The Dissertation Committee for Sang-Hoon Park certifies that this is the approved version of the following dissertation:

Quantifying Perceptual Contrast: The Dimension of Place of Articulation

Committee:

Robert T. Harms, Co-Supervisor

Björn E. Lindblom, Co-Supervisor

Randy L. Diehl

Harvey M. Sussman

Richard P. Meier

Quantifying Perceptual Contrast: The Dimension of Place of Articulation

by

Sang-Hoon Park, B.A., M.A.

Dissertation

Presented to the Faculty of the Graduate School of

the University of Texas at Austin

in Partial Fulfillment

of the requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2007

Dedication

To My Family

Acknowledgements

I would like to express my sincerest appreciation to Dr. Björn Lindblom who led me into this wonderful world of Phonetics and guided me through the course of the research not only as a supervisor but as a role model. Without his invaluable comments, encouragements and support, this work would not have been possible.

I am much indebted to Dr. Robert Harms. As a Graduate Advisor and a supervisor, he provided precious comments as well as encouragements and help. He showed me how a scholar should be awake.

My thanks go to Dr. Randy Diehl who provided me with invaluable comments especially for the experimental design. He also kindly offered me to use the lab facilities and an opportunity to work in his lab.

Dr. Harvey Sussman's Neurolinguistics class was one of the most interesting and impressive courses I have ever taken. His witty comments on my thesis also helped me focus on the topic.

I also thank Dr. Richard Meier who willingly agreed to be my committee member and gave me support at every stage of the research.

The life in the Linguistics department was enriched and animated by my dear colleagues and friends: Jeong-Hoon Lee, Hansang Park, Incheol Choi, Chisung Oh, Q Wan Kim, Dong-Rhin Shin, Young-Ok Cho, So-Hee Kim, Seung-Eun Chang, Hyun-Jong Hahm, Hee Young Lyu, Jinung Kim, Jihwan Kim, Mi Jang, Yunjung Bae and Taesun Moon, to name a few.

Finally, my deepest thanks and love will go to my wife, KYUNG PARK and my daughter, JI-WOON.

Quantifying Perceptual Contrast: The Dimension of Place of Articulation

Publication No.______________

Sang-Hoon Park, Ph.D.
The University of Texas at Austin, 2007

C0-supervisors: Robert T. Harms and Björn E. Lindblom

This study investigates the role of perceptual distinctiveness in consonant inventories. While distinctiveness appears to play a role in the shaping of vowel systems, a literature review indicates that its status in consonant selections remains unclear.

To address this issue I used speech materials recorded by a trained phonetician containing 35 CV syllables with seven places of articulation (bilabial, dental, alveolar, retroflex, palatal, velar and uvular) and five vowels: [i] [ɛ] [a] [ɔ] and [u].

Detailed acoustic measurements were performed: formant patterns at vowel onsets (loci) and vowel midpoints, transitions rates and burst spectra. To validate the speech material, comparisons were made with published data and with formant frequencies derived by means of an articulatory model.

Perceptual data were collected on these 35 syllables. Multiple Regression analyses were performed with the coded dissimilarities as the dependent variable and with (combinations of) formant-based distances, time constant differences and burst differences as the independent variables. The results indicated that acoustic measurements could be successfully used to help explain listener responses.

Optimal place sets were obtained from a rank ordering of the CV syllables with respect to ‘individual salience’ (defined as the sum of a syllable’s perceptual distance to other places in the same vowel context) and from a replication of the Liljencrants & Lindblom systemic criterion of maximizing distances within all vowel pairs. Instead of the typologically prevalent pattern of [b d g], predictions were found to be vowel-dependent and to often favor CV:s located at the ‘corners’ of the acoustic F3-F2 space, viz., uvular, palatal and retroflex.

This finding leads to a conclusion that distinctiveness alone is unlikely to account for how languages use place of articulation in voiced stops. For more successful attempts, future work should be directed towards defining and incorporating production constraints such as ‘ease of articulation’.

Table of Contents

List of Tables..xii

List of Figures..xvii

Chapter 1: Introduction...1

1.1 background...1

1.2 Goals..7

Chapter 2: Experimental Data Collection and Data Analyses...............................10

2.1 Materials..10

2.2 Acoustic analysis..11

2.2.1 Formant patterns...11

2.2.1.1 Measurements..11

2.2.1.2 F3 vs. F2 space at CV boundary...15

2.2.1.3 Locus equations...17

2.2.1.4 Comparison with previous theoretical predictions.................23

2.2.2 Time constants of formant transitions...30

2.2.3 Burst spectra...39

2.3 Perceptual data..43

2.3.1 Subjects and language groups...43

2.3.2 Procedures...48

2.3.2.1 Training sessions..48

2.3.2.2 Identification test..49

2.3.2.3 Dissimilarity judgment task..50

2.3.3 Results...51

2.3.3.1 Response bias...52

2.3.3.2 Identification test results...60

2.3.3.3 Dissimilarity judgment task results..73

2.3.3.4 Effect of language groups...84

Chapter 3: Acoustic Distances as a Measure of Perceptual Contrast...................86

3.1 Development of a measure of acoustic distance...86

3.1.1 Formant distances...86

3.1.2 Time constant differences..91

3.1.3 Burst spectra...98

3.2 Results...100

3.2.1 Formants...103

3.2.2 Formants and time constants..107

3.2.3 Formants, time constants and bursts...111

3.2.4 Summary of the results..115

Chapter 4: Articulatory Distances Combined with Acoustic Distances as an Extended Measure of Perceptual Contrast...118

4.1 Mirror neurons..118

4.2 Development of a measure of articulatory distance.................................125

4.2.1 X-ray data on CV:s...125

4.2.2 Articulatory distance defined as an RMS difference between articulatory contours..127

4.3 Results...128

Chapter 5: Preferred Consonant Place in Each Vowel Context...........................132

5.1 Simulating preferred vowel systems: the Liljencrants and Lindblom approach...132

5.2 Preferred consonant places based on maximal distinctiveness..............136

5.3 Replication of the method of Liljencrants and Lindblom.......................144

Chapter 6: Summary and conclusions..150

Appendix A: Formant patterns (i) at CV boundary and (ii) at vowel mid-point For each CV...156

Appendix B: Results of single factor ANOVA for the time constants..................191

Appendix C: Confusion matrices before symmetrization...................................193

Appendix D: Dissimilarity matrices before symmetrization..............................203

Appendix E: Results of multiple regression analyses with six independent variables (Formant-based distance, the time constant distances for F1, F2, and F3, burst spectra distance and articulatory distance)..213

Bibliography...216

Vita...225

List of Tables

Table 2.1: 35 CV syllables (consonants of 7 places of articulation x 5 vowels)......10

Table 2.2: Formants 1-4 measured at CV boundary and vowel mid-point...........15

Table 2.3: Locus equations and R^2 for the seven places of articulation...............22

Table 2.4: Parameters fixed for the APEX simulation of dorsal and bilabial stops..26

Table 2.5: Variable parameters for dorsal and bilabial stops..............................26

Table 2.6: Formant time constants (F1, F2 and F3) for all CVs............................36

Table 2.7: Example of similarity matrix for response bias calculation.................53

Table 2.8: Example of maximum response bias (dissimilarity judgment task)....55

Table 2.9: Response bias for identification test...56

Table 2.10: Example of maximum response bias (identification test)..................57

Table 2.11: Response bias for the dissimilarity judgment task.............................59

Table 2.12: Symmetrized responses for the identification task (Korean subjects)..62

Table 2.13: Symmetrized responses for the identification task (English subjects)...65

Table 2.14: Symmetrized responses for the identification task (Hindi subjects)..67

Table 2.15: Symmetrized responses for the identification task (Spanish subjects)..70

