1987a, b) SLM model, inasmuch as it predicts considerable perceptual difficulty in discerning differences between two similar sounds in L1 and L2.

For vowel stimuli in *Series 3* (/ɔː/ vs. /ɒ/), the results indicate a high sensitivity to the spectral differences between the vowel contrast. As learners’L1 lacks these vowels in its vowel inventory, it may be that Algerian learners have kept their discrimination sensitivity in the mid-high and low back area on the perceptual map, while demonstrating some preference of reliance on temporal features in vowel perception in this area, regardless of any presumed age effect or perceptual development. Additionally, the French vowel contrast (/o/ vs. /ɔ/ as in /bo/ vs. /bɔl/, respectively) may be responsible for this maintained perceptual sensitivity, which is likely why the Algerian learners could discern the qualitative differences for this RP English prototypic vowel pair.

The sensitivity indices for the pair of ambiguous vowel sounds (14_15 & 15_14 and 15_16 & 16_15) were significantly different (*MD* = .41, *p* < .000), though the spectral distance in both combined pairs is equal, which suggests a categorical perceptual effect, with probably the vowel /ɔː/ occupying a larger perceptual space than /ɒ/. This finding provides further confirmation for Algerian learners' preference for temporal effect in perception. That is, in the absence of similar L1 sounds to refer to when discriminating vowels, Algerian learners rely more on quantitative features in the discrimination of the vowel rather than the qualitative ones, i.e. the spectral properties. This suggestion should not come as a surprise as Algerian learners' L1 exploits less the qualitative dimension on the perceptual map and makes up for it by using the temporal dimension. Nonetheless, it may be argued that the perceptual assimilation pattern for this pair is of the *Two-Category Assimilation Type (TC Type),* as suggested by Best (1995).

Vowel contrast in *Series 4* (/ɑː/ vs. /ʌ/) is the most singular pair amongst all. The findings suggest complete insensitivity to discriminate the spectral difference between the two RP prototypic vowels. The probably small spectral distance between these two contrasted vowels may be the cause of that insensitivity. The perceptual assimilation pattern falls well within then *Single Category Assimilation (SC Type)* as predicted by the Perceptual Assimilation Model of L2 phonological acquisition, where L2 contrasts are assimilated within the same L1 category, and that they are both equally good exemplars of the L1 sound or both equally deviant. In this view, it is quite hard to determine whether /ɑː/ and /ʌ/ are assimilated equally as good exemplars or deviant of the Arabic vowel /aa/ or the French vowel /a/. It is to emphasise here that Algerian learners of English linguistic experience makes them have qualitatively four similar vowels in the central low and mid low area on the perceptual map (where spectral distance among vowels along F_1 and F_2 is not considerable), creating a crowded vowel space for perception that is likely difficult to improve with learning experience. A similar finding to ours was reported by Flege (2003b), who claimed that native Italian adults initially "have great difficulty discriminating English /ɒ/-/ʌ/ because instances of both English vowels tend to be assimilated by a single Italian vowel (usually /a/)" and that "single-category assimilation appears to persist for some late learners over many decades of English use" (p. 339).

The high sensitivity displayed in discriminating the qualitative differences in the prototypic vowel contrast in *Series 5* (/ɑː/ vs. /æ/) provides support for our previous claim. A large difference along the F_2 dimension was enough for Algerian learners to discriminate (/ɑː/ vs. /æ/) with a pooled sensitivity index *d'* of 4.63 and demonstrate, through considerable discrimination of other ambiguous synthetic vowel token between the pair prototypic vowels, a seemingly categorical perception effect midway between the synthetic sounds 25 and 27. One further comment to make is that, if the spectral distance between /ɑː/ and /æ/ is enough for Algerian learners to discriminate them, then it is likely to suggest that they will be able to discriminate between their L1 /aa/ (F_1 675 Hz, F_2 1200 Hz) and its French counterpart /a/ (F_1 684 Hz, F_2 1256 Hz) with a high sensitivity. Further support to the role of spectral difference in discriminating RP English prototypic vowel contrasts so far is found in Series 6 (/æ/ vs. /ʌ/), where Algerian learners demonstrated an important pooled sensitivity index *d'* of 2.84. As displayed in Table 51, the mean discrimination rate for (/ɑː/ vs. /æ/) was statistically higher than mean discrimination for (/æ/ vs. /ʌ/) ($MD = .11, p < .01$), offering much evidence for the previous claim that the larger the spectral difference, the higher the discriminate rate and the perceptual sensitivity as well. This time again, the pair (/æ/ vs. /ʌ/) falls likely under the *Two-Category Assimilation (TC Type).*

With respect to the prototypic pair in *Series 7* (/e/ vs. /ɜː/), pooled sensitivity index *d'* of 2.08 indicates a considerable perceptual ability among Algerian learners to discriminate accurately the spectral difference between vowels in the contrast. Nonetheless, the null-pooled sensitivity index near the centre of the spectral distance between /e/ vs. /ɜː/ indicates the small perceptual space occupied by these prototypic vowels. Interestingly, when compared to the contrast (/e/ vs. /ɜː/) that has a similar F_2, the pair (/ɜː/ vs. /ʊ/) with similar F_1 is discriminated with a higher sensitivity index *d'* of 5.76. This fact demonstrates probably learners' higher perceptual sensitivity in discriminating contrasts when the spectral distance is more important in F_1 than in F_2.

The higher sensitivity index for the pair in *Series 8* (/ɜː/ vs. /ʊ/) suggests that these prototypic vowels are assimilated to two different L1 categories as predicted by the *Two-Category Assimilation (TC Type).* However, Algerian learners' L1 does not have analogous sound to the English vowel /ɜː/. In this case, it is likely that learners' discrimination would have been based on vowel sounds that do not exist in their L1 and would therefore belong to the French vowel inventory such as the vowel /œ/ in the word /kœʁ/ (meaning 'heart') or the vowel /e/ as in the word /le/ (meaning 'the'). Given the lack of substantial evidence in favour of our above-mentioned explanation, we further suggest that, in the absence of a long-term representation of vowels, vowel stimuli are likely processed on the spot, and discrimination is likely realised through a cognitive effort as a function of simulation of articulatory movements required for vowel stimuli production.

Finally, the RP prototypic vowel contrast in *Series 9* (/ɜː/ vs. /uː/) demonstrated once more significant perceptual abilities showed by Algerian learners in discriminating some RP English vowels non-existent in their L1 vowel inventory. It is to emphasise herein that the presumed similarity between the Arabic and English vowel pairs (/u/ and /uw/ vs. /ʊ/ and /uː/), is indeed misleading, as it is based on phonetic notation only. The spectral differences between these vowel pairs is visibly important and should not be trivialised on whatever ground, as illustrated in the perceptual map in *Figure 29*. Respective spectral difference between /u/ and

/ʊ/ is 105 Hz in F_1 and 607 Hz in F_2 and 101 Hz in F_1 and 812 Hz in F_2 between /uw/ and /uː/. Importantly, it is to recall herein that the Algerian learners' perceptual preference for important spectral distance in F_1 than F_2 to discriminate accurately vowel contrasts is further confirmed.

In sum, overall results suggest a high perceptual sensitivity among Algerian learners of English to discriminate several RP English prototypic vowel contrasts, except for the contrast (/ɑː/ vs. /ʌ/) whose sensitivity index *d'* was null. In view of PAM or SLM's frameworks, the obtained results suggest the existence of English phonetic categories within the Algerian learners' perceptual space. Learners' perceptual abilities and sensitivity measurements showed a similar pattern of findings for the pairs (/ɜː/ vs. /ʊ/), (/ɜː/ vs. /uː/), (/ɪ vs. /iː/), (/æ/ vs. /ɑː/), and (/ɔː/ vs. /ʌ/). However, other pairs showed a different pattern of discrimination: (/æ/ vs. /ʌ/) > (/e/ vs. /ɜː/) > (/uː/ vs. /ʊ/) > (/ɑː/ vs. /ʌ/).

As the origins of these considerable discrimination abilities necessitate plausible explanations, it is to note that two possible explanations are to advance indeed. First, Algerian learners' discrimination abilities are likely to stem from learners' rich linguistic experience, bearing on Arabic and French vowel inventories, learned prior to pubescent age. The second plausible explanation is that Algerian learners' L1 phonological system, with its economic use of the spectral space on the perceptual map, offers its users the perceptual advantage to form new L2 vowel with further language learning experience. The empirical findings reported herein do not support the NLM hypothesis that L1 speech learning may lead to warping or filtering of L2 input, causing therefore difficulty in an L2 sound formation. As regards the present research objectives for the same different discrimination test, the obtained empirical findings provide evidence for the Algerian learners' highly considerable perceptual abilities in discriminating accurately between RP English prototypic vowels, relying most exclusively on intrinsic spectral differences in the presented stimuli. However, it was clearly demonstrated that empirical threshold of sensitivity was rather a function of the vowel contrast and its position on the perceptual map, with significantly higher discrimination hit rates strongly correlated with (a) differences in F_1 rather than in F_2 and (b) occupation of a larger perceptual space by the vowel compared to another.

Understood as achievement in *Speaking and Phonetics* course, Algerian learners' speaking proficiency in English seems not to correlate with their perceptual discrimination abilities, likely calling into question both the relevance and precedence of speech perception to speech production. Our study findings do not corroborate Flege's (1992) Speech Learning Model (SLM) positing that accurate perception of an L2 sounds is required for their accurate production. However, our findings are not inconsistent with Zampini's (1998) findings denying the strong correlation between learners' perceptual and production abilities. Zampini's (1998) study, involving the perception and production of Spanish stops among a group of adult, native English speakers, showed that the latter's perception and production of Spanish stops did not strongly correlate. Some of the adult, native English speakers with Spanish-like perceptual abilities did not exhibit parallel or corresponding production abilities, as some of them could identify Spanish stops at shorter VOT boundaries and produce them with longer VOT and vice versa. Providing further support to our findings and previous claim, an analysis of variance of discrimination abilities demonstrated important, non-significant overlapping differences in discrimination of combined pairs of synthetic stimuli among the three speaking proficiency groups, as discrimination hit rates overlapped across vowel contrasts.

To conclude, we believe there is sizeable evidence to suggest that Algerian learners' inventory size does not seem to reduce their discrimination abilities of English vowel contrasts based on spectral contrasts alone. The small number of vowels in the Algerian learners' L1 has a non-significant effect on the use of cue-weighting perceptual processing or strategies in English vowel perception. Our statement is indeed consistent with Elvin, Escudero, and Vasiliev's (2014) claim that actually "cross-linguistic acoustic properties, rather than cross-linguistic vowel inventory sizes, successfully predict non-native discrimination difficulty" (p. 1). Whether special attention should be given to the possible effect of French vowel system has on Algerian learners' discrimination abilities, this remains a legitimate inquiry open to debate and rather requires a longitudinal investigation within a developmental perspective.

9.2 Discussion of Isolated Vowel Tokens Identification Results

Findings of isolated vowel tokens identification experiment indicate different perceptual abilities of Algerian learners in accurately identifying the synthetic vowel tokens, demonstrating considerable variability across vowels and temporal manipulations. The low sensitivity indices for vowel identification indicate a constant perceptual difficulty and confusion in identifying accurately short vowels compared to their long counterparts. Indeed, short vowel identification rates seem to be rather static across all temporal manipulations. However, with long vowels, identification seems to improve as a function of temporal manipulation. That is, the longer the duration of the vowel, the more it is accurately identified. Learners' perceptual abilities in identifying short vowels indicate this identification pattern: /ʊ/ > /ɪ/ > /e/ > /ʌ/ > /æ/ > /ɒ/, with the vowel /ɒ/ being identified significantly at the lowest rate amongst all. However, learners' perceptual abilities indicate this identification pattern among long or tense vowels: /iː/ > /ɔː/ > /ɑː/ > /ɜː/ > /uː/, with the vowel /uː/ identified significantly at the lowest rate amongst all.

Given the fact that Algerian learners' L1 vowel system makes use of both spectral and temporal contrasts, it is to note that Algerian learners of English demonstrated a varied pattern of use of both spectral and temporal cues in isolated vowel tokens' identification. Obtained findings isolated vowel tokens' identification suggest significant differences between perceptual abilities of learners in discriminating spectral differences in RP English pairs of prototypic vowels and in identifying them as a function of their temporal variability. Learners' discrimination abilities are far much better than their identification abilities. Difficulty in identification of vowels manifests itself in misidentification and confusion. As displayed in confusion matrices, there was not misidentification of prototypic vowels along their temporal properties alone, but there was confusion across vowel categories, suggesting important difficulty of the task and serious identification overlapping for all vowels. Vowel misidentifications included several confusion pairs such as (/ɪ/ and /iː/), (/ʊ/ and /uː/), (/ɪ/ and /ə/), (/uː/ and /e/), (/ʊ/ and /e/), (/e/ and /ə/), (/ʌ/ and /ɒ/), (/ɒ/ and /æ/), (/ʌ/ and /æ/), (/ɒ/ and /ɔː/), (/æ/ and /e/), (/ʌ/ and /e/), (/ɑː/ and /æ/), and few other negligible misidentifications. Although discrimination of the spectral differences in some of the above mentioned vowel contrasts proved excellent, identification of individual prototypic vowels proved poor. Nonetheless, some precision is necessary to state here with regard to short vowel identification compared to long ones. The short vowels demonstrated a high variability in identification and confusion as well. That is, the short vowels were not only confused with their long counterparts, but were also confused across each other. Some of the vowel confusions along the vowel temporal feature were (/ɪ/ and /iː/), (/ʊ/ and /uː/), (/ɒ/ and /ɔː/), and (/ɑː/ and /æ/). Some of the cross-vowel confusions were (/ɪ/ and /ə/), (/uː/ and /e/), (/ʊ/ and /e/), (/ʌ/ and /e/), and (/ʌ/ and /ɒ/).