Table 2.16: Symmetrized responses for the identification task
(All subjects)..72

Table 2.17: Perceptual dissimilarity matrix for Korean subjects..........................74

Table 2.18: Perceptual dissimilarity matrix for English subjects.........................76

Table 2.19: Perceptual dissimilarity matrix for Hindi subjects............................79

Table 2.20: Perceptual dissimilarity matrix for Spanish subjects........................81

Table 2.21: Perceptual dissimilarity matrix averaged over language groups........83

Table 2.22: Correlation coefficients among language groups in
the dissimilarity judgment task...84

Table 3.1: Distance matrices based on formant 1-3 (in Mel)...............................90

Table 3.2: Distance matrices for formant time constants....................................97

Table 3.3: Distance matrices for bursts...100

Table 3.4: Results of multiple regression analysis with formant distance
as a predictor (vowels pooled)..104

Table 3.5: Results of multiple regression analysis with formant distance
as a predictor (/i/ vowel)..104

Table 3.6: Results of multiple regression analysis with formant distance
as a predictor (/e/ vowel)..105

Table 3.7: Results of multiple regression analysis with formant distance
as a predictor (/a/ vowel)...106

Table 3.8: Results of multiple regression analysis with formant distance
as a predictor (/o/ vowel)...106

Table 3.9: Results of multiple regression analysis with formant distance as a predictor (/u/ vowel)..107

Table 3.10: Results of multiple regression analysis with formant distance and time constant as predictors (vowels pooled)....................................108

Table 3.11: Results of multiple regression analysis with formant distance and time constant as predictors (/i/ vowel)..109

Table 3.12: Results of multiple regression analysis with formant distance and time constant as predictors (/e/ vowel)...109

Table 3.13: Results of multiple regression analysis with formant distance and time constant as predictors (/a/ vowel)...110

Table 3.14: Results of multiple regression analysis with formant distance and time constant as predictors (/o/ vowel)..110

Table 3.15: Results of multiple regression analysis with formant distance and time constant as predictors (/u/ vowel)..111

Table 3.16: Results of multiple regression analysis with formant distance, time constant and burst spectrum distance as predictors (vowels pooled)..112

Table 3.17: Results of multiple regression analysis with formant distance, time constant and burst spectrum distance as predictors (/i/ vowel)...113

Table 3.18: Results of multiple regression analysis with formant distance,
time constant and burst spectrum distance as predictors
(/e/ vowel)..113

Table 3.19: Results of multiple regression analysis with formant distance,
time constant and burst spectrum distance as predictors
(/a/ vowel)..114

Table 3.20: Results of multiple regression analysis with formant distance,
time constant and burst spectrum distance as predictors
(/o/ vowel)..115

Table 3.21: Results of multiple regression analysis with formant distance,
time constant and burst spectrum distance as predictors
(/u/ vowel)..115

Table 4.1: Articulatory distances among seven stop articulations......................127

Table 4.2: Results of multiple regression analysis with formant distance,
time constant, burst spectrum distance and articulatory
distance as predictors (vowels pooled)...128

Table 5.1: Example of triangular perceptual distance matrix.............................138

Table 5.2: Example of square perceptual distance matrix.................................138

Table 5.3: Rectangular matrices for the predicted acoustic distance.................140

Table 5.4: Predicted consonant places in five vowel environments....................141

Table 5.5: Preferred consonant places predicted by the rank ordering
of the individual CV salience in Ci..143

Table 5.6: Preferred consonant places predicted by the rank ordering of the individual CV salience in Cɛ..143

Table 5.7: Preferred consonant places predicted by the rank ordering of the individual CV salience in Ca..143

Table 5.8: Preferred consonant places predicted by the rank ordering of the individual CV salience in Cɔ 143

Table 5.9: Preferred consonant places predicted by the rank ordering of the individual CV salience in Cu..144

Table 5.10: Predicted consonant places in five vowel environments where the number of consonant is three (predicted by the method of Liljencrants and Lindblom 1972)..............................145

Table 5.11: Preferred consonant places predicted by the method of Liljencrants and Lindblom (1972) in Ci..146

Table 5.12: Preferred consonant places predicted by the method of Liljencrants and Lindblom (1972) in Cɛ..146

Table 5.13: Preferred consonant places predicted by the method of Liljencrants and Lindblom (1972) in Ca..147

Table 5.14: Preferred consonant places predicted by the method of Liljencrants and Lindblom (1972) in Cɔ..147

Table 5.15: Preferred consonant places predicted by the method of Liljencrants and Lindblom (1972) in Cu..147

List of Figures

Figure 2.1: (a) Formant pattern at vowel mid-point, (b) Formant pattern at CV boundary..13

Figure 2.2: F3 vs. F2 space at stop release...16

Figure 2.3: The waveform and spectrogram of /ba/...19

Figure 2.4: Locus equations for different places of articulation............................21

Figure 2.5: The F2 vs. F3 space for the APEX simulation of stop consonants......27

Figure 2.6: Example of exponential curves that only differ in the value of alpha (time constant, see Equation 2-1)..31

Figure 2.7: Steps involved in fitting an exponential to a formant transition........33

Figure 2.8: The best fitting exponential curve (i.e. simulation of formant transition) for F2 of [ɖu] compared with the curves with different alpha values..38

Figure 2.9: Burst spectra for the dental, alveolar, retroflex, palatal, velar and uvular consonants (in blue) compared with the burst spectrum for the labial consonant.....................................41

Figure 3.1: Hypothetical F2 transitions associated with two CV syllables...........87

Figure 3.2: Plots of averaged dissimilarity scores against formant-based distances compared with upside down exponential curve (6*[$\mathbf{1\text{-}e}^{(\text{-k* acoust dist})}$])..101

Figure 3.3: R^2 values with different predictor variables for the different vowel contexts..116

Figure 4.1: 3-dimensional confusion matrices obtained for identification task..124

Figure 4.2: R^2 values with different predictor variables for the different vowel contexts..130

Figure 5.1: Predicted acoustic distance (derived from Multiple regression analyses using locus distances, time constant differences, burst spectra differences) plotted against the perceptual distance judgements coded as LN(1-perceptual distance/6).............132

Figure 5.2: R^2 values with different predictor variables for the different vowel contexts..133

Figure 5.3: F3 vs F2 space at stop release..142

Chapter 1: Introduction

1.1 Background

This study investigates the problem of defining a measure of perceptual contrast for voiced stops varying along the dimension of place of articulation. There has long been interest in the role of phonetic factors in shaping the sound inventories of the languages. Since obviously different meanings must be conveyed by distinct sound patterns, the role of distinctiveness has long been recognized in phonology (Jakobson, 1941; Martinet, 1955; de Groot, 1931). It was discussed by Moulton (1962). Wang (1968) applied it in an interpretation of formant frequency data.

Liljencrants and Lindblom (1972) was one of the first studies to propose that vowel inventories are shaped by a preference for perceptually maximally distinct vowels. They conducted numerical simulations to derive systems drawn from a space of possible vowels, so as to show maximal perceptual contrasts.

After Liljencrants and Lindblom (1972), the role of speech perception in phonology has been explored by many researchers that include Flemming (2005), Hume and Johnson (2001) and Ten Bosch (1991) among others. The work of Ohala (1981, 1993) presented a perceptually motivated account of sound change that is based on listener errors. A further example is *auditory enhancement theory* (Diehl, Kluender and Walsh, 1990; Diehl and Kingston, 1991) that emphasizes the perceptual role of acoustic redundancy and suggests that perceptual needs determine articulatory patterns.

Compared with the actual vowel inventories of the world's languages, the simulations of Liljencrants and Lindblom (1972) were found to be generally successful except that, for seven or more vowels per inventory (e.g. Italian), they predicted more high vowels than are typically attested. The problem of "too many high vowels" reveals an asymmetry between the contrasts along the open-close (sonority) and the front-back (chromaticity) dimensions. One attempt to solve this problem was the Grenoble's approach (e.g. Schwartz et al., 1997) which introduced the λ coefficient (a number lower than 1) and which was used to weight higher formants so that they made a smaller contribution to the distinctiveness measure than the first formant. However, while the introduction of λ improved the predictions it failed to explain the asymmetry. This weight factor was an ad-hoc stipulation.

To remedy this problem in a principled way, Lindblom (1986) adopted a measure of auditory distance based on whole spectra rather than formant frequencies. Recently Diehl, Lindblom and Creeger (2003) and Lindblom, Diehl and Creeger (2006) took a further step in improving auditory realism by introducing the notion of Dominant Frequency, a measure derived from the zero-crossing frequencies observed at the output of auditory filters.

While the simulations of vowel systems must be said to have been quite successful, they suffer from the limitation of being based on steady state vowels which are rare in natural speech. It would, therefore, be desirable to have a more general measure which could be applied also to time-varying patterns such as diphthongs and CV syllables. Our present understanding of these spectro-temporal aspects of speech is highly incomplete. Nevertheless, in the present work an attempt is made to expand the Liljencrants and Lindblom (1972) to CV syllables.

If it is assumed that distinctiveness is at work also in shaping consonant systems, several puzzles arise. Typological data indicate (Maddieson, 1984; Maddieson and Precoda, 1989) that the preferred places of articulation turn out to include primarily *labial*, *dental/alveolar* and *velar*. Based on the UCLA Phonological Segment Inventory Database (henceforth, UPSID), Maddieson (1984) suggested that the most common place system used 3 places: bilabial, dental/alveolar and velar. UPSID shows that, of 317 languages, 314 languages (99.1%) have bilabial, 316 languages (99.7%) have dental or alveolar and 315 languages (99.4%) have velar stops while only 47 languages (14.8%) have uvular stops and 3 languages have pharyngealized stops. Maddieson (1984) also said that for fricatives, the most common places were labio-dental, dental/alveolar and palatal. Of the 296 languages that have one or more voiceless fricatives, 266 languages (89.0%) have dental/alveolar, 146 languages (49.0%) have palatal and 135 languages (45.0%) have labio-dental fricatives while only 29 languages (9.0%) have uvular and only 13 languages (4%) have pharyngeal fricatives.

It is surprising that segments such as [ɖ], [ɢ] and [ɟ] are not more popular because when they are produced before vowels their auditory salience can be quite striking. Apparently syllables like [ɢi] and [ɟu] are less frequently attested in natural languages than sequences such as palatal [ɟi] and velar [gu]. It is clear that, in articulatory space [ɢi] and [ɟu] travel farther than [gu] and [ɟi] which are "assimilatory". If perceptual salience and contrast are highly valued, why is the former pair rare or absent while the latter pair seems to be the norm?

In fact, Maddieson (1984: 16) made a similar point about vowels:

> The most frequent vowel inventory is /i, e, a, o, u/, not **/i** ẽ a̤ o̰ **u**ˤ**/** where each vowel not only differs in quality but is distinctively plain, nasalized, breathy, laryngealized and pharyngealized. Yet this second set of vowels surely provides for more salient distinctions between them and approaches maximization of contrast more than the first set whose differences are limited to only the primary dimensions conventionally recognized for vowel quality. (Maddieson 1984: 16)

Ohala (1980) further suggests:

[The research of Lindblom and his colleagues] would most satisfying if we could apply the same principles to predict the arrangement of consonants, i.e., posit an acoustic-auditory space and show how the consonants position themselves so as to maximize the inter-consonantal distance. Were we to attempt this, we should undoubtedly reach the patently false prediction that a 7 consonant system should include something like the following set:

ɗ   k’   ts   ɬ   m   r |

Languages which do have few consonants, such as the Polynesian languages, do not have such an exotic consonant inventory. In fact, the languages which do possess the above set (or close to it), such as Zulu, also have a great many other consonants of each type, i.e., ejectives, clicks, affricates, etc. Rather than maximum differentiation of the entities in the consonant space, we seem to find something approximating the principle which would be characterized as "maximum utilization of the available distinctive features". This has the result that many of the consonants are, in fact, perceptually quite close — differing by a minimum, not a maximum number of distinctive features.

Does this mean that consonant inventories are structured according to different principles from those which apply to vowel inventories? Could it mean that the "spaces" both consonants and vowels

> range in, are limited by the auditory features (= parameters) recognized by the particular language? Or does it mean that we are asking our questions about segment inventories in the wrong way? (Ohala 1980: 185)

The above observations lead us to ask whether the distinctiveness alone is sufficient for predicting the CV inventories.

1.2 Goals

This study is an attempt to address the problem of the role of distinctiveness in the patterning of place contrasts in stop consonant inventories. Firstly, I will consider how to define perceptual distinctiveness. My aim is to come up with a minimally ad-hoc definition, that is a definition which is based on phonetic factors alone and independent of the typological phonological patterns to be explained.

In the present thesis I will show that perceptual dissimilarity judgments can indeed be meaningfully quantified using acoustic and auditory properties of the sound stimuli. The work also includes multiple regression analyses set up to correlate data on perceptual confusions and dissimilarity judgments with a composite distance measure which combines information on spectral, dynamic and burst characteristics of the selected CV syllables. These multiple regression analyses provide an indication of the relative contributions of those attributes to the composite measure.

Secondly, in so far I succeed in attaining that goal I will also apply the approach of Liljencrants and Lindblom (1972) to a set of CV syllables in order to shed some preliminary light on the role of perceptual distinctiveness in consonant place systems. This question underlying this second part can be stated as follows:

> *If consonant (CV) systems were seen as adaptations to a demand for perceptual contrast, what would these systems be like?*

The present study is one of very few attempts to seriously address the role of time-variations of speech signals in measures of perceptual contrast and to apply such a measure in a Liljencrants & Lindblom (1972) type of simulation to CV syllables rather than to steady state vowels (For a recent attempt to address the role of speech signal dynamics, see Al-Tamimi (2007)).

Of great relevance to the present effort is the work by Diana Krull (1988, 1990) who used acoustic data to predict perceptual confusions between the Swedish stops [b, d, ɖ, g] in systematically varied vowel contexts. She calculated acoustic distances based on: (1) filter band spectra; (2) F2 and F3 at the CV boundary and in the middle of the following vowel; (3) the duration of the burst (= transient + noise section). The predictions were improved when time-varying properties of the stimuli were included in the distance measures. The highest correlation was obtained with the formant-based model in combination with burst length data. The asymmetries in the listeners' confusions were also shown to be predictable, given acoustic data on the following vowel.

Chapter 2: Experimental Data Collection and Data Analyses

2.1 Materials

The speech materials are 35 CV Syllables (in Table 2.1) recorded by a phonetician who is a native speaker of Icelandic and was trained in the British tradition as a speaker of “universal phonetics.” The 35 CV syllables consist of consonants of seven places of articulation ([b, d̪, d, ɖ, ɟ, ɡ, ɢ]) and five vowels ([i, ɛ, a, ɔ, u]) as shown in Table 2.1.

	bilabial	dental	alveolar	retroflex	palatal	velar	uvular
[i]	bi	d̪i	di	ɖi	ɟi	ɡi	ɢi
[ɛ]	bɛ	dɛ	dɛ	ɖɛ	ɟɛ	ɡɛ	ɢɛ
[a]	ba	d̪a	da	ɖa	ɟa	ɡa	ɢa
[ɔ]	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	ɡɔ	ɢɔ
[u]	bu	d̪u	du	ɖu	ɟu	ɡu	ɢu

Table 2.1 35 CV syllables (consonants of 7 places of articulation x 5 vowels)