Aggregate mean identification of isolated vowel tokens in the 7-*AFC* test demonstrated an overreliance on temporal cues in accurate identification of vowels tokens. To begin, the vowel token /iː/ was accurately identified at a considerable hit rate of .69 compared to other vowel tokens in the experiment set. Misidentification of /iː/ with /ɪ/ in 26% of the cases indicates considerable difficulty Algerian learners had in identification of /iː/ at shorter durations, likely suggesting both (a) the inherent differences in discrimination and identification perceptual processes and (b) the relative difficulty of an L2 vowel identification. Similarly, participants identified accurately the vowel /ɪ/ with a hit rate of .34 and misidentified it with /ə/, /e/ and /ʊ/ in 24, 19 and 15% of the cases, respectively, offering further confirmation for difficulty of L2 vowel tokens identification. It is to emphasise that misidentification of /ɪ/ with /ə/ and /ʊ/ is unsurprising, as the vowel /ɪ/ is used to transcribe a range of vowel sounds close to /ɪ/ and /ʊ/ when they are not word-final (Flemming & Johnson, 2007), reflecting the probable overlap in the range of qualities used for these three vowels. Though this fact applies specifically to reduced vowels in American English, it is to recall that 29 participants reported themselves as being more familiar to American English than RP English. Therefore, there is reason to believe that frequent exposure to American English might have attuned participants' perceptual abilities in favour of American English phonological system. This pattern of identification is quite similar with the vowel /ʊ/, as it was misidentified with /uː/ (the long or tense counterpart), /ə/ (spectrally similar) and /e/ (relatively spectrally different) in 8, 15 and 29%, respectively. The large misidentification of /ʊ/ with /e/ is hard to explain in terms of spectral similarities or vowel space overlap alone and may require a special attention. Although the vowel case /e/ was accurately identified with a hit rate of only .33, it was hardly unpredictably with /ɜː/ and /ə/ in 10 and 48% of the cases, respectively. Participants might have misidentified the vowel /e/ with (a) the vowel /ə/ because of this latter's spectral high variability and (b) the vowel /ɜː/ because of the spectral similarity along F_1.

Similar pattern of low accurate identification rates were found among the vowels /ɒ/, /ʌ/ and /æ/. In fact, the vowel /ʌ/ was identified accurately with a hit rate of .25 only, while it was misidentified with /æ/ and /e/ in 27 and 42% of the cases, respectively. Being the most singular vowel amongst all, the vowel /ɒ/ was identified accurately with the least hit rate of .07, while it was misidentified with /æ/ and /ʌ/ in 35 and 45% of the cases. Following the same identification tendency, the vowel/æ/ was identified accurately with a hit rate of .17, while it was misidentified with /ʌ/ and /e/ in 16 and 65% of the cases. These findings suggest that Algerian learners had a considerable difficulty in vowel tokens' identification of vowel tokens that are located in the low area on the perceptual map. Although the tense vowels /ɑː/ and /ɔː/

were accurately identified with higher rates of .58 and .60, they were misidentified with vowels located in the low area on the perceptual map, namely /æ/ and /ʌ/ for /ɑː/ and /ɒ/ for /ɔː/, with the latter misidentified largely with its short or lax counterpart.

Based on these empirical findings, it is to suggest that learners' identification abilities vary between discrimination and identification. Regardless of the low rate of identification of isolated vowel tokens, the findings give support and empirical evidence in favour of the psychoacoustic theories of speech perception, as they claim a dual perceptual system for speech. That is, perception of speech elements is subject to a dual system, the first of which is purely auditory in which auditory stimuli is processed within a general modality, called auditory processing mode. However, the second is subject to a processing based on long-term memory representations of these speech elements, called phonetic processing mode (Pisoni, 1973). Therefore, the findings would not come as a surprise and difficulty in vowel tokens' identification would not be a function of vowel recognition alone, but also of a memory decay problem. Further support for this claim is found in improvement of vowel identification with the so-called long vowels. This is to confirm that perception of speech elements does not depend only on an auditory mode, but extends to a whole dynamics of physical properties of the speech (acoustic aspect) and its abstract representations in memory, likely to suggest the presence of an intermediate cognitive interface mediating processing of physical characteristics of stimuli and corresponding identification. Phonetic processing is a function of stored representations of L1 sound categories or possibly L2 as well. Vowel identification findings provided here depicts a complex picture about the way Algerian learners identified RP English prototypic vowels, as to whether they relied on their L1 (Arabic) vowel abstract representations, their L2 (French) vowel long-term representations, or even likely abstract representations of RP English vowels. Nonetheless, it is to admit that it is difficult to incline for either of the claims for discrimination data proved rather excellent. In addition, the intrinsic differences in the three phonological systems of Arabic and French make it difficult to decide. As regards the previous claim, it is to suggest that Algerian learners may have a perceptual advantage in using both spectral and temporal information in processing RP English prototypic vowels. Originating in spectral and temporal contrasts from Arabic and spectral contrasts in French (having a large vowel inventory size), it is reasonable to believe that Algerian learners might have developed a perceptual system that is capable of exploiting and integrating temporal and spectral information in a complex way to processing vowels, with preference towards weighting temporal cues in identification.

Regarding speaking proficiency, it is to note that two central findings are to be reported herein. First, the correlations between isolated vowel tokens' identification and Algerian learners' speaking proficiency were not significant. Second, significant differences in isolated vowel tokens' identification among the three proficiency groups were not found. Our findings corroborate Cebrian's (2006) findings on the effect of L2 experience in identification of English vowel continuum from /iː/ to /ɪ/ by native Catalan speakers. Differences in identification of English vowel continuum among English and non-English learners, were found non-significant, as both groups of learners relied more on temporal cues in identifying /iː/ to /ɪ/, in contrast to native English speakers. Unfortunately, these findings lend less empirical support to the role and effect of L2 experience in development of perceptual abilities among L2 learners.

Contrary to discrimination reaction time findings in the discrimination test, reaction time in synthetic vowel tokens identification seems to vary systematically across vowels and temporal manipulations. Reaction time pattern for isolated vowels' identification followed this pattern: /e/ > /ɜː/ > /uː/ > /ɪ/ > /ʊ/ > /ʌ/ > /ɒ/ > /æ/ > /ɑː/ > /iː/ > /ɔː/, with the latter identified in the shortest time around 2.53 sec. Once more, reaction time indicated a perceptual preference among Algerian learners of English in identifying the three long or tense vowels /ɔː/, /iː/ and /ɑː/ compared to the two other tense vowels /uː/ and /ɜː/ and other short vowels. Importantly however, obtained findings do not suggest the existence of relationship between accurate vowel identification hit rate and mean reaction time.

To conclude, we believe there is substantial evidence to suggest that isolated, synthetic vowel tokens’ identification is a function of a complex use of spectral and temporal cues relative to Algerian learners’ L1, while any probable use of spectral use relative to French language still needs further firm investigation. While Algerian learners’ L1 signals vowel contrasts both spectrally and temporally, our empirically based findings suggest predominant use of temporal cues to identify accurately vowels in the absence of sufficient spectral cues. Nonetheless, sufficient spectral cues for inexistent vowels in L1 may not entirely ensure accurate vowel identification. Furthermore, speaking proficiency does neither lend support to the belief that accurate L2 perception precedes production nor to the belief that L2 perception and production are significantly correlated. Sensitivity to both spectral and temporal cues in Algerian learners’ L1 may not ensure them accurate identification of English vowels. In addition, Algerian learners’ experience with English may not enhance their perceptual abilities.

9.3 Discussion of in-Context Vowel Identification Results

Results of in-context vowel identification revealed a largely different pattern with vowels identified at high accuracy rates. Aggregate mean identification of vowels and pooled sensitivity index *d’* across several vocabulary series demonstrated Algerian learners’ significant perceptual abilities in identifying RP English vowels in varying phonetic contexts. Findings showed a little variation with regard to accuracy pattern among vowels compared to that of isolated vowel identification, as long vowels did not show the same advantage. The identification pattern for pooled data in all experimental conditions was /ɜ:/ > /e/ > /ɔ:/ > /ɪ/ > /ʌ/ > /i:/ > /ɑ:/ > /ɒ/ > /æ/. The vowels /ʊ/ and /u:/ are not listed for they were included in the experiment in limited contexts only. This is not to claim that their occurrence is very limited, but that material stimuli in the third experiment were orientated by the findings of the same different vowel discrimination experiment.

Although identification pattern of vowels in context reveals a new turn, it should rather be interpreted with caution, as word tokens within word series were not truly representative of all vowels. Some of the vowels were experimented on several times compared to others that were included in few experimental manipulations, raising the issue of representativity and less the issue of data reliability. In spite of that, the vowels /æ/, /ɒ/ and /ɑː/ were the least accurately identified ones, respectively, consistently suggesting probable difficulty in identifying vowels spectrally located in the mid-low central and back positions on the perceptual map.

Regarding the word token /b.n/ in *Word Series 1*, for instance, only the tense vowels /ɔː/ and /ɑː/ were accurately identified above .70 hit rate, while other vowels /æ/, /ʌ/ and /ɒ/ were accurately identified below .62 hit rate. The vowel /ʌ/ was the least accurately identified and misidentified with /æ/ in 50% of the responses, suggesting the high level of confusion Algerian learners had in distinguishing between /ʌ/ and /æ/. A similar pattern of findings was observed for the word token /m.ʧ/ in *Word Series 6*, where /ʌ/ was accurately identified below .60 hit rate and misidentified with /æ/ in 41% of the responses. Further findings from other word series involving mid low and low, central vowels lend further support for our claim of the considerable confusion in distinguishing between /ʌ/ and /æ/, except in the word token /s.kə/. In the latter word token, the vowels were unsurprisingly accurately identified above .90 hit rate. We believe that the few number of occurrences of other vowels in this word token caused Algerian learners to attend more to differences between the two vowels and to reduce the number of alternative response choices. That is, occurrence of more than two vowels in the same word token may cognitively overload Algerian learners' decision. Carrying on with mid low back vowels, it stands clearly from the findings that the vowel /ɒ/ was accurately identified at lower hit rates as against its long or rather tense counterpart /ɔː/, while it was misidentified with /ʌ/ and /æ/ in a considerable percentage of responses. As for high front vowels /ɪ/ and /iː/, it stand clear they were identified accurately above a .85 hit rate in all experimental conditions, demonstrating a perceptual easiness in distinguishing between this vowel contrast. This perceptual advantage is not significantly exceptional as Algerian learners L1 uses temporal cues to signal lexical contrasts, as in the words /ʔalif/ (the first letter of the Arabic alphabet) and /ʔalijf/ (meaning ‘pet’). High accuracy of distinction between the vowels /ɪ/ and /iː/ corresponds almost perfectly to discrimination accuracy as previously seen.

As displayed in confusion matrices, interaction between consonants and vowels seems to have caused improvement in learners’ accurate identification of vowels, indicating higher perceptual sensitivity to the occurrence of vowels in a phonetic context. This suggests that consonant and vowel interaction may result in an effective use of information inherent in temporal and spectral properties that learners tend to employ optimally in identification. As demonstrated by our empirical findings, this information dynamics had a significant effect on identification accuracy and reaction time as well. As against isolated vowel tokens’ identifications, Algerian learners of English identified vowels in context with a significantly higher accurate of .49 in both 1- and 2-syllable words. As for reaction time, vowels in context were identified in a shorter reaction time, with a time difference of .71 sec. in 1-syllable words and .64 sec. in 2-syllable words. Once more in this experiment, there were non-significant effects of proficiency in vowel in-context identification, except for the vowel /ɒ/ that was identified in 2-syllable words at a significantly higher accuracy rate among high-proficiency as against low-proficiency groups ($MD = .09, p < .05$).

When all experimental manipulations were combined, statistical analysis of aggregate data showed some significant effects on vowel identification and reaction time. First, it was found that vowel length affected significantly the vowel’s identification accuracy rate but not reaction time, following the pattern /iː/ > /ɔː/ > /ɑː/ > /ɜː/ > /ɪ/ > /e/ > /ʌ/ > /æ/ > /ɒ/, with /iː/

being identified significantly at a higher accuracy rate and /æ/ and /ɒ/ being identified at significantly lower accuracy rates than other vowels. /Second, it was found phonetic context affected significantly both vowel's identification accuracy rate and reaction time, with vowels in context significantly identified at higher accuracy rates and in shorter reaction times. Third, vowel context interaction was found to affect significantly both vowel's identification accuracy rate and reaction time.

Theoretical claims and research findings lend support to our empirical findings. Speech perception research suggests that acoustic (i.e. spectral and temporal) correlates of a vowel vary as a function of the preceding or following consonants. Stated differently, spectral and temporal characteristics of a vowel are likely to vary as a function of coarticulation effects resulting from joint influences of the vowel's neighbouring sounds, essentially consonants (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Rakerd, 1984). Presumably, little or much of perceivers' adjustment to any vowel's spectral and temporal variation and identification accuracy may be a function of variation importance. If variation is negligible, then it would simply be neglected and vowels' identification accuracy would be high. If not, then spectral and temporal variation would be closely attended to, and identification accuracy would consequently vary as a function of the perceiver's sensitivity. Accordingly, identification of vowels in the two previous experiments (isolated vowel tokens and in-context vowels) substantially provides more evidentiary weight to the second claim. Both isolated and in-context vowel identification findings give empirical evidence in favour of the importance of acoustic variation resulting from coarticulation effects in determining the amount of attention required for the vowel identification task and accuracy performance.