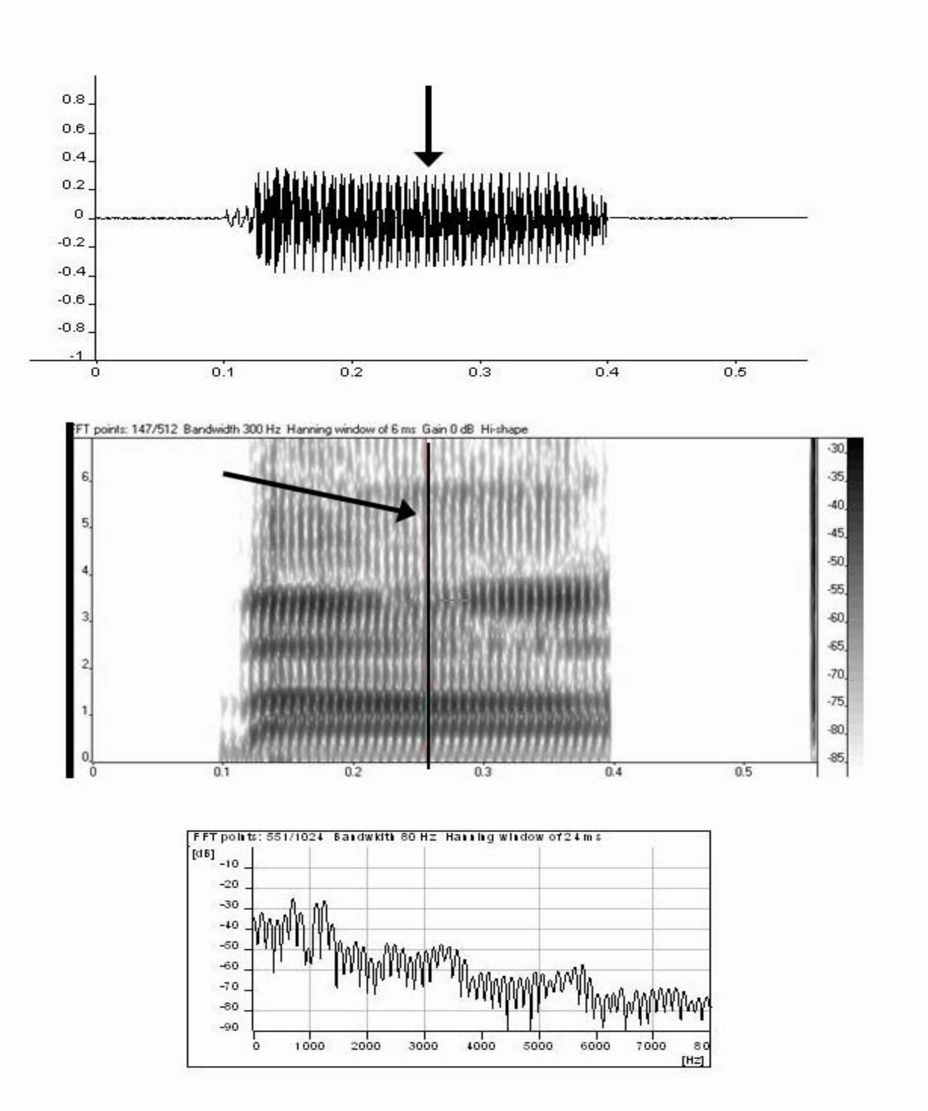
2.2 Acoustic analyses

The question may arise about how valid these utterances are, in other words, whether the speaker succeeded in varying place as instructed. To answer this question, formant measurements were made and compared with (i) published data and (ii) the output of an articulatory model APEX (Lindblom & Sundberg, 1971; Branderud et al., 1998; Stark et al., 1999; Ericsdotter, 2005).

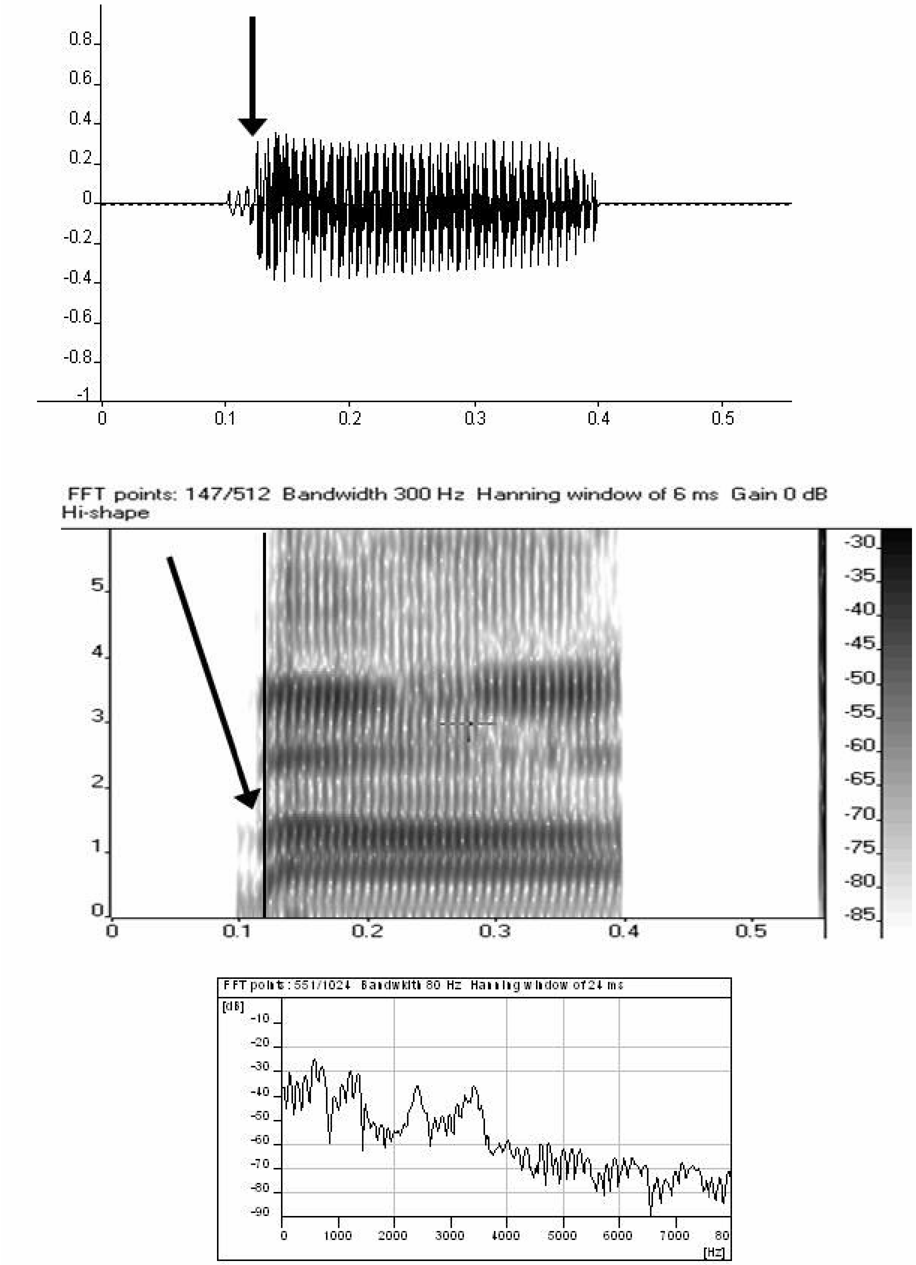
2.2.1 Formant patterns

2.2.1.1 Measurements

The frequencies of formants 1-4 were obtained at the onset and mid-point of each vowel in each CV syllable using Soundswell (Version 4.0). The measurements were made by hand using wide-band (300 Hz) spectrograms and consulting FFT displays (discrete power spectrum of a frame of sampled data). For the FFT display 25ms window (bandwidth of 80 Hz) was used to clearly identify each harmonic. The onset of vowel was defined as a point immediately after the burst and vowel mid-point was defined as a point where there were no formant transitions in the spectrogram. The measurement points for vowel-onset and mid- vowel points are illustrated in Figure 2.1.



(a)



(b)

Figure 2.1 (a) Formant pattern at vowel mid-point

(b) Formant pattern at CV boundary

The vertical lines indicate the point at which the formant measurements were made.

Formant measurements for F1, F2, F3 and F4 are shown in Table 2.2.

	Vowel Onset				Mid-vowel			
	F1	F2	F3	F4	F1	F2	F3	F4
bi	215	2215	2445	3452	301	2344	3282	3729
d̪i	215	1913	2493	3373	387	2278	3160	3611
di	279	1910	2407	3233	408	2210	3181	3463
ɖi	258	1892	2515	3412	365	2193	3074	3511
ɟi	343	2171	3267	3511	408	2278	3223	3492
gi	301	1481	2064	3392	430	2296	2816	3577
ɢi	322	1139	2472	3412	515	2214	2967	3531
bɛ	387	1548	2493	3509	688	2021	2708	3591
dɛ	430	1652	2429	3611	645	1999	2730	3599
dɛ	408	1655	2511	3531	645	2021	2751	3541
ɖɛ	430	1720	2407	3490	602	2042	2730	3660
ɟɛ	236	2278	3203	3516	602	2106	2730	3569
gɛ	322	1397	2386	3277	645	2035	2751	3607
ɢɛ	387	1225	2321	3392	731	1956	2558	3553
ba	387	1182	2407	3340	817	1247	2538	3450
d̪a	451	1548	2493	3670	860	1311	2429	3539
da	430	1677	2644	3511	838	1311	2429	3481
ɖa	387	1634	2106	3476	851	1290	2364	3482
ɟa	322	2149	3095	3469	795	1311	2386	3437
ga	279	1526	2128	3362	860	1311	2300	3318
ɢa	387	1502	2171	3343	946	1373	2343	3473
bɔ	322	860	2493	3073	559	989	2601	3015
d̪ɔ	430	1290	2450	3596	494	946	2429	3407
dɔ	365	1416	2601	3447	559	944	2408	3370
ɖɔ	344	1440	1548	3358	494	752	2450	3184

ɟɔ	301	1889	2128	3348	494	880	2472	3273
gɔ	344	1009	2193	3206	537	858	2429	3134
ɢɔ	387	944	2257	3222	580	987	2558	3254
bu	301	731	2257	3118	387	645	2364	3035
d̪u	365	1268	2278	3516	451	752	2386	3401
du	300	1354	2193	3328	387	688	2322	3279
ɖu	365	1245	1612	2321	472	751	2150	2976
ɟu	279	2193	2279	3410	365	709	2322	3313
gu	301	858	2149	3154	430	730	2257	3093
ɢu	279	731	2253	3277	387	709	2382	3088

Table 2.2 Formants 1-4 measured at CV boundary and vowel mid-point

2.2.1.2 F3 vs. F2 space at CV boundary

Fant (1973) examined the discriminative power of the second and third formant frequencies of Swedish voiced and voiceless stops with three places of articulation (labial, dental, and velar) followed by nine vowels by plotting F2 and F3 against each other. He found that voiceless consonants were better differentiated by F2 and F3 than voiced stops. In the case of voiced stops, F2-F3 points varied considerably with the following vowels. Krull (1988) plotted F2 and F3 at the CV boundary with four places of articulation (labial, dental, retroflex and velar). The retroflex consonants showed considerable overlap with labials and dentals.

A similar F2-F3 plot (Figure 2.2) was obtained from my measurements with the stops of seven places of articulation (labial, dental, alveolar, retroflex, palatal, velar, and uvular) followed by five vowels ([i], [ɛ], [a], [ɔ], and [u]).

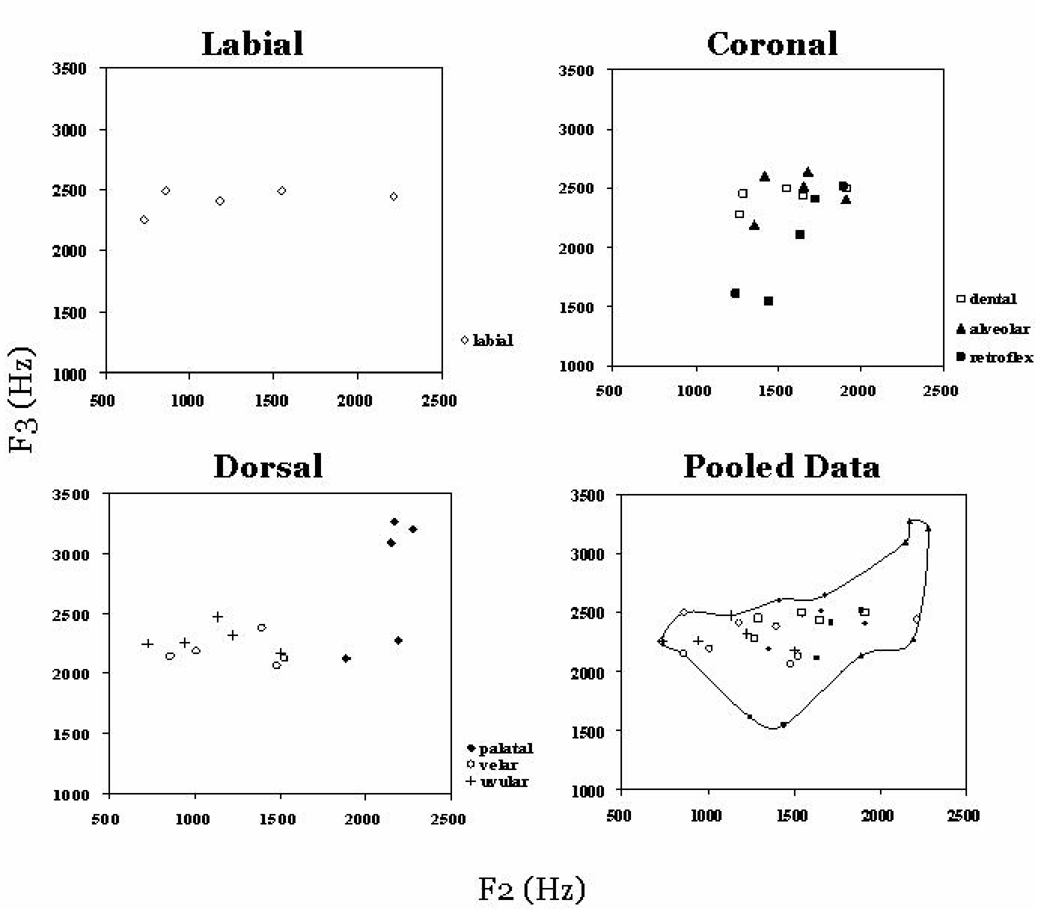


Figure 2.2 F3 vs. F2 space at stop release.

In the diagram with pooled data the outermost points were connected with a smoothed curve so as to enclose all measurements.

It can be seen that labials spread broadly over the F2 range while F3 shows little variation. Coronals (dental, alveolar, and retroflex) span a fairly wide F3 range (1500-2700 Hz) but show a relatively small F2 range (1200-2000 Hz). Low F3 values are associated with [ɖ] in the back vowel contexts. Dorsals except palatals, that is velars and uvulars, are also widely spread in F2 and show considerable overlap with labials. Palatals form the most distinctive group with higher F2's. These F2-F3 measurements will be examined in detail in 2.2.1.4.