As coarticulation effect were probably important, it was found that Algerian learners of English experientially took significantly longer time to attend carefully to isolated vowel tokens in the multiple temporal manipulations (understood as reaction time) before they could identify the vowel. Respective identification accuracy rates of isolated vowel tokens lend additional support to the importance of coarticulation effects and spectral and temporal variation. Likewise, it was found that Algerian learners of English experientially took Algerian learners significantly shorter time to attend carefully to in-context vowels before they could identify them at significantly higher accuracy rates. These findings are consistent with Rakerd's (1984) concerning the contextual influence of the presence of neighbouring consonants can have on vowel perception. Carrying out two experiments on scaling of similarity relationships among a set of American English vowels occurring in isolation and within a /dVd/ consonantal context, Rakerd (1984) reported that "there was significantly greater agreement among individuals who heard vowels in consonantal context than there was among those who heard isolated vowels" (p. 128), suggesting the deployment of different perceptual strategies in processing stimuli in both conditions.

9.4 Discussion of Attitudes towards English Pronunciation Learning Results

Algerian learners' perceptions and beliefs on the importance of English pronunciation learning and instruction demonstrated consistent findings with similar studies in the field. Algerian learners of English approved of the importance of pronunciation within the learning context, showing a highly positive attitude towards sounding like a native and the need for authentic pronunciation in case of social integration within an English-speaking community. These beliefs were combined with a considerable belief in their abilities to mastering English pronunciation, irrespective of what is commonly held about the impossibility of achieving a native-like accent. Algerian learners of English also showed their dissatisfaction with accented and distorted pronunciation, favouring thus the role of unaccented English in their learning context and equally appreciating skilled pronunciation. Achievement of an unaccented or an English native-like accent may be a common attribute among English L2 learners. Coskun (2001) reported that Turkish future teachers of English perceived that the goal of a pronunciation course was to attain a native-like accent. The Algerian learners' profile with regard to their held attitudes towards English pronunciation resembles that reported by Waniek-Klimczak, Rojczyk, and Porzuczek (20015) with Polish students majoring in English. The latter were also reported to be very positive about the role of pronunciation in the learning context and to be self-confident enough to have good command of English pronunciation. Nonetheless, common findings in the present research and that one within the Polish context is that, while learners strived for skilled and less accented pronunciation, they were satisfied with their pronunciation. Special to the Algerian context, however, is learners' reluctance to get involved in self-initiated pronunciation learning. Unfortunately, the small number of male participants in our sample made it non-significant to conduct further analysis on attitudinal variation across gender. Learners' expressed reluctance to more autonomous learning of English pronunciation may be due to the absence of clearly stated objectives with regard to teaching both *English Phonetics and Phonology* and *Speaking and Phonetics* courses in the university. Interest in theoretical teachings of articulatory phonetics and phonology and the type of assessment of acquired concepts may be the cause of such unwillingness to be actively involved in self-directed pronunciation learning. Transcription activities, absence of clearly designed pronunciation learning tasks and inadequate measurement of pronunciation skills in oral expression course may cause learners to trivialise this aspect of the language, irrespective of their perception of its importance. Thus, they would have a very instrumental motivation obeying to an exam-orientated pedagogy, which may jeopardise the real objectives of learning that is to develop skills required for life.

As for relationships among attitudinal factors and speaking proficiency, we have to emphasize that only two out of seven attitudinal factors were significantly correlated with speaking proficiency, understood as level of achievement in *Speaking and Phonetics* course, plus other significant interrelationships among attitudinal factors. Positive significant correlations between speaking proficiency and attitude towards achievement of a native-like accent (r = .41, 2-tailed *p* < .01) and attitude towards active learning of pronunciation (r = .34, 2-tailed *p* < .05) suggest the significant directive and mediating power of Algerian learners' perceptions of achieving a native-like accent and active learning of pronunciation in their learning efforts. This claim is further confirmed by the significant positive correlation between the above cited attitudinal factors (r = .34, 2-tailed *p* < .05), suggesting that the more positive attitude Algerian learners hold towards achievement of a native-like accent, the more positive attitude they hold towards active learning of pronunciation. Further reported attitudes towards aspects of pronunciation and instruction tend to be uniform and signal less ambivalence in Algerian learners' attitudes, except for attitude towards importance of perceptual abilities in pronunciation mastery. As Algerian learners' attitudes towards importance of pronunciation (mainly in an instruction setting) and importance of pronunciation in social integration were positively correlated, their attitudes towards importance of perceptual abilities in pronunciation mastery correlated negatively, though non-significantly, with both attitudes towards active learning of pronunciation and importance of pronunciation in social integration.

Regarding variation in attitude towards investigated aspects of pronunciation learning and instruction, it was found that the same attitudinal factors, correlating positively significantly with speaking proficiency, differed significantly among proficiency groups. High proficiency group held significantly a more positive attitude towards achievement of a native-like accent (*M* = 4.42) as against poor proficiency group (*M* = 3.65). Similarly, high proficiency group held significantly a more positive attitude towards active learning of pronunciation (*M* = 3.38) as against poor proficiency group (*M* = 2.76). However, there were non-significant differences in attitude towards achievement of a native-like accent and active learning of pronunciation among mediocre and high proficiency groups. Once more, variation findings among proficiency groups imply the mediating directive power of the two significant attitudinal factors in eliciting systematic learning behaviours among high proficiency group capable of affecting their speaking proficiency compared to other groups.

Findings of multiple regression analysis among attitudinal factors (taken as independent variables) and speaking proficiency (taken as a dependent variable) showed a highly significant model, with attitude towards active learning of pronunciation as the only significant predictor of speaking proficiency. This indicates that more positive attitude towards active learning of pronunciation predicts improvement in speaking proficiency. This finding is important as it lends empirical support to role of perceptions of the self as a responsible agent for one's learning and, therefore, for learning outcomes and achievements.

9.5 Recommendations for Further Research

Both exploratory and confirmatory nature of our present research makes our recommendations determined by theoretical rather than pedagogical interest, as it is quite early at that stage to suggest any pedagogical recommendations, as any of the latter should build primarily on prerequisite empirical data to conceive of empirically grounded educational practices and comprehensively informed L2 education.

One of the recommendations to suggest at that stage is to urge the need to investigate Algerian learners' perceptual abilities of French vowels to make sure they are firmly formed, represented and distinguished from Algerian learners' L1. The findings of such an investigation would form an empirical basis for a reliable interpretation of results of potential research on Algerian learners' perception of English sounds, as these findings would form empirically based reference points for any comparison. Further investigations are manifestly required to draw an empirically based perceptual map of vowel representations typical of the Algerian learner, specifying spectral and temporal characteristics of vowels and accurately delimiting their boundaries. This involves a series of investigations. First, highly reliable research on Algerian learners' perceptual abilities must pass by a serious investigation of the spectral and temporal properties of Algerian learners' L1 vowels that are unfortunately neglected or, in the best case, taken for granted. This is to be followed by a preliminary work on the vowel inventory of the Algerian dialect that would essentially provide guidelines to undertake longitudinal and developmental research on the change of Algerian learners' perceptual and production abilities in both French and English.

Given the special status of English in the Algerian educational system, we believe there is a need for research to ensure whether English vowels' perception is mediated by the filter of the Arabic vowel system, French vowel system, an intermediate system between Arabic and French or completely a new language system, similar or equivalent to that of English. We highly recommend the use of synthetic vowel continua in isolation or in consonantal context (without temporal manipulation) to delimit vowel categories as formed and represented by the Algerian learner and figure out probable overlapping or crowded vowel space. As perceivers tend to perceive vowels less discretely than consonants, we suggest the use of vowel similarity or difference scaling to have a better understanding of sensitivity change and avoid, therefore, categorical data.

As for the presumed role of L2 experience and instruction in improving perceptual and production abilities among learners, it is necessary to investigate the spectral and temporal characteristics of vowels from the three languages as produced by different types of learners selected along some criteria to meet research interests. Findings of such a research would reveal much about learners' production abilities and provide empirical data to assess the spectral and temporal accuracy with which vowels from Arabic, French and English are actually produced. Research interests may equally involve linguistic and psychological criteria, such as age of exposure to foreign language, amount of L2 input, L1 frequency use, L2 frequency use in social interaction, etc. This would allow us compare probable perceptual and production differences and investigate their relationship and variation as a function of the previously cited criteria.

More research on Algerian learners’ attitudes towards English pronunciation learning and instruction is needed. It is necessary to have a deeper understanding of the three constructs making Algerian learners’ overall attitude. One suggestion for further research is to compare learners’ attitudes towards English pronunciation among high school and university students from other majors. Research on the beliefs of high school and university students from other majors could reveal much about the effect of educational practices on learners’ attitudes and the way they are initially shaped. Potential findings may allow us gain further insights into whether learners’ attitudes towards English pronunciation are amenable to change through formal instruction.

General conclusion

This section brings our research to the field by reporting the main conclusion about our research questions investigating the perceptual abilities of Algerian learners of English with respect to RP English vowels and exploring their attitudes towards RP English pronunciation learning and instruction within the university setting. Research on English phonological acquisition or learning in the Algerian university context is to date more limited than research on the other aspects of the language, and the undertaken research in the field is even more limited to production than to perception abilities. Perceptual abilities of Algerian learners of English are either taken for granted to be intact, and, if not, are left simply uninvestigated or surrounded by theoretical assumptions and speculations without immediate intention to investigate the issue. The present research aimed to contribute to the field of English speech perception among Algerian learners of English by examining their perception of RP English prototypic monophthongs that differ quantitatively and qualitatively from any repertoire of vowels learners might have formed in their L1 (Arabic) and French, being the first foreign language to which they have been introduced. Limiting choice to investigating learners' perceptual abilities in discriminating and identifying RP English pure vowels alone was driven by empirical findings demonstrating the less categorical nature of vowel perception as against consonants and the richness of Algerian learners' dialect in consonantal sounds. The second part of the research concerned itself with exploring learners' attitudes towards English pronunciation learning and instruction within the formal educational setting in Algeria. In light of the above, the present research draws the following conclusions arranged in the order of the research questions that motivated the research.

Algerian learners of English demonstrated important perceptual abilities in discriminating among RP English prototypic pure vowels. Algerian learners showed high perceptual sensitivity to discriminate among several vowel contrasts, except for the contrast

(/ɑː/ vs. /ʌ/) to which they demonstrated no perceptual sensitivity. Overall findings suggest a perceptual advantage Algerian learners have at discriminating differences between some vowels on the basis of spectral differences alone. This fact implies that Algerian learners still maintain the so-called perceptual flexibility or sensitivity with which they either preserve or recover the ability to discriminate L2 sound contrasts. This perceptual flexibility is operational across various areas on the vowel perceptual map, except for the above-mentioned vowel pair. Algerians learners' discrimination abilities did not correlate with their speaking proficiency and did not differ as a function of the latter.

Findings of isolated vowel tokens identification test demonstrated a specific picture for this perceptual flexibility or sensitivity in that it revealed the way Algerian learners showed effect of both temporal and spectral information in identification perceptual task, with a consistently maintained identification accuracy and substantially higher sensitivity when using temporal information. This indicates quite a developed or still developing perceptual system that can integrate features from both the L1 and the L2 in discriminating and identifying vowels in a further language. Nonetheless, isolated vowel tokens' identification did not correlate significantly with speaking proficiency and did not differ as a function of the latter as well. The Speech Learning Model and the Perceptual Assimilation Model give support to present findings.

In-context vowel identification findings showed significant improvement in learners' identification accuracy and perceptual sensitivity as a function of vowel occurrence within a consonantal context. A consistently maintained identification accuracy and significantly higher sensitivity were observed for almost vowels, with a significant reduction in reaction time as well. These findings suggest that accuracy in vowel identification is the result of a complex use of spectral and temporal information, on the one hand, and a special processing of information, on the other. Indeed, the first ones concerns formant dynamics, as an inherent task of the auditory system, and the second one concerns linguistic processing, as specialised for processing of human language sounds as they form an integral part of an abstract system. As against previous findings, consonantal context proves to improve sensitivity accuracy and increase identification accuracy. Additionally, identification of vowels in context did not differ significantly as a function of speaking proficiency. It is to emphasise herein that, reported perceptual flexibility demonstrated by Algerian learners of English in our research work should be cautiously interpreted, given the important linguistic input and theoretical knowledge they received on the matter.

Findings about the Algerian learners' attitude profile towards English pronunciation learning and instruction suggests their overall positive profile towards learning this aspect of the language, by reporting a high sense of self-confidence and perceived capacity to learn English pronunciation and achieve a native-like accent. The concept of the native speaker seems to be present in them when reflecting on learning English pronunciation and the equal importance of native-like pronunciation for both learning and teaching the language. Nonetheless, learners' perceptions and beliefs about the importance of English pronunciation did not tend to match their explicitly expressed willingness to involve in self-initiated pronunciation learning activities. As for relationships among investigated attitudinal factors and speaking proficiency, only attitude towards achievement of native-like accent and active pronunciation learn did significantly positively correlate with speaking proficiency. However, only attitude towards learning active pronunciation learning proved a significant predictor of speaking proficiency, as proficiency increased with a stronger positive attitude.