2.2.1.3 Locus equations

When the F2 onset of the transition from a given consonant is plotted as a function of the F2 at the mid-point of the following vowel, a linear and tight cluster of data points is obtained under a wide range of conditions. The straight lines that describe such data are known as locus equations (henceforth, LEs). They are of the form $F2(onset) = k*F2(vowel) + c$ (where constants k and c represent the slope of the regression line and its intercept respectively). These slopes have been found to vary systematically with the place of the consonant (Sussman, McCaffrey & Matthews, 1991; Sussman, Fruchter, Hilbert and Sirosh, 1998). They also provide an indication of the degree of coarticulation (Krull 1988: 66-71). The constant *k*, which is the slope of the LE, can vary between 0 and 1. When *k*=1, the formant frequency at the locus is the same as that of the target indicating a maximal coarticulation effect, while *k*=0 means that the formant frequency at the locus stays the same regardless of the following vowel thus indicating a no-coarticulation effect.

This metric was applied to the present place data to examine how slopes and intercepts vary with place of articulation. An example is shown in Figure 2.3.

The LE phenomenon can be interpreted in terms of simplified rules of thumb of acoustic theory that relate cavity shapes to formant frequencies (Fant, 1960). The second formant frequency reflects the front-back position of the tongue body in a fairly straightforward way. Since consonants are normally coarticulated with, in other words anticipate, the following vowel, F2onset (F2 measured at the CV boundary) will to some extent reflect the F2 value at the following vowel mid-point (F2vowel). It is this coarticulatory organization that shows up as linear patterns in F2onset vs. F2vowel plots and that LEs capture elegantly in terms of only 2 numbers: slope and intercept.

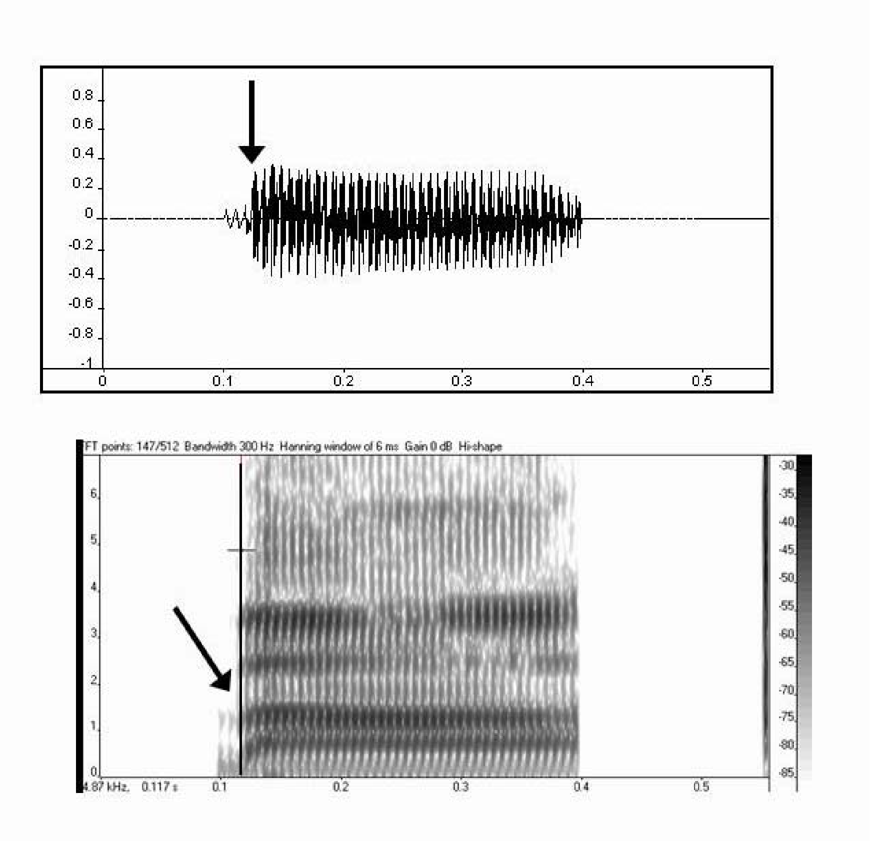


Figure 2.3 The waveform and spectrogram of /ba/

The arrow head indicates the point where the F2onset and F3onset ("locus" values) were measured.

Vocal tract cavities are not acoustically autonomous. There is interaction. This becomes evident when F3onset is plotted against F2vowel as illustrated in Figure 2.4. As shown by the top cluster of points, a reasonably linear pattern is obtained for F3onset vs. F2vowel indicating that also F3onset is influenced by the front-back dimension although to a lesser degree than F2.

The LEs for F2onset vs. F2vowel and F3onset vs. F2vowel with R^2 values for the seven places of articulation are shown in Table 2.3. The labial LE is steeper (slope=0.81) with lower intercept (132 Hz) than dental LE (slope=0.39, intercept=962 Hz) or alveolar LE (slope=0.31, intercept=1164 Hz). This observation is compatible with those of Sussman et al. (1991) where for a male speaker the slope of LE for labial is 0.813 with an intercept of 231 Hz while the slope for alveolar is 0.394 with an intercept of 1217 Hz. The dorsal consonants are better described by using two LEs distinguishing the front vowel and back vowel contexts (slope before front vowel: 0.32, intercept=746 Hz; slope before back vowel: 1.15, intercept=23Hz) as in Sussman et al. (1991), where the slope for velars before back vowels is 0.963 with an intercept of 487 Hz while the slope for front vowels context is 0.222 with an intercept of 2179 Hz (the number is based on one female speaker). The present data for the uvular place also motivate separate LEs for the front and back vowel contexts. The slope is steeper before back vowels (slope=1.18) and flatter (slope=-0.13) before front vowels.