References

Ackerman, B. P. (1990). Attention and memory in context-independent and context-interactive situations. In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 289-304). North-Holland: Elsevier Science Publishers B. V.

Al-Ani, S. (1970). *Arabic phonology*. Paris: Mouton.

Albarracín, D., Johnson, B. T., Zanna, M. P., & Kumkale, G. T. (2005). Attitudes: Introduction and scope. In D. Albarracín, B. T. Johnson & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 3-19). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Amaro, J. C., Flynn, S., & Rothman, J. (2012). Third language (L3) acquisition in adulthood. In J. C. Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 1-6), Amsterdam: John Benjamins Publishing Company.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC. Author.

Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*. 89(4), 369-406.

Anderson, J. R. (1983). *The architecture of cognition*. Harvard University Press.

Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology*, summer 1992, 105(2), 165-180.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, 51(4), 355-365.

Anderson, J. R. (2015). *Cognitive psychology and its implications* (8th ed.). Worth Publishers.

Anderson, J. R., & Bower, G. H. (1974). A propositional theory of recognition memory. *Memory & Cognition*, 2(3), 406-412.

Archibald, J. (1993). Language learnability and L2 phonology: The acquisition of metrical parameters. Springer Science+Business Media Dordrecht.

Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, 114(3), 632-656.

Aspray, W. (1990). *John Von Neumann and the origins of modern computing*. Massachusetts: The MIT Press.

Baddeley, A. D. (1999). *Essentials of human memory*. Psychology Press.

Balas, A. (2018). English vowel perception by Polish advanced learners of English. *Canadian Journal of Linguistics*, 63(3), 309-338.

Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning and Verbal Behaviour*, 22, 88-104.

Baran-Łucarz, M. (2017). FL pronunciation anxiety and motivation: Results of a mixed-method study. In E. Pierchurska-Kuciel, E. Szymańska-Czaplak, & M. Szyszka (Eds.), *At the crossroads: Challenges of foreign language learning* (pp. 107-133). Switzerland: Springer.

Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. *Journal of Personality and Social Psychology*, 43(3), 425-436.

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3-51). The Guilford Press.

Bargh, J. A., & Cohen, J. L. (1978). Mediating factors in the arousal-performance relationship. *Motivation and Emotion*, 2(3), 243-257.

Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends in Cognitive Sciences*, 16(12), 593-605.

Barwise, J. (1977). An introduction to first-order logic. In J. Barwise (Ed.), *Handbook of mathematical logic* (pp. 5-46). North-Holland: Elsevier Science Publishers B. V.

Beghoul, Y. (2007). *The phonological interlanguage of the undergraduate students at the Department of English, University of Mentouri, Constantine*. Unpublished doctoral dissertation. Mentouri University of Constantine, Algeria.

Belmechri, F., & Hummel, K. (1998). Orientations and motivation in the acquisition of English as a second language among high school students in Quebec City. *Language Learning*, 48(2), 219-244.

Bernstein, L. E. (2005). Phonetic processing by the speech perceiving brain. In D. B. Pison & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 79-98). Blackwell Publishing.

Best, C. T. (1991). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *Haskins Laboratories Status Report on Speech Research, SR-107/108*, 1-30.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204) Timonium, Maryland: York Press.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Haskins Laboratories Status Report on Speech Research, SR-109/110*, 89-108.

Bever, T. G., Lackner, J. R., & Kirk, R. (1969). The underlying structures of sentences are the primary units of immediate speech. *Perception & Psychophysics*, 5(4), 225-234.

Bialystok, E. (1991). Metalinguistic dimensions of bilingual language proficiency. In E. Bialystok (Ed.), *Language processing in bilingual children* (p. 113-140). Cambridge: Cambridge University Press.

Bickford, A. C., & Floyd, R. (2006). *Articulatory phonetics* (6th ed.). Dallas, Texas: SIL International.

Birdsong, D. (1999). Introduction: Whys and why nots of the critical period hypothesis for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 1-22). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Boersma, P. (2013). The use of Praat in corpus research. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford handbook of corpus phonology* (pp. 343-360). Oxford, England: Oxford University Press.

Boersma, P., & Hamann, S. (2009). Introduction: Models of phonology in perception. In P. Boersma & S. Hamann (Eds.), *Phonology in perception* (pp. 1-24). Berlin: Mouton de Grutyer.

Boersma, P., & Weenik, D. (2016*). PRAAT: Doing phonetics by computer* (Version 6.0.14) [Software]. The University of Amsterdam: The Netherlands. Available from http://www.fon.hum.uva.nl/praat/.

Bond, Z. S., & Garnes, S. (1980). Misperceptions of fluent speech. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 115-132). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133-160). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Bornstein, M. H. (1990). Attention in infancy and the prediction of cognitive capacities in childhood. In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 3-20). North-Holland: Elsevier Science Publishers B. V.

Borrell, A. (1990). Perception et (re)production dans l'apprentissage des langues étrangères. Quelques réflexions sur les aspects phonético-phonologiques. *Revue de Phonétique Appliquée*, 95-97, 107-114.

Breckler, S. J. (2004). Hold still while I measure your attitude. In G. Haddock & G. R. Maio (Eds.), *Contemporary perspectives on the psychology of attitudes* (pp. 77-91). Hove: Psychology Press.

Brière, E. (1966). An investigation of phonological interference. *Language*, 42(4), 768-796.

Broadbent, D. E. (1958). *Perception and communication*. Pergamon Press.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language. The acquisition of a hierarchy of habits. *The Psychological Review*, 6(4), 345-375.

Burke, R. S. (1990). A cognitive-developmental approach to studying attention deficits. In In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 365-382). North-Holland: Elsevier Science Publishers B. V.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast in French-English bilinguals. *The Journal of the Acoustical Society of America*, 54(2), 421-428.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34, 372-387.

Cenoz, J., & Lecumberri, L. G. (1999). The acquisition of English pronunciation: Learners' views. *International Journal of Applied Linguistics,* 9(1), 3-17.

Chapman, M. (1990). Cognitive development and the growth of capacity: Issues in neo-Piagetian theory. In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 263-288). North-Holland: Elsevier Science Publishers B. V.

Chomsky, N. (1957). Review of verbal behavior. *Language*, 35(1), 26-58.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Chomsky, N. (2002). *Syntactic structures*. Berlin: Mouton de Gruyter.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Chow, S. L. (1986). Automatic detection, consistent mapping, and training. *Bulletin of the Psychonomic Society*, 24(6), 431-434.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. Harcourt Brace Jovanovich.

Cole, J., & Hualde, J. I. (2011). Underlying representations. In M, van Oostendrop, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology*, Vol. 1 (pp. 241-266). Blackwell Publishing.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133-163). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review*, 81(4), 348-374.

Cole, R. A., Rudnicky, A. I., Zue, V. W., & Reddy, D. R. (1980). Speech as patterns on paper. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 3-50). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Conrad, C. (1974). Context effects in sentence comprehension: A study of the subjective lexicon. *Memory & Cognition*, 2(1A), 130-138.

Copeland, J. (2004). Artificial intelligence. In J. Copeland (Ed.), *The essential Turing: Seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life: Plus the secrets of Enigma* (pp. 353-361). Oxford University Press.

Corder, S. P. (1975). Error analysis, interlanguage and second language acquisition. *Language Teaching*, 8(4), 201-218.

Corder, S. P. (1993). A role for the mother tongue. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language learning* (Rev. ed.) (pp. 18-31). Amsterdam: John Benjamins Publishing Company.

Coskun, A. (2011). Future English teachers' attitudes towards EIL pronunciation. *Journal of English as an International Language*, 6(2), 46-68.

Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th ed.). Blackwell Publishing.

Cutler, A., & Norris, D. (1988). Role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology*, 14(1), 113-121.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). Syllable differing role in the segmentation of French and English. *Journal of Memory and Language*, 25, 385-400.

Cutting, J. E. (1978). There may be nothing peculiar to perceiving in a speech mode. In J. Requin (Ed.), Attention and performance VII (pp. 229-243). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cutting, J. E., & Rosner, B. S. Categories and boundaries in speech and music*. (1974). *Perception & Psychophysics*, 16(3), 564-570.

Cutting, J. E., Rosner, B. S., & Foard, C. F. (1976). Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology*, 28, 361-378.

D'Aloisio, A., & Klein, R. M. (1990). Aging and the deployment of visual attention. In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 447-466). North-Holland: Elsevier Science Publishers B. V.

DeKeyser, R. M. (2007). Skill acquisition theory. In B. Van Patten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97-113). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *The Journal of the Acoustical Society of America*, 27(4), 769-773.

Deterding, D. (2006). The north wind versus a wolf: Short texts for the description and measurement of English pronunciation. *Journal of International Phonetic Association*, 36(2), 187-196.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70(1), 80-90.

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, 1(2), 121-144.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55(1), 149-179.

Dörnei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Dresher, B. E. (2011). The phoneme. In M, van Oostendrop, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology*, Vol. 1 (pp. 26). Blackwell Publishing.

Eagly, A. H., & Chaike, S. (1993). *The psychology of attitudes*. Orlando, Florida: Hardcourt Brace Jovanovich.

Ekman, P. (1994). All emotions are basic. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion* (pp. 15-19). Oxford: Oxford University Press.

Ellis, R. (1999). *The study of second language acquisition*. Shanghai Foreign Language Education Press.

Elvin, J., Escudero, P., & Vasiliev, P. (2014). Spanish is better than English for discriminating Portuguese vowels: acoustic similarity versus vowel inventory size. *Frontiers in Psychology*, 5, 1-8.

Eysenck, M. W. (1982). *Attention and arousal: Cognition and performance*. Berlin, Heidelberg: Springer-Verlag.

Eysenck, M. W., & Keane, M. T. (2010). *Cognitive psychology: A student's handbook* (6th ed.). Psychology Press.

Fennel, C. T. (2012). Habituation procedures. In E. Hoff (Ed.), *Research methods in child language: A practical guide* (pp. 3-16). Blackwell Publishing.

Ferreira, F., & Anes, M. (1994). Why study spoken language? In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 33-56). London, England: Academic Press.

Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention, and behaviour: An introduction to theory and research. Reading, Massachusetts: Addison-Wesley.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Belmont, California: Cole Publishing Company.

Flege, J. E. (1981). The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly*, 15(4), 443-455.

Flege, J. E. (1987a). Effects of equivalence classification on the production of foreign language speech sounds. In A. R. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 9-39). Dordrecht: Foris Publications.

Flege, J. E. (1987b). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47-65.

Flege, J. E. (1987c). A critical period for learning to pronounce foreign languages? *Applied Linguistics*, 8, 162-177.

Flege, J. E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Heubner & C. A. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theory* (pp. 249-289). Philadelphia: John Benjamins Publishing Company.

Flege, J. E. (1992). Speech learning in a second language. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 565-604). Timonium, Maryland: York Press.

Flege, J. E. (1995). Second language speech learning: Learning, theory, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, Maryland: York Press.

Flege, J. E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 101-131). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Flege, J. E. (2003a). A method for assessing the perception of vowels in a second language. In E. Fava & A. Mioni (Eds.), Issues in clinical linguistics. Padova, Italy: Unipress.

Flege, J. E. (2003b). Assessing constraints on second-language segmental production and perception. In A. Eyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319-355). Berlin: Mouton de Gruyter.

Flege, J. E. (2005, April). Origins and development of the Speech Learning Model. Lecture presented at the 1st ASA Workshop on L2 speech Learning, Simon Fraser University, Vancouver, BC, Canada, 14-45 April. Retrieved from http://www.jimflege.com/ (date last viewed June 02, 2018).

Flege, J. E. (2008). Child-adult differences in second language phonological learning: The role of cross-language similarity. *Language and Speech*, 51(4), 317-342.

Flamming, E., & Johnson, S. (2007). Rosa's roses: reduced vowels in American English. *Journal of the International Phonetic Association*, 37(1), 83-96.

Fodor, J. A. (1983). *The modularity of mind*. Massachusetts: The MIT Press.

Foss, D. J., Harwood, D. A., & Blank, M. A. (1980). Deciphering decoding decisions: Data and devices. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 165-199). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Foster, M., & Sherrington, C. S. (1897). *A textbook of physiology. Part III: The central nervous system* (7th ed.). MacMillan and Co.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3-28.

Fowler, C. A. (1989). Real objects of speech perception: A commentary on Diehl and Kluender. *Ecological Psychology*, 1(2), 145-160.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of Acoustical Society of America*, 99(3), 1730-1741.

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36(2, 3), 171-195.

Frazier, L. (1995). Issues of representation in psycholinguistics. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 1-27). New York: Academic Press.

Fry, D. B. (1979). *The physics of speech*. Cambridge, England: Cambridge University Press.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language & Speech*, 5, 171-189.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361-377.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London, England: Edward Arnold.

Gardner, R. C., & Lambert, w. E. (1972). *Attitudes and motivation in second language learning*. Rowley, Massachusetts: Newbury House Publishers.

Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development*. New York: Oxford University Press.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, 49, 585-612.

Goodman, J. C., Lee, L., & DeGroot, J. (1994). Developing theories of speech perception: Constraints from developmental data. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 3-33). Cambridge, Massachusetts: The MIT Press.

Gottfried, T. L., & Beddor, P. S. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech, 31(1)*, 57-75.

Graf, P., Tuokko, H., & Gallie, K. (1990). Attention deficits in Alzheimer's disease and related dementias. In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 527-544). North-Holland: Elsevier Science Publishers B. V.