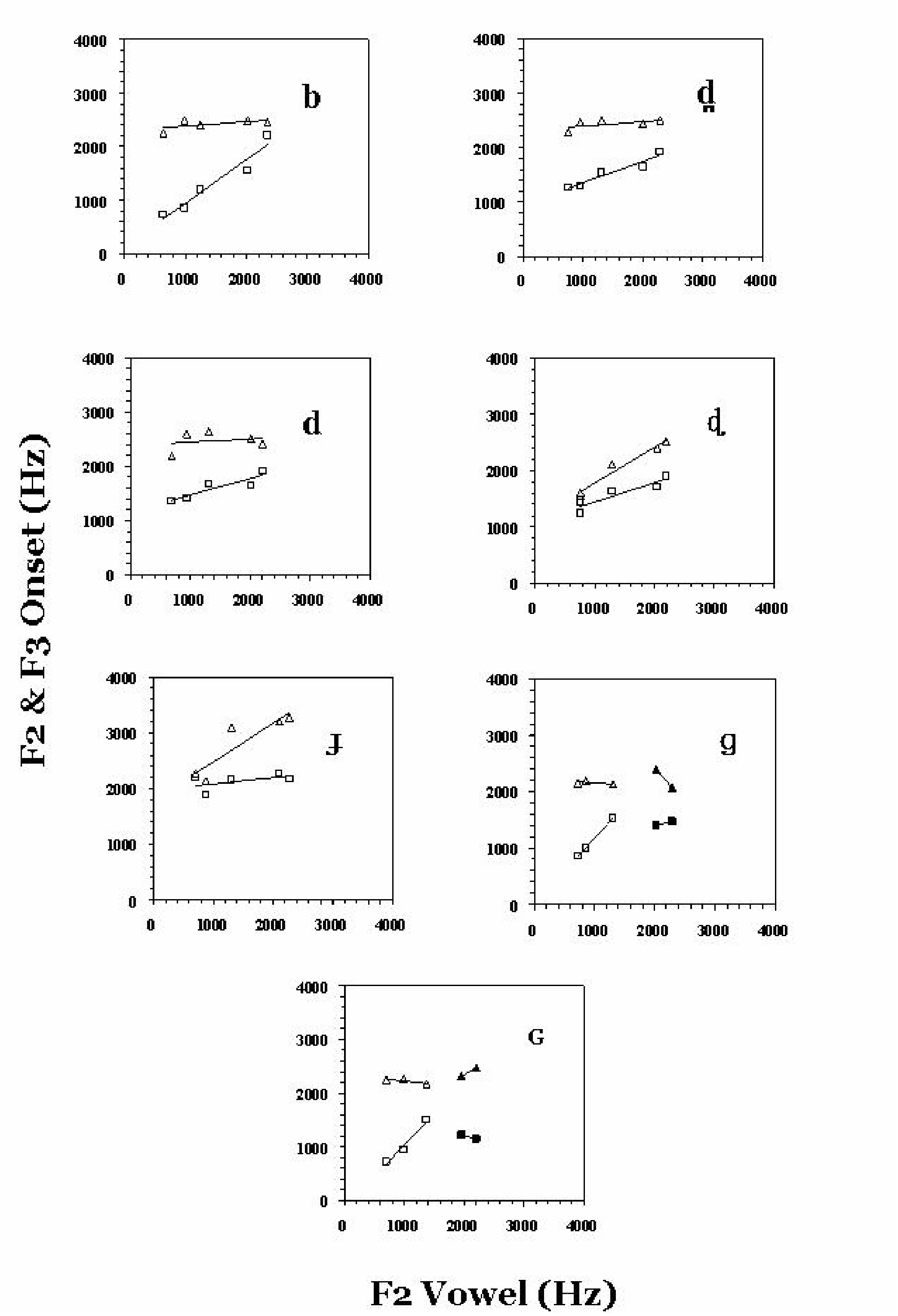


Figure 2.4 Locus equations for different places of articulation

R^2 values are generally high for F2onset-F2vowel (labial: 0.93, dental: 0.93, alveolar: 0.82, retroflex: 0.86, velar before back vowel: 1.0, uvular before back vowel: 0.98). The flat LE of the palatal stop is low (0.28). R^2 values for velar and uvular stops before front vowels were not considered because with only two measurement points a R^2 score of 1 would be obtained. The LE slope, y-intercept and R^2 values obtained for each place of articulation are shown in Table 2.3.

	F2		F3	
Bilabial	y = 0.81x + 132	$R^2 = 0.93$	y = 0.08x + 2302	$R^2 = 0.35$
Dental	y = 0.39x + 962	$R^2 = 0.93$	y = 0.08x + 2314	$R^2 = 0.35$
Alveolar	y = 0.31x + 1164	$R^2 = 0.82$	y = 0.06x + 2392	$R^2 = 0.04$
Retroflex	y = 0.34x + 1112	$R^2 = 0.86$	y = 0.63x + 1146	$R^2 = 0.96$
Palatal	y = 0.11x + 1977	$R^2 = 0.28$	y = 0.69x + 1794	$R^2 = 0.80$
Velar (back)	y = 1.15x + 23	$R^2 = 1.00$	y = -0.06x + 2219	$R^2 = 0.36$
Velar (front)	y = 0.3201x + 746	$R^2 = 1.00$	y = -1.23x + 4897	$R^2 = 1.00$
Uvular (back)	y = 1.18x - 148	$R^2 = 0.98$	y = -0.13x + 2361	$R^2 = 0.80$
Uvular (front)	y = -0.13x + 2361	$R^2 = 1.00$	y = 0.58x + 1179	$R^2 = 1.00$

Table 2.3 Locus equations and R^2 for the seven places of articulation

It is also worth noting that since most F3onsets show a limited range, R^2 values for F3onset-F2vowel are relatively low (labial: 0.35, dental: 0.35, alveolar: 0.04, velar before back vowels: 0.36). In contrast those for retroflex (0.96) and palatal (0.80) show larger F3 spans and thus have significantly higher R^2's.

In conclusion it can be safely claimed that the materials that I used in this research are compatible with previous locus equation findings.

2.2.1.4 Comparison with previous theoretical predictions

Klatt & Stevens (1969) and Stevens (1998) conducted simulations to predict F1-F4 patterns based on simplified tube models using parameters such as length of constriction and the cross-sectional area of the constriction.

My own approach to derive predictions of formant onset values was based on the APEX articulatory model. APEX (Lindblom & Sundberg, 1971; Branderud et al., 1998; Stark et al., 1999; Ericsdotter, 2005) is a model that derives the frequencies of the first four formants from input specifications of the shape and position parameters for lips, tongue body, tongue blade elevation and protrusion, jaw aperture and larynx height.

APEX was used in two separate simulations to derive F-patterns of (i) dorsal consonants and (ii) for coronal articulations.

Using APEX Engstrand, Frid, and Lindblom (2007) examined the acoustic properties of various points of articulation in dorsal and coronal rhotics (i.e. various /r/ sounds). Their aim was to look for perceptual overlap and similarity between dorsals and coronals that might make these categories ambiguous and lead to a reinterpretation of their places of articulation. For the coronals, they used articulatory data consisting of 400 tongue shapes obtained from an X-ray film of a Swedish speaker (Lindblom, 2003). These tongue-body shapes were specified numerically by parameters derived from a Principal Components analysis. To examine the acoustic consequences of changes in place of tongue blade articulation they selected a retroflex [r] occurring after an [ɑː] vowel as a reference point. The tongue blade portion of this configuration was systematically varied according to geometric rules along a continuum simulating five different places of articulation ranging from dental to extreme retroflex. The APEX model was used to derive formant patterns from the lateral profiles of these articulations (Figure 12.3. in Engstrand et al., 2007). Three of the points analyzed in the Engstrand study are relevant to my own project (i.e. data points for dental, alveolar, and retroflex). Their F2 and F3 frequencies are plotted on a F2-F3 plane (Figure 2.5).

For dorsals, Engstrand et al. (2007) simulated the formant patterns of eleven data points ranging from palatal to pharyngeal. A constant value of 0.25 cm^2 was used for the constriction area at these places. All articulatory specifications for the APEX simulations of these articulations were made available to me by the authors so as to allow me to replicate and extend their findings. I selected nine configurations that I judged to be representative of a palatal to uvular series. Using APEX I generated formant patterns and plotted F2 and F3 values on the above mentioned F2-F3 chart (Figure 2.5).

One problem in using the Engstrand et al. (2007) data is that neutral lip conditions were assumed throughout the continuum. In my own simulations I decided to also include a palatal stop produced with spread lips. Accordingly there were ten data points for bilabials and dorsals. The palatal with spread lips was identical to the neutral palatal except for its larger lip opening area. The parameters used in the APEX simulation are listed in Tables 2.4 and 2.5.

elevation	0
protrusion	0
jaw	7
larynx	85
displacement	1

Table 2.4 Parameters fixed for the APEX simulation of dorsal and bilabial stops

The specifications are given in mm except for displacement and position that are normalized and dimensionless dimensions.

position	-1	-1	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6
lip	S	N	N	N	N	N	N	N	N	N