Gray, J. A., & Wedderburn, A. A. I. (1960) Shorter articles and notes: Grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, 12(3), 180-184.

Guttentag, R. E., & Ornstein, P. A. (1990). Attentional capacity and children's memory strategy use. In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 305-320). North-Holland: Elsevier Science Publishers B. V.

International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

Hebb, D. (1949). *The organization of behavior: A neurophysiological theory*. New York: John Wiley & Sons.

Herschensohn, J. (2007). *Language development and age*. Cambridge University Press.

Hickock, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393-402.

Hirst, W., Neisser, U., & Spelke, E. (1978). Divided attention. *Human Nature*, June, 54-61.

Hochberg, I. (1975). Most comfortable listening for the loudness and intelligibility of speech. *Audiology*, 14, 27-33.

IBM Corporation. Released 2017. *IBM SPSS Statistics for Windows* (Version 25.0). Armonk, New York: IBM Corporation.

Jaccard, J., & Blanton, H. (2005). The origins and structure of behavior: Conceptualising behavior in attitude research. In D. Albarracín, B. T. Johnson & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 125-172). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110 (3), 306-340.

Jakobson, R. (1978). *Six lectures on sound and meaning* (J. Mepham, Trans.). Cambridge, Massachusetts: The MIT Press.

Jakobson, R., & Halle, M. (1956). *Fundamentals of language*. The Hague: Mouton & Co.

James, W. (1890). *The Principles of psychology*. Chicago: William Benton. Encyclopaedia Britannica.

James, W. (1910). *Psychology*. New York: Henry Holt and Company.

Jastrow, J. (1906). *The subconscious*. Boston: Houghton, Mifflin and Company.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60-99.

Johnson, K. (2003). Acoustic and auditory phonetics. Blackwell Publishing

Johnson, K., & Johnson, H (Eds.). (1998). *Encyclopaedic dictionary of applied linguistics*. Blackwell Publishing.

Johnson, K., & Ralston, J. V. (1994). Automaticity in speech perception: Some speech/nonspeech comparisons. *Phonetica*, 51, 195-209.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Review of Psychology*, 37, 43-75.

Jones, D. (1967). *The phoneme: Its nature and use* (1st ed.). Cambridge, England: Cambridge University Press.

Jusczyk, P. W., Rosner, B. S., Cutting, J. E., Foard, C. F., & Smith, L. B. (1977). Categorical perception of nonspeech sounds by 2-month-old infants. *Perception & Psychophysics*, 21(1), 50-54.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, New Jersey: Prentice-Hall.

Kelley, H. H. (1989). Foreword. In A. W. Kruglanski, *Lay epistemics and human knowledge: cognitive and motivational bases* (pp. vii-ix). New York: Springer Science+Buisness Media.

Kluender, K. R., & Kiefte, M. (2006). Speech perception within a biologically realistic information-theoretic framework. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.) (pp. 153-199). San Francisco: Academic Press.

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford, England: Pergamon Press.

Krashen, S. D. (1989). *Language acquisition and language education: Extensions and applications*. London, England: Prentice Hall International.

Kuhl, P. K. (1984). The perception of speech in early infancy. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice*, Vol. 10 (pp. 1-47). Orlando, Florida: Academic Press.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93-107.

Kuhl, P. K. (1993). Developmental speech perception: Implications for models of language impairment. *Annals New York Academy of Sciences*, 14, 248-263.

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, 4, 812-822.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B*, 363, 979-1000.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, Vol. 255, 606-608.

Ladefoged, P. (1996). *Elements of acoustic phonetics* (2nd ed.). Chicago: The University of Chicago Press.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: The University of Michigan Press.

Lane, H. (1962). Psychological parameters of vowel perception. *Psychological Monographs: General and Applied*, 76(44), 1-25.

Leather, J. (1999). Second-language speech research: An introduction. *Language Learning*, 49(Suppl.1), 1-56.

Leather, J. (2003). Phonological acquisition in multilingualism. In M. P. G. María & M. L. G. Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 23-58). Multilingual Matters.

Lenneberg, E. H. (1967). *Biological foundations of language*. John Wiley & Sons.

Liberman, A. M. (1970). The grammars of speech and language[1]. *Cognitive Psychology*, 1(4), 301-323.

Liberman, A. M. (1992). Plausibility, parsimony, and theories of speech. In J. Alegria, D. Holender, J. Junça de Morais., & M. Radeau (Eds.), *Analytic approaches to human cognition* (pp. 25-40). Amsterdam: Elsevier Science Publishers B. V.

Liberman, A. M. (1996). Speech: A special code. Cambridge, MA: The MIT Press.

Liberman, A. M. (1998). When theories of speech meet the real world. *Journal of Psycholinguistic Research*, 27(2), 111-122.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised*. *Cognition*, 21, 1-36.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, 243, 489-494.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P, & Studdert-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431-461.

Liberman, A. M., Harris, K. S., Hoffman, H. S., &Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boudaries[1]. *Journal of Experimental Psychology*, 54(5), 358-368.

Lieberman, P. (1970). Toward a unified phonetic theory*. *Linguistic Inquiry*, 1(3), 307-322.

Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge: Cambridge University Press.

Lindblom, B. E. F. (1963). Spectrographic analysis of vowel reduction. *The journal of the Acoustical Society of America*, 35(11), 1773-1781.

Lindblom, B. E. F., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *The journal of the Acoustical Society of America*, 42(4), 830-843.

Lisker, L., Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384-422.

Lisker, L., Abramson, A. (1970). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague*, *1967*. Prague: Academia.

Logan, G. D. (1988). Toward an instance theory of automatisation. *Psychological Review*, 95(4), 492-527.

Logan, G. D. (1990). Repetition priming and automaticity: common underlying mechanisms? *Cognitive Psychology*, 22, 1-35.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 883-914.

Logan, G. D., Taylor, S. E., & Etherton, J. L. (1999). Attention and automaticity: Toward a theoretical integration. *Psychological Research*, 62(2-3), 165-181.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, 12, 251-285.

Lovelace, E. A. (1990). Cognitive aging: A summary overview. In E. A. Lovelace (Ed.), *Aging and cognition: Mental processes, awareness and interventions* (pp. 407-434). North-Holland: Elsevier Science Publishers B. V.

Lovie, A. D. (1983). Attention and behaviourism—fact and fiction. *British Journal of Psychology*, 74(3), 301-310.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, 15, 215-233.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd edi.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Macmillan, N. A., Braida, L. D., & Godelberg, R. F. (1987). Central and peripheral processes in the perception of speech and nonspeech sounds. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 28-45). Dordrecht: Martinus Nijhoff Publishers.

Major, R. C. (2001). Foreign accent: *The ontogeny and phylogeny of second language phonology*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Malim, T. (1994). *Cognitive processes: Attention, perception, memory, thinking and language* (1st ed.). The Macmillan Press.

Mallick, D. B., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: Contributions of participants, stimuli, and response type. *Psychological Bulletin Review*, 22(5), 1299-1307.

Marchal, A. (2009). *From speech physiology to linguistic phonetics*. John Wiley & Sons.

Marean, G. C., Werner, L. A., & Kuhl, P. K. (1992). Vowel categorization by very young infants. *Developmental Psychology*, 28(3), 396-405.

Marler, P. (1970). A comparative approach to vocal learning: Song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology Monograph*, 71(2), 1-25.

Masgoret, A., -M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning*, 53, 167-210.

Massaro, D. W. (1975). Language and information processing. In D. W. Massaro (Ed.), *An information-processing analysis of speech perception, reading, and psycholinguistics* (pp. 3-28). New York: Academic Press.

Massaro, D. W. (1987). Psychophysics versus specialised processes in speech perception: An alternative perspective. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 46-65). Dordrecht: Martinus Nijhoff Publishers.

Mayberry, R. I. (1994). The importance of childhood to language acquisition: Evidence from American Sign Language. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 57-90). Cambridge, Massachusetts: The MIT Press.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11(1), 122-134.

McClelland, J. D., & Elman, J. L. (1986). Interactive processes in speech perception: The TRACE Model. In D. E. Rumelhart, J. L. McClelland & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (pp. 58-121). Massachusetts: The MIT Press.

McClelland, J. D., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart, J. L. McClelland & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 3-44). Massachusetts: The MIT Press.

McGuire, G. (2010). Methods in speech perception: A brief primer on experimental designs for speech perception research [PDF document]. Retrieved from https://people.ucsc.edu/~gmcguir1/experiment_designs.pdf (date last viewed June 02, 2018).

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, Vol. 264, December 23/30, 746-748.

Messaoud-Galusi, S., Hazan, V., & Rosen, S. (2011). Investigating speech perception in children with dyslexia: Is there evidence of consistent deficit in individuals. *Journal of Speech, Language, and Hearing Research*, 54, 1682-1701.

Miller, J. L., & Eimas, P. D. (1994). Observations on speech perception, its development, and the search for a mechanism. J. C. Goodman & H. C. Nusbaum (EDS.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 37-55). Cambridge, Massachusetts: The MIT Press.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, 132(2), 297-326.

Mullennix, J. W., Sawusch, J. R., & Garrison, L. F. (1992). Automaticity and the detection of speech. *Memory & Cognition*, 20(1), 40-50.

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, 36(1), 39-66.

Neisser, U. (2014/1967). *Cognitive psychology*. Psychology Press.

Neufeld, G. G. (1984). Phonological asymmetry in second-language learning and performance. *Language Learning*, 38(4), 531-559.

Newell, A., & Rosenbloom, R. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, 65(3), 151-161.

Noels, K. A. (2001). Learning Spanish as a second language: Learners’ orientations and perceptions of their teachers’ communication style. *Language Learning*, 51(1), 107-144.

Nyagaard, L. C., & Pisoni, D. B. (1995). Speech perception: New directions in research and theory. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 63-96). New York: Academic Press.

O'Callaghan, C. (2009). Sounds and events[1]. In M. Nudds & C. O'Callaghan (Eds.), *Sounds and perception: New philosophical essays* (pp. 26-49). Oxford University Press.

O'Callaghan, C., & Nudds, M. (2009). Introduction: The philosophy of sounds and auditory perception. In C. O'Callaghan & M. Nudds (Eds.), *Sounds and perception: Philosophical essays* (pp. 1-25). Oxford University Press.

O'Malley, J. M., & Chamot, A. U. (1990). Learning strategies in second language acquisition. Cambridge: Cambridge University Press.

Oskamp, S., & Schultz, P. Z. (2005). *Attitudes and opinions* (3rd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Boston: Heinle & Heinle Publishers.

Pashler, H. E. (1998). *The psychology of attention*. Cambridge, Massachusetts: The MIT Press.

Pasnau, R. (2009). What is sound? *The Philosophical Quarterly*, 49(196), 309-324.

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. (G. V. Anrep, Trans., & Ed.). Oxford University Press: Humphrey Milford.

Pawlak, M., Mystkowska-Wiertelak, A., & Bielak, J. (2015). Exploring advanced learners' beliefs about pronunciation instruction and their relationship with attainment. In E. Waniek-Klimczak & M. Pawlak (Eds.), *Teaching and researching the pronunciation of English* (pp. 3-22). Dordrecht: Springer.

Pawlak, M., Mystkowska-Wiertelak, A., & Bielak, J. (2015). Exploring advanced learners' beliefs about pronunciation instruction and their relationship with attainment. In E.

Waniek-Klimczak & M. Pawlak (Eds.), *Teaching and researching the pronunciation of English* (pp. 3-22). Dordrecht: Springer.

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton, New Jersey: Princeton University Press.

Petzold, C. (2008). *The annotated Turing: A guide tout through Alan Turing's historic paper on computability and the Turing machine*. Wiley Publishing.

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of Phonetics*, 18, 375-394.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253-260.

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition*, 3(1), 7-18.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 61(5), 1352-1361.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of Acoustic Society of America*, 55(2), 328-333.

Poeppel, D., & Hackl, M. (2008). The functional architecture of speech perception. In J. R. Pomerantz (Ed.), *Topics in Integrative Neuroscience: From cells to cognition* (pp. 154-180). Cambridge University Press.

Posner, M. I., & Boies, S. J. (1971). Components of attention. Psychological Review, 78(5), 391-408.

Potter, J. (1998). Discursive social psychology: From attitudes to evaluative practices. In W. Stroebe & M. Hewstone (Eds.), European Review of Social Psychology, 9(1), 233-266.

Pullum, G. K. (2011). On the mathematical foundations of syntactic structures. *Journal of Logic, Language and Information*, 20(3), 277-296.

Rabbit, P., & Vyas, S. (1979). Memory and data-driven control of selective attention in continuous tasks. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 33(2), 71-87.

Rakerd, B. (1984). Vowels in consonantal context are perceived more linguistically than are isolated vowels : Evidence from an individual differences scaling study. *Perception & Psychophysics*, 35(2), 123-136.

Raphael, L. J., Borden, G. J., & Harris, K. S. (2011). *Speech science primer: Physiology, acoustics, and perception of speech* (6th ed.). Philadelphia: Lippincott Williams & Wilkins.

Rato, A., Rauber, A. S., Kluge, D. C., & Santos, G. R. (2015). Designing speech perception tasks with TP. In J. A. Mompean & J. Fouz-Gonzáles (Eds.), *Investigating English Pronunciation: Trends and directions* (pp. 295-313). Hampshire, England: Palgrave Macmilan.

Rauber, A. S., Rato, A., Kluge, D. C., Santos, G. R., & Figueiredo, M. (2016). *TP* (Version 3.1) [Software]. Brazil: WORKEN Enngenharia de Sistemas. Available from http://www.worken.com.br/tp/tp_install.html.

Remez, R. E. (1994). A guide to research on the perception of speech. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 145-172). London, England: Academic Press.

Remez, R. E. (2005). Perceptual organisation of speech. In D. B. Pison & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 28-50). Blackwell Publishing.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice*, Vol. 10 (pp. 243-335). Orlando, Florida: Academic Press.

Repp, B. H. (1987). The role of psychophysics in understanding speech perception. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 66-79). Dordrecht: Martinus Nijhoff Publishers.

Rogers, T. T., & McClelland, J. D. (2004). *Semantic cognition: A parallel distributed processing approach*. Massachusetts: The MIT Press.

Rosen, R. S., DeLouise, M-K., & Daley, K. (2014). Native language, target language, and the teaching and learning of American Sign Language vocabulary. In D. McKee, R. S. Rosen, & R. McKee (Eds.), *Teaching and learning signed languages* (pp. 145-174). London, Engalnd: Palgrave Macmillan.

Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. B. Pison & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 51-78). Blackwell Publishing.

Rubin, P., Turvey, M. T., & van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, 19(5), 394-398.

Ruthruff, E., Allen, P. A., Lien, M-C., & Grabbe, J. (2008). Visual word recognition without central attention: Evidence for greater automaticity with greater reading ability. *Psychological Bulletin & Review*, 15(2), 337-343.

Ryle, G. (2009/1949). *The concept of mind*. Routledge.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, 12(4), 110-114.

Schimmack, U., & Crites, Jr., S. L. (2005). The structure of affect. In D. Albarracín, B. T. Johnson & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 397-435). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Schneider, F. W., & Kintz, B. L. (1967). An analysis of the incidental-intentional learning dichotomy[1]. *Journal of Experimental Psychology*, 73(1), 85-90.

Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing: behavior, theory, and biological mechanisms. *Cognitive Science*, 27, 525-559.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84(1), 1-66.

Schneider, W., Pimm-Smith, M., & Worden, M. (1994). Neurobiology of attention and automaticity. *Current Opinion in Neurobiology*, 4, 177-182.

Schouten, M. E. H. (1987). Speech perception and the role of long-term memory. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 3-27). Dordrecht: Martinus Nijhoff Publishers.

Scott, B. L. (1980). Speech as patterns in time. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 51-71). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Scovel, T. Foreign accents, language acquisition, and cerebral dominance. *Language Learning*, 69(3&4), 245-253.

Scruton, R. (2009). Sounds as secondary objects and pure events. In M. Nudds & C. O'Callaghan (Eds.), *Sounds and perception: New philosophical essays* (pp. 50-68). Oxford University Press.

Searle, C. L., Jacobson, J. Z., & Kimberley, B. P. (1980). Speech as patterns in the 3-space of time and frequency. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 73-102). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3(3), 243-261.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84(2), 127-190.

Shiffrin, R. M., & Schneider, W. (1984). Theoretical note: Automatic and controlled processing revisited. *Psychological Review*, 91(2), 269-276.

Simon, E., & Taverniers, M. (2011). Advanced EFL learners' beliefs about language learning and teaching: A comparison between grammar, pronunciation, and vocabulary. *English Studies*, 92(8), 896-922.

Simon, H. A., & Newell, A. (1970). Human problem solving: The state of the theory in 1970. *American Psychologist*, 26(2), 145-159.

Singleton, D. (2003). Critical period or general age factor(s)? In M. del P, García Mayo, & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 3-22). Clevedon: Multilingual Matters.

Singleton, D. (2003). Critical period or general age factor(s)? In M. del. P. Garcia María & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 3-22). Multilingual Matters.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

Smith, B. C. (2009). Speech sounds and the direct meeting of minds. In M. Nudds & C. O'Callaghan (Eds.), *Sounds and perception: New philosophical essays* (pp. 183-210). Oxford University Press.

Smolensky, P. (1986). Information processing in dynamical systems: foundations of harmony theory. In D. E. Rumelhart, J. L. McClelland & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 194-281). Massachusetts: The MIT Press.

Sorensen, R. (2009). Hearing silence: The perception and introspection of absences. In C. O'Callaghan & M. Nudds (Eds.), *Sounds and perception: Philosophical essays* (pp. 126-145). Oxford University Press.

Stevens, K. N. (2006). Features in speech perception and lexical access. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 125-155). Blackwell Publishing.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., & Öhman, S. E. G. (1969). Crosslanguage study of vowel perception. *Language & Speech*, 12, 1-23.

Stone, G. O. (1986). An analysis of the delta rule and the learning of statistical associations. In D. E. Rumelhart, J. L. McClelland & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 444-459). Massachusetts: The MIT Press.

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 3-45). Baltimore: York Press.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39(4), 456-466.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, XVIII (6), 643-662.

Studdert-Kennedy, M. (1976). Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 243-294). New York: Academic Press.

Styles, E. A. (2006). *The psychology of attention* (2^{nd} ed.). Psychology Press.

Szczygłowska, T. (2017). Advanced learners' attitudes towards their target language pronunciation: A study with Polish students of English. *Linguistica Silesiana*, 38, 373-401.

Szyszka, M. (2017). *Pronunciation learning strategies and language anxiety: In search of an interplay*. Switzerland: Springer.

Thagard, P. (2005). *Mind: Introduction to cognitive science* (2nd ed.). Massachusetts: The MIT Press.

Tipper, S. P., & McLaren, J. (1990). Evidence for efficient visual selectivity in children. In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 197-210). The Netherlands: Elsevier Science Publishers.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. Journal of Acoustical Society of America, 88(1), 97-100.

Trehub, S. E., & Trainor, L. J. (1990). Rules for listening in infancy. In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 87-120). The Netherlands: Elsevier Science Publishers.

Treisman, A. M. (1964). Monitoring and storage or irrelevant messages in selective attention[1]. *Journal of Learning and Verbal Behavior*, 3, 449-459.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97-136.

Treisman, A. M., Vieira, A., & Hayes, A. (1992). Automaticity and preattentive processing. *American Journal of Psychology*, 105(2), 341-362.

Trubetzkoy, N. S. (1969). *Principles of phonology* (C. A. M. Baltaxe, Trans.). University of California Press.

Tubach, J-P. (1989). *La parole et son traitement automatique*. Paris: Masson.

Von Neumann, J. (1958). *The computer and the brain*. Yale University Press.

Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Development*, 81(2), 517-527.

Waniek-Klimczak, E., Rojczyk, A., & Porzuczek, A. (2015). 'Poliglish' in Polish eyes: What English studies majors think about their pronunciation in English. In E. Waniek-

Klimczak & M. Pawlak (Eds.), *Teaching and researching the pronunciation of English* (pp. 23-34). Dordrecht: Springer.

Watson, J. B. (1930). *Behaviorism*. London, Engalnd: Kegan Paul, Trench, Trubner & Co.

Werker, J. F., & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In C. A. Ferguson, L, Menn, & C, Stoel Gammon (Eds.), *Phonological development: Models, research, and implications* (pp. 285-311). Timonium, Maryland: York Press.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganisation during the first year of life. *Infant Behavior and Development*, 7, 49-63.

Werker, J. F., & Tees, R. C. (1992). The organization and reorganization of human speech perception. *Annual Review of Neuroscience*, 15, 377-402.

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, 50, 509-535.

Wertheimer, M. (1938). Laws of organization in perceptual forms (W. D. Ellis, Trans.). In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71-88). London, England: Routledge & Kegan Paul.

Wilder, L. (1975). Articulatory and acoustic characteristics of speech sounds. In D. W. Massaro (Ed.), *An information-processing analysis of speech perception, reading, and psycholinguistics* (pp. 31-76). New York: Academic Press.

Wode, H. (1994). Speech perception and the learnability of languages. *International Journal of Applied Linguistics*, 4(2), 143-168.

Wundt, W. (1897). *Outlines of psychology*. (C. H. Judd, Trans.). Leipzig: Germany: Wilhelm Engelmann.

Wyer, Jr., R. S., & Albarracín, D. (2005). Belief formation, organization, and change: Cognitive and motivational influences. In D. Albarracín, B. T. Johnson & M. P. Zanna (Eds.), *The* *handbook of attitudes* (pp. 273-322). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Zampini, M. L. (1998). The relationship between the production and perception of L2 Spanish stops. *Texas Papers in Foreign Language Education*, 3(3), 85-100.

Zisper, D., & Rabin, D. E. (1986). P3: A parallel network simulating system. In D. E. Rumelhart, J. L. McClelland & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. *Volume 1: Foundations* (pp. 488-506). Massachusetts: The MIT Press.

Appendix 1. Reported Attitudes per Statement

Frequency of Participants' Scores per Statement

	5-point Likert scale				
Survey statements	1	2	3	4	5
1. How do you feel when you use correct RP English?	1.9	1.9	1.9	45.3	49.1
2. *How do you feel when your classmates mispronounce RP English?* (-)	71.7	26.4	0.0	1.9	0.0
3. How do you feel when a teacher comments negatively on your pronunciation?	37.7	5.7	45.3	5.7	5.7
4. *How do you feel when you listen to distorted English input?* (-)	86.8	5.7	3.8	1.9	0.0
5. What do you think of possessing skilled RP pronunciation habits for a future teacher of English?	0.0	5.7	17.0	50.9	26.4
6. What do you think of possessing skilled RP pronunciation habits in comparison to accurate grammar of English?	1.9	18.9	64.2	9.4	5.7
7. What do you think of possessing skilled RP pronunciation habits in comparison to rich vocabulary?	0.0	11.3	47.2	20.8	20.8
8. What do you think of possessing skilled RP pronunciation habits in comparison to fluency?	0.0	11.3	71.7	7.5	9.4
9. What do you think of English material used in teaching phonetics and phonology of English?	0.0	3.8	22.6	32.1	41.5
10. What do you think of achieving a native-like accent?	0.0	5.7	17.0	45.3	32.1
11. What do you think of the need for native speakers around to improve one's pronunciation?	0.0	1.9	1.9	20.8	75.5
12. What do you think of skilled English pronunciation if you want to integrate in an English speaking community?	0.0	3.8	18.9	45.3	32.1
13. Perception of English sounds is responsible for pronunciation problems.	3.8	7.5	11.3	73.6	3.8
14. What do you think of the need for effective pronunciation feedback?	0.0	0.0	1.9	34.0	64.2
15. How intense is your motivation to have skilled pronunciation habits?	1.9	1.9	5.7	37.7	52.8
16. *A teacher in middle and secondary schools does not need skilled English pronunciation habits.* (-)	49.1	43.4	1.9	3.8	1.9
17. How often do you train yourself at correct pronunciation?	1.9	11.3	43.4	26.4	17.0
18. How often do you listen to authentic English to improve your perceptual abilities?	0.0	11.3	56.6	22.6	9.4
19. How often do you self-monitor your pronunciation?	5.7	35.8	34.0	18.9	5.7
20. How often do you interact in group to check each other's pronunciation?	20.8	41.5	26.4	9.4	1.9

Values, 1 and 5, are opposite ends of a continuum of satisfaction (1, 2, 3, 4), importance (5, 6, 7, 8, 12), probability (10), agreement (13, 16), usefulness (11, 14), frequency (17, 18, 19, 20), and intensity (15). The value 5 indicates the continuum extreme positive end, while 1 indicates the extreme negative of it.

Appendix 2. Synthesised Vowels Frequencies

Formant Frequencies of Synthesised RP Monophthong Vowel Tokens Used in Experiment 1

Series N$_{o}$.	Sound N$_{o}$.	F$_1$ (Hz)	F$_2$ (Hz)	Series N$_{o}$.	Sound N$_{o}$.	F$_1$ (Hz)	F$_2$ (Hz)	Series N$_{o}$.	Sound N$_{o}$.	F$_1$ (Hz)	F$_2$ (Hz)
Series 1	S$_{1}$	296	2241	Series 4	S$_{18}$	680	1193	Series 7	S$_{36}$	532	1656
	S$_{2}$	310	2184		S$_{19}$	675	1219		S$_{37}$	529	1594
	S$_{3}$	325	2126		S$_{20}$	671	1245		S$_{38}$	526	1532
	S$_{4}$	339	2069		S$_{21}$	666	1270		S$_{39}$	522	1470
	S$_{5}$	353	2011		S$_{22}$	661	1296		S$_{40}$	519	1408
	S$_{6}$	367	1954								
	S$_{7}$	382	1896	Series 5	S$_{23}$	680	1193	Series 8	S$_{41}$	519	1408
	S$_{8}$	396	1839		S$_{24}$	678	1255		S$_{42}$	479	1408
					S$_{25}$	676	1317		S$_{43}$	439	1408
Series 2	S$_{9}$	386	1587		S$_{26}$	674	1379		S$_{44}$	399	1408
	S$_{10}$	389	1527		S$_{27}$	671	1441		S$_{45}$	359	1408
	S$_{11}$	392	1467		S$_{28}$	669	1503				
	S$_{12}$	395	1407		S$_{29}$	667	1565	Series 9	S$_{46}$	519	1408
									S$_{47}$	486	1453
Series 3	S$_{13}$	480	857	Series 6	S$_{30}$	676	1565		S$_{48}$	453	1498
	S$_{14}$	521	898		S$_{31}$	673	1511		S$_{49}$	419	1542
	S$_{15}$	562	938		S$_{32}$	670	1457		S$_{50}$	386	1587
	S$_{16}$	602	979		S$_{33}$	667	1404				
	S$_{17}$	643	1019		S$_{34}$	664	1350				
					S$_{35}$	661	1296				

Appendix 3. Descriptive Statistics for Discrimination Hit Rates

Descriptive Statistics for Discrimination Hit Rate (original data)

Stimulus pair	*M*	*SD*	Variance
1_8	0.97	0.07	0.005
8_1	0.99	0.05	0.002
2_7	0.95	0.09	0.009
7_2	0.96	0.09	0.008
3_6	0.83	0.17	0.029
6_3	0.87	0.15	0.021
4_5	0.25	0.28	0.080
5_4	0.23	0.23	0.055
4_4	0.92	0.15	0.022
9_12	0.73	0.33	0.110
12_9	0.68	0.31	0.098
10_11	0.39	0.23	0.053
11_10	0.37	0.27	0.072
11_11	0.82	0.16	0.027
13_17	0.98	0.06	0.004
17_13	0.98	0.06	0.003
14_16	0.89	0.17	0.030
16_14	0.87	0.19	0.038
14_15	0.28	0.24	0.056
15_14	0.28	0.24	0.058
15_16	0.73	0.26	0.070
16_15	0.64	0.30	0.091
14_14	0.90	0.11	0.012
18_22	0.45	0.27	0.072
22_18	0.45	0.28	0.080
19_20	0.29	0.22	0.048
20_19	0.28	0.20	0.041
19_21	0.29	0.21	0.045
21_19	0.32	0.22	0.048
20_21	0.33	0.22	0.049
21_20	0.31	0.23	0.053
21_21	0.68	0.22	0.047
23_29	0.98	0.04	0.002
29_23	0.97	0.05	0.003
24_28	0.96	0.10	0.010
28_24	0.96	0.09	0.007
26_25	0.38	0.25	0.064
25_26	0.38	0.27	0.074
25_27	0.70	0.26	0.067
27_25	0.70	0.27	0.070
26_27	0.34	0.27	0.072
27_26	0.30	0.21	0.044
25_25	0.86	0.16	0.025

Note. N = 52

Appendix 3. Descriptive Statistics for Discrimination Hit Rates

Descriptive Statistics for Discrimination Hit Rate (original data)

Stimulus pair	*M*	*SD*	Variance
30_35	0.87	0.22	0.049
35_30	0.86	0.20	0.038
31_34	0.78	0.23	0.051
34_31	0.72	0.26	0.069
32_33	0.34	0.20	0.042
33_32	0.36	0.20	0.038
33_33	0.80	0.17	0.028
36_40	0.77	0.26	0.065
40_36	0.75	0.26	0.069
38_37	0.34	0.19	0.037
37_38	0.31	0.20	0.040
37_39	0.51	0.26	0.069
39_37	0.49	0.28	0.079
38_39	0.32	0.23	0.053
39_38	0.33	0.23	0.054
39_39	0.46	0.28	0.078
41_45	0.98	0.06	0.003
45_41	0.99	0.03	0.001
42_43	0.83	0.17	0.029
43_42	0.89	0.17	0.028
42_44	0.93	0.11	0.013
44_42	0.93	0.14	0.019
43_44	0.13	0.16	0.027
44_43	0.12	0.14	0.019
44_44	0.93	0.12	0.013
46_50	0.98	0.04	0.002
50_46	0.98	0.06	0.004
47_49	0.93	0.12	0.016
49_47	0.97	0.09	0.007
47_48	0.91	0.15	0.022
48_47	0.94	0.13	0.017
49_48	0.12	0.15	0.021
48_49	0.12	0.16	0.024
48_48	0.93	0.11	0.011

Note. N = 52

Appendix 4. Descriptive Statistics for Mean Reaction Time

Descriptive Statistics for Reaction Time (original data)

Stimulus pair	*M*	*SD*	Variance
1_8	5.05	3.99	15.88
8_1	5.31	3.67	13.49
2_7	5.27	3.40	11.58
7_2	4.79	3.60	12.95
3_6	6.46	4.38	19.20
6_3	5.36	3.99	15.95
4_5	5.47	3.60	12.94
5_4	6.52	4.22	17.78
4_4	5.15	3.15	9.94
9_12	4.64	3.21	10.28
12_9	4.57	3.50	12.26
10_11	5.12	4.15	17.26
11_10	5.13	3.79	14.39
11_11	4.60	3.29	10.84
13_17	3.96	2.77	7.69
17_13	3.98	2.63	6.91
14_16	5.11	3.53	12.49
16_14	5.03	3.79	14.35
14_15	4.42	2.96	8.76
15_14	4.93	3.72	13.87
15_16	3.59	2.86	8.19
16_15	3.91	2.72	7.42
14_14	5.25	3.52	12.39
18_22	5.32	3.93	15.47
22_18	5.51	4.27	18.26
19_20	5.26	3.22	10.38
20_19	5.33	3.52	12.38
19_21	5.66	4.69	22.03
21_19	5.19	3.23	10.44
20_21	6.44	4.73	22.42
21_20	6.07	4.53	20.49
21_21	6.13	4.18	17.45
23_29	4.57	3.52	12.39
29_23	4.48	3.07	9.41
24_28	4.23	3.39	11.47
28_24	4.83	4.11	16.86
26_25	3.95	2.66	7.09
25_26	4.47	3.08	9.50
25_27	4.48	3.04	9.25
27_25	4.17	3.31	10.98
26_27	5.05	3.99	15.88
27_26	5.31	3.67	13.49
25_25	5.27	3.40	11.58

Note. N = 52

Appendix 4. Descriptive Statistics for Mean Reaction Time

Descriptive Statistics for Reaction Time (original data)

Stimulus pair	*M*	*SD*	Variance
30_35	4.94	3.34	11.13
35_30	4.24	2.61	6.81
31_34	4.50	2.94	8.65
34_31	4.99	3.94	15.50
32_33	4.90	3.56	12.69
33_32	4.30	2.59	6.72
33_33	4.30	2.83	8.01
36_40	6.33	4.11	16.91
40_36	4.94	3.01	9.05
38_37	4.53	3.27	10.69
37_38	5.25	4.40	19.39
37_39	5.34	3.57	12.74
39_37	5.79	3.44	11.87
38_39	5.64	4.02	16.18
39_38	5.28	3.72	13.81
39_39	5.24	3.65	13.30
41_45	5.42	4.30	18.51
45_41	4.81	3.17	10.06
42_43	5.43	3.29	10.82
43_42	3.60	2.67	7.12
42_44	4.38	2.99	8.93
44_42	4.77	3.70	13.67
43_44	3.95	2.78	7.72
44_43	4.86	3.29	10.83
44_44	4.01	2.61	6.79
46_50	4.72	3.04	9.23
50_46	5.69	4.01	16.11
47_49	3.87	2.42	5.85
49_47	4.26	3.51	12.33
47_48	4.35	3.38	11.45
48_47	4.63	2.86	8.19
49_48	5.35	3.53	12.44
48_49	4.56	3.23	10.41
48_48	4.46	3.40	11.58

Note. N = 52

Appendix 5. Descriptive Statistics for Mean Discrimination Hit Rate

Descriptive Statistics for Discrimination Hit Rate (modified data)

Stimulus pair	*M*	*SE*	Variance
1_8 & 8_1	0.98	0.01	0.003
2_7 & 7_2	0.96	0.01	0.007
3_6 & 6_3	0.85	0.02	0.019
4_5 & 5_4	0.24	0.03	0.058
4_4	0.92	0.02	0.022
9_12 & 12_9	0.71	0.04	0.096
10_11 & 11_10	0.38	0.03	0.053
11_11	0.82	0.02	0.027
13_17 & 17_13	0.98	0.01	0.003
14_16 & 16_14	0.88	0.02	0.029
14_15 & 15_14	0.28	0.03	0.045
15_16 & 16_15	0.68	0.04	0.067
14_14	0.90	0.02	0.012
18_22 & 22_18	0.45	0.03	0.061
19_20 & 20_19	0.28	0.03	0.037
19_21 & 21_19	0.30	0.03	0.033
20_21 & 21_20	0.32	0.03	0.044
21_21	0.68	0.03	0.047
23_29 & 29_23	0.98	0.01	0.002
24_28 & 28_24	0.96	0.01	0.006
25_26 & 26_25	0.38	0.03	0.055
25_27 & 27_25	0.70	0.03	0.054
26_27 & 27_26	0.32	0.03	0.043
25_25	0.86	0.02	0.025
30_35 & 35_30	0.86	0.03	0.039
31_34 & 34_31	0.75	0.03	0.052
32_33 & 33_32	0.35	0.02	0.029
33_33	0.80	0.02	0.028
36_40 & 40_36	0.76	0.03	0.061
37_38 & 38_37	0.33	0.02	0.027
37_39 & 39_37	0.50	0.03	0.060
38_39 & 39_38	0.32	0.03	0.042
39_39	0.46	0.04	0.078
41_45 & 45_41	0.99	0.01	0.001
42_43 & 43_42	0.86	0.02	0.020
42_44 & 44_42	0.93	0.02	0.013
43_44 & 44_43	0.13	0.02	0.018
44_44	0.93	0.02	0.013
46_50 & 50_46	0.98	0.01	0.002
47_49 & 49_47	0.95	0.01	0.009
47_48 & 48_47	0.93	0.02	0.014
48_49 & 49_48	0.12	0.02	0.020
48_48	0.93	0.01	0.011

Note. N = 52

Appendix 5. Descriptive Statistics for Mean Discrimination Reaction Time

Descriptive Statistics for Discrimination Reaction Time (modified data)

Stimulus pair	*M*	*SE*	Variance
1_8 & 8_1	5.18	0.38	7.508
2_7 & 7_2	5.03	0.35	6.472
3_6 & 6_3	5.91	0.47	11.667
4_5 & 5_4	6.00	0.41	8.807
4_4	5.15	0.44	9.938
9_12 & 12_9	4.60	0.34	6.103
10_11 & 11_10	5.13	0.37	7.218
11_11	4.60	0.46	10.840
13_17 & 17_13	3.97	0.26	3.479
14_16 & 16_14	5.07	0.39	7.820
14_15 & 15_14	4.68	0.32	5.462
15_16 & 16_15	3.75	0.29	4.367
14_14	5.25	0.49	12.386
18_22 & 22_18	5.41	0.42	9.282
19_20 & 20_19	5.29	0.35	6.510
19_21 & 21_19	5.43	0.39	7.903
20_21 & 21_20	6.26	0.50	13.218
21_21	6.13	0.58	17.454
23_29 & 29_23	4.52	0.31	5.046
24_28 & 28_24	4.53	0.32	5.176
25_26 & 26_25	4.21	0.27	3.722
25_27 & 27_25	4.33	0.33	5.597
26_27 & 27_26	4.59	0.28	4.207
25_25	4.50	0.41	8.649
30_35 & 35_30	4.95	0.36	6.666
31_34 & 34_31	4.30	0.27	3.929
32_33 & 33_32	5.63	0.30	4.777
33_33	4.53	0.45	10.692
36_40 & 40_36	5.30	0.35	6.487
37_38 & 38_37	5.72	0.38	7.378
37_39 & 39_37	5.26	0.35	6.496
38_39 & 39_38	5.11	0.36	6.720
39_39	5.43	0.46	10.824
41_45 & 45_41	3.99	0.26	3.555
42_43 & 43_42	4.36	0.34	5.993
42_44 & 44_42	4.43	0.32	5.301
43_44 & 44_43	5.21	0.33	5.633
44_44	3.87	0.34	5.850
46_50 & 50_46	4.31	0.34	6.104
47_49 & 49_47	4.99	0.31	5.075
47_48 & 48_47	4.51	0.30	4.577
48_49 & 49_48	4.71	0.35	6.496
48_48	4.72	0.42	9.385

Note. N = 52

Appendix 6. Sphericity tests for Mean Hit Rate across Pairs of Vowel Tokens

Sphericity Test for Mean Hit Rate for Combined Pairs Series 1

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 1	.068	133.557	5	.000	.452	.460	.333

Sphericity Test for Mean Hit Rate for Combined Pairs Series 2

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 2	1.000	.000	0	.	1.000	1.000	1.000

Sphericity Test for Mean Hit Rate for Combined Pairs Series 3

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 3	.334	54.562	5	.000	.578	.596	.333

Sphericity Test for Mean Hit Rate for Combined Pairs Series 4

Within subjects effects	Mauchly's W	Approx.Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 4	.855	7.764	5	.170	.907	.963	.333

Sphericity Test for Mean Hit Rate for Combined Pairs Series 5

Within subjects effects	Mauchly's W	Approx.Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 5	.158	91.041	9	.000	.961	.735	.250

Sphericity Test for Mean Hit Rate for Combined Pairs Series 6

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 6	.724	16.126	2	.000	.784	.804	.500

Sphericity Test for Mean Hit Rate for Combined Pairs Series 7

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 7	.808	10.576	2	.061	.869	.920	.333

Sphericity Test for Mean Hit Rate for Combined Pairs Series 8

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 8	.302	59.474	5	.000	.567	.585	.333

Sphericity Test for Mean Hit Rate for Combined Pairs Series 9

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 9	.133	100.441	5	.000	.457	.466	.333

Sphericity Test for Mean Hit Rate for Main Prototypic Vowel Pairs

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Main vowel pairs	.000	569.654	35	.000	.409	.440	.125

Appendix 6. Sphericity tests for Mean Reaction Time across Pairs of Vowel Tokens

Sphericity Test for Mean Reaction Time Series 1

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 1	.906	4.928	5	.425	.940	1.000	.333

Sphericity Test for Mean Reaction Time Series 2

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 2	1.000	.000	0	.	1.000	1.000	1.000

Sphericity Test for Mean Reaction Time Series 3

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 3	.904	5.023	5	.413	.936	.996	.333

Sphericity Test for Mean Reaction Time Series 4

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 4	.873	6.758	5	.215	.922	.980	.333

Sphericity Test for Mean Reaction Time Series 5

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 5	..867	7.031	9	.634	.938	1.000	.250

Sphericity Test for Mean Reaction Time Series 6

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 6	.940	3.0710	2	.215	.944	.979	.500

Sphericity Test for Mean Reaction Time Series 7

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 7	.899	5.313	5	.379	.935	.995	.333

Sphericity Test for Mean Reaction Time Series 8

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 8	.877	6.5000	5	.261	.931	.991	.333

Sphericity Test for Mean Reaction Time Series 9

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Series 9	.930	3.620	5	.379	.952	1.000	.333

Sphericity Test for Mean Reaction Time for All Pairs

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
All pairs	.000	1152.264	902	.000	.489	.833	.024

Within-Subjects Effects for Mean Reaction Time for All Vowel Pairs

Source: Vowel pair	*df*	*F*	*Sig.*	ES η^2	Observed power [a]
Sphericity Assumed	42	2.869	.000	.053	1.000
Greenhouse-Geisser	20.536	2.869	.000	.053	1.000
Huynh-Feldt	34.988	2.869	.000	.053	1.000
Lower-bound	1	2.869	.096	.053	.383

[a]. Computed using alpha = .05

Appendix 7. Sphericity Tests for Mean Hit Rate across Short Vowels

Sphericity Test for Mean Identification of Short Vowels at 90 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.392	46.915	14	.000	.752	.818	.200

Sphericity Test for Mean Identification of Short Vowels at 110 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.368	50.091	14	.000	.727	.789	.200

Sphericity Test for Mean Identification of Short Vowels at 130 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.437	41.476	14	.000	.770	.839	.200

Sphericity Test for Identification Hit Rate for Short Vowels at 150 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.333	55.088	14	.000	.729	.791	.200

Sphericity Test for Mean Identification of Short Vowels at 170 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.429	42.429	14	.000	.825	.905	.200

Sphericity Test for Mean Identification of Short Vowels at 190 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.328	55.835	14	.000	.746	.832	.200

Appendix 8. Sphericity Tests for Mean Hit Rate across Long Vowels

Sphericity Test for Mean Identification of Long Vowels at 150 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.225	75.235	9	.000	.577	.606	.250

Sphericity Test for Mean Identification of Long Vowels at 170 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.596	26.124	9	.002	.777	.832	.832

Sphericity Test for Mean Identification of Long Vowels at 190 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.890	5.861	9	.754	.950	1.000	.250

Sphericity Test for Mean Identification of Long Vowels at 210 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.745	14.847	9	.095	.864	.933	.250

Sphericity Test for Mean Identification of Long Vowels at 230 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.499	35.095	9	.000	.801	.860	.250

Sphericity Test for Mean Identification of Long Vowels at 250 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.504	34.499	9	.000	.779	.835	.250

Sphericity Test for Mean Identification of Long Vowels at 270 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.356	52.095	9	.000	.710	.755	.250

Sphericity Test for Mean Identification of Long Vowels at 290 ms

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.331	55.785	9	.000	.703	.747	.250

Sphericity Test for Aggregate Mean Identification of Short Vowels

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.428	42.557	14	.000	.752	.818	.200
Duration	.444	40.705	14	.000	.712	.771	.200
Vowel*Duration	.000	626.931	324	.000	.436	.560	.040

Sphericity Test for Aggregate Mean Identification of Long Vowels

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.492	35.765	9	.000	.785	.842	.250
Duration	.011	221.530	27	.000	.466	.500	.143
Vowel*Duration	.000	835.054	405	.000	.441	.587	.036

Sphericity Test for Aggregate Identification of All Vowels

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	.019	193.158	54	.000	.562	.637	.100

Appendix 9. Sphericity Tests for Vowel Mean Identification in Consonantal Context

Sphericity Test for Mean Identification of Vowel (Word Series 1)

					Epsilon		
Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
/b n/	.444	40.924	9	.000	.811	.872	.250

Sphericity Test for Mean Identification of Vowel (Word Series 3)

					Epsilon		
Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
/ˈm .tə/	.652	21.687	5	.001	.825	.869	.333

Sphericity Test for Mean Identification of Vowel (Word Series 5)

					Epsilon		
Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
/ˈb .tə/	.963	1.879	2	.387	.965	1.000	.500

Sphericity Test for Mean Identification of Vowel (Word Series 6)

					Epsilon		
Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
/m ʧ/	.889	5.977	2	.050	.900	.931	.500

Sphericity Test for Mean Identification of Vowel (Word Series 9)

					Epsilon		
Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
/b d/	.436	41.858	9	.000	.779	.835	.250

Sphericity Test for Mean Identification of Vowel (Word Series 10)

					Epsilon		
Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
/b t/	.657	21.155	9	.012	.833	.897	.250

Sphericity Test for Mean Identification of Vowel (Word Series 14)

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
/k b/	.880	6.473	5	.263	.931	.990	.333

Sphericity Test for Mean Identification of Vowel (Word Series 15)

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
/k d/	.914	4.570	2	.102	.921	.953	.500

Sphericity Test for Mean Identification of Vowel (Word Series 17)

Within subjects effects	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
/b k/	.840	8.867	2	.012	.862	.889	.500

Appendix 10. Sphericity Tests for Vowel Mean Identification and Reaction Time and their Interaction

Sphericity Test for Vowel Mean Identification and Reaction Time and Their Interaction

						Epsilon		
Within subjects effects	Measure	Mauchly’s W	Approx. Chi-Square	*df*	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Vowel	Mean identification	.122	103.175	35	.000	.692	.784	.125
	Mean reaction time	.376	48.096	35	.070	.783	.902	.125
Context	Mean identification	.868	7.215	2	.027	.883	.912	.500
	Mean reaction time	.963	1.919	2	.383	.964	1.000	.500
Vowel*Context	Mean identification	.000	.	135	.	.475	.565	.063
	Mean reaction time	.001	332.444	135	.000	.547	.668	.063

Appendix 11. List of Word Tokens

Series	*Words*				
Series 1	/bæn/	/bʌn/	/bɑːn/	/bɒn/	/bɔːn/
Series 2	/ʧæd/	/ʧɑːd/	/ʧæt/	/ʧɑːt/	
Series 3	/ˈmætə/	/ˈmʌtə/	/ˈmɑːtə/	/ˈmɔːtə/	
Series 4	/gæb/	/gɑːb/	/bæʤ/	/bʌʤ/	
Series 5	/ˈbætə/	/ˈbʌtə/	/ˈbɑːtə/	/ˈflætə/	/ˈflʌtə/
Series 6	/mæʧ/	/mʌʧ/	/mɑːʧ/		
Series 7	/bɪn/	/biːn/	/ɪʧ/	/iːʧ/	
Series 8	/ˈlɪvə/	/ˈliːvə/	/pɪʧ/	/piːʧ/	
Series 9	/bæd/	/bʌd/	/bɑːd/	/bɒd/	/bɔːd/
Series 10	/bæt/	/bʌt/	/bɑːt/	/bɒt/	/bɔːt/
Series 11	/ˈfeðə /	/ˈfɜːðə/	/wɜːnt/	/went/	
Series 12	/bɪʧ/	/bʊʧ/	/hɪd/	/h̲ʊd̲/	/sɪt/
Series 13	/sʊt/	/suːt/	/fʊl/	/fuːl/	
Series 14	/kæb/	/kɑːb/	/kʌb/	/kɒb/	
Series 15	/kɒd/	/kɑːd/	/kɔːd/	/kɒn/	/kɔːn/
Series 16	/ˈhɒtɪ/	/ˈhɑːtɪ/	/ kɒn/	/kæn/	
Series 17	/bæk/	/bʌk/	/bɑːk/	/ʧɔːk/	/ʧʌk/
Series 18	/kæb/	/kɑːb/	/kʌb/	/kɒb/	
Series 19	/fɒnd/	/fʌnd/	/ʃɒt/	/ʃʌt/	
Series 20	/ˈsʌkə/	/ˈsɒkə/	/ˈsʌkə/	/ˈsɒkə/	
Series 21	/ˈbætl/	/ˈbɒtl/	/ˈwægl/	/ˈwɒgl/	

Résumé

Ce travail étudie la perception des voyelles prototypiques de l’anglais standard britannique appelé RP parmi un groupe de 53 étudiants de la 3ème année d’anglais à l'École Normale Supérieure à Constantine, et d’explorer leurs attitudes vis-à-vis de la prononciation de l’anglais. L’étude emploie un protocole expérimental mixte pour étudier les phénomènes prévus. Trois conditions expérimentales ont été manipulées dans cette étude : (a) la discrimination de la distance spectrale entre certaines voyelles prototypiques de l’anglais, en utilisant le protocole A*X*, (b) l'identification des voyelles prototypiques de l’anglais en isolation à travers une manipulation temporelle, en utilisant un test à choix multiple imposé, et (c) ce même dernier test pour l'identification des voyelles en contextes phonétiques. La présente recherche emploie un questionnaire des attitudes afin d’explorer ces derniers dans le contexte de la prononciation de l’anglais. Les résultats du premier test ont montré une sensibilité perceptuelle remarquable aux distances spectrales parmi 8 contrastes, avec un *d’* s'étendant entre 1.87 pour (/uː/contre /ʊ/) et 5.76 pour (/ɜː/contre/ʊ/), et à un indice *d’* nul pour le contraste (/ɑː/contre/ʌ/). Le deuxième test d’identification de voyelles en isolation a montré l’ordre d'identification suivant : /ʊ/ > /ɪ/ > /e/ > /ʌ/ > /æ/ > /ɒ/, avec /ɒ/ comme étant significativement la voyelle la moins identifiée. Pour les voyelles longues, l’ordre a été comme suit : /iː/ > /ɔː/ > /ɑː/ > /ɜː/ > /uː/, avec /uː/ significativement la moins identifiée. Dans le test d’identification de voyelles en contexte, l’ordre d’identification a été comme suit : /i ː/ > /ɔː/ > /ɑː/ > /ɜː/ > /ɪ/ > /e/ > /ʌ/ > /æ/ > /ɒ/ (/ʊ/, et /uː/ n'ont pas été inclues dans l'analyse). Un aperçu global des résultats du questionnaire montre une attitude positive globale parmi les étudiants vis-à-vis de l’apprentissage de la prononciation de l’anglais et une réticence ressentie à participer à des activités auto-initiées d’apprentissage de la prononciation de l’anglais. La recherche interprète les résultats obtenus dans une perspective de recherche de la perception de la parole.

ملخص

يتناول هذا البحث إدراك الحروف الصائتة للغة الإنجليزية البريطانية لدى 53 طالب جزائري من السنة الثالثة الإنجليزية في المدرسة العليا للأساتذة قسنطينة ومواقفهم تجاه نطق اللغة الإنجليزية. استخدم البحث تصميم بحث مختلط للتحقيق في المتغيرات حيث تم التعامل مع ثلاثة شروط تجريبية: (أ) اختبار ثنائية الاختلاف أو التشابه للتمييز بين المسافة الطيفية بين 9 ازواج من الحروف الصائتة الإنجليزية، (ب) اختبار بديل قسري لتحديد الحروف الصائتة الإنجليزية منعزلة عبر التغيرات الزمنية، (ج) اختبار بديل قسري لتحديد الحروف الصائتة النموذجية في سياقات صوتية مختلفة. كما استخدم البحث استبيانا لاستكشاف مواقف المتعلمين الجزائريين من نطق اللغة الإنجليزية. في اختبار ثنائية الاختلاف أو التشابه، أظهرت نتائج اختبار التمييز حساسية الإدراك العالية للمسافات الطيفية بين 8 ازواج من الحروف الصائتة النموذجية، مع مؤشر حساسية *d’* يتراوح بين 1.87 لـ (/uː/ ضد /ʊ/) و5.76 لـ (/ɜː/ ضد /ʊ/)، ومؤشر حساسية *d’* منعدم لـ (/ɑː/ ضد /ʌ/). أظهر الاختبار البديل القسري الأول لتحديد الحروف المعزولة عبر التغيير الزمني نمط التحديد التالي: /ɒ/ < /æ/ < /ʌ/ < /e/ < /ɪ/ < /ʊ/، مع كون /ɒ/ الأقل تحديدا. بالنسبة للحروف الصائتة الطويلة، كان النمط: /uː/ < /ɜː/ < /ɑː/ < /ɔː/ < /iː/، مع كون/uː/ الأقل تحديدا. في الاختبار البديل القسري الثاني، تتبع الحروف الصائتة في السياق نمطًا مختلفًا عن الاختبارين السابقين، حيث تم تحديد جميع الحروف الصائتة بدقة وبنسبة إصابة تتراوح بين .74 إلى .98. كما أظهر الاستطلاع موقفا إيجابيا بشكل عام تجاه تعلم النطق باللغة الإنجليزية وتعليمه واحجام محسوس في المشاركة في أنشطة مستقلة لتعلم النطق. ناقش البحث النتائج التي تم الحصول عليها ضمن النظريات الشائعة حول إدراك الكلام عبر اللغة، مكانة نطق اللغة الإنجليزية بين الطلاب الجزائريين، وأوصى بإجراء مزيد من الأبحاث حول إدراك الكلام في السياق الجزائري.