Table 2.5 Variable parameters for dorsal and bilabial stops

S: Spread lip, N: Neutral lip

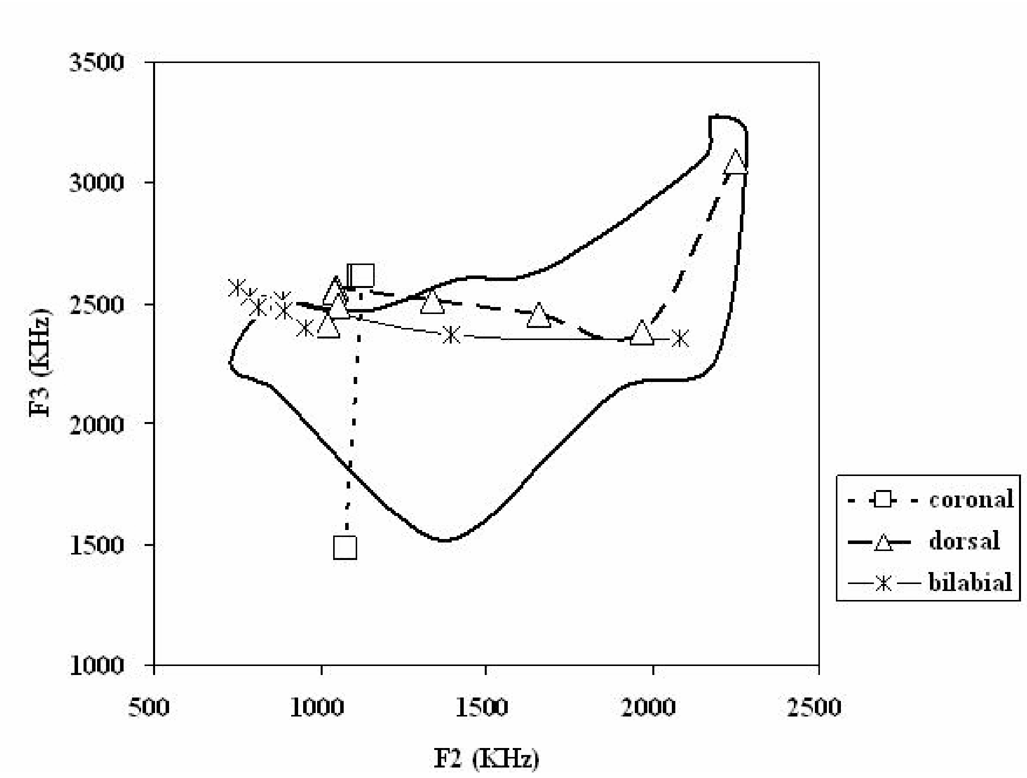


Figure 2.5 The F2 vs. F3 space for the APEX simulation of stop consonants

For comparison the F3-F2 space presented with the pooled data in Figure 2.2 is also shown.

The area enclosed by the thick solid curve represents the F2-F3 space of my formant measurements as previously presented in Figure 2.2. The symbols indicate the F3-F2 values obtained in the APEX simulations.

In comparing the APEX results with the spectrographic measurements we should bear in mind that the former were derived from X-ray data on an [r] observed in a single context [ɑː]. The measured data come from syllables with five different vowels and should therefore be expected to show a greater range due to coarticulation, especially in F2.

Open squares show the predicted coronals. The dental and alveolar stops are located at the top of the left half. The retroflex articulation is found at the bottom.

Each triangle indicates F2 and F3 at a point from (spread) palatal at the rightmost position to uvular at the leftmost end. Since bilabials were not considered by Engstrand et al. (2007), I simulated bilabials using the same parameters as I used for the dorsals except for the lip opening which was set at 0.16 cm^2 (= nearly complete closure).

The APEX simulation of different places of articulation shows reasonable compatibility with my measurements. Not only do the predicted F2's and F3's fall not far from the measured F2-F3 space, they also indicate a similar pattern.

The downward drop in the measured F2-F3 contour is due to the low F2 and F3 for the retroflex before back vowels. This effect is correctly predicted in the APEX simulation. Good agreement is also obtained for the uvulars. In my measurements, uvulars before front vowels and /a/ have significantly high F3 and F2 which is well predicted by APEX simulation.

My previous observation that coronals (dental, alveolar, and retroflex) occupy a wide F3 range but a relatively small F2 range is confirmed by the vertical alignment of the coronal consonants (dotted line). Labials (solid line) and dorsals (broken line) form horizontal configurations (i.e. large F2 variation but small F3 variation). This result agrees nicely with the patterning of labial and dorsal measurements in Figure 2.2.

In summary we note that there is satisfactory qualitative agreement among APEX simulation results, the present CV measurements as well as published information. This finding suggests that the speech materials used in this study are sufficiently representative of variations in place of articulation and that they can be meaningfully explored with respect to perceptual properties.

2.2.2 Time constants of formant transitions

In Lindblom (1963 a,b) formant undershoot (displacement of formant from target value) was shown to depend on vowel duration and consonant context and was described in terms of an exponential function with good accuracy. The success of this method is probably related to the fact that the time variations of CV and VC formant transitions often closely resemble decaying exponentials. A CV transition could thus be expected to be fairly well represented by the following equation.

$$Fn(t) = (Fn(onset)-Fn(V))*e^{-\alpha t}+Fn(V) \qquad \text{(Equation 2-1)}$$

where Fn(t) stands for *n*th formant at time point (t), Fn(onset) is the *n*th formant at CV boundary and Fn(V) is the *n*th formant at steady state vowel.

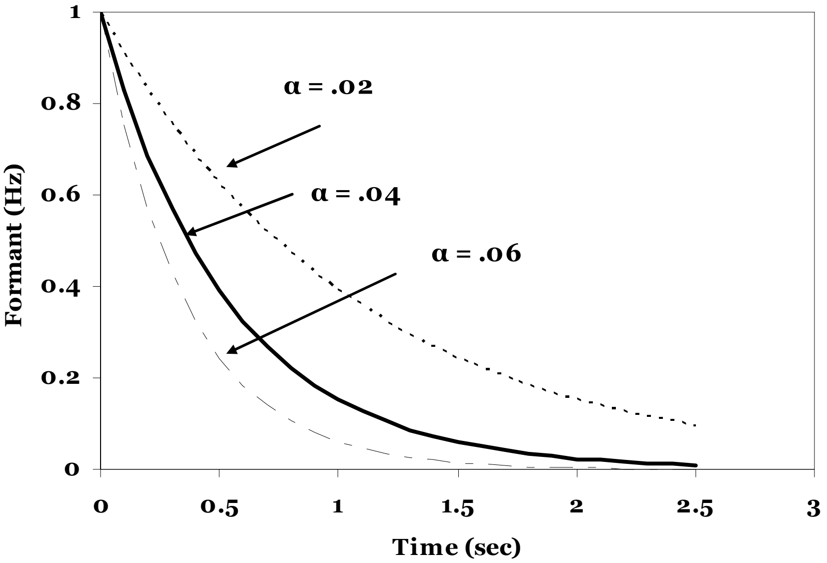


Figure 2.6 Example of exponential curves that only differ in the value of alpha (time constant, see (Equation 2-1))

Figure 2.6 illustrates the effect of varying the time constant α of (Equation 2-1). This number determines how fast the exponential curve approaches to the asymptote, that is, in this case the zero line (abscissa). Applied to formant transitions, it determines how fast the formant curve reaches the formant target (i.e. steady state). In other words, given the locus and target value of the formant, we can describe the formant transition [Fn(t)] as a decaying exponential that starts at the locus and whose asymptote is Fn(V). (Equation 2-1) can be rewritten as follows, where the rate of formant frequency change is captured by α.

$$[Fn(t)-Fn(V)]/[Fn(onset)-Fn(V)] = e^{-\alpha t} \qquad \text{(Equation 2-2)}$$

By taking the natural logarithm of both sides of the equation, we obtain:

$$LN([Fn(t)-Fn(V)]/[Fn(onset)-Fn(V)]) = -\alpha t \qquad \text{(Equation 2-3)}$$

This result suggests that by plotting the value of the left-hand side of (Equation 2-3) against time (t), we can find α by a linear regression analysis.

The steps of calculating the alpha for a given CV are thus as follows (Figure 2.7). First, you measure formants at CV boundary and several consecutive points along the time course of the transition until you reach to the point where formants do not change any more. The diagram in the right-hand side of the first row shows the result of this first step applied to F2. Then, you subtract F2(V) from F2 at every time point. This will give you a cluster of data points approaching zero as a function of time. Next, divide F2(t)-F2(V) by F2(onset)-F2(Vt). This is a normalization procedure restricting numbers to a range between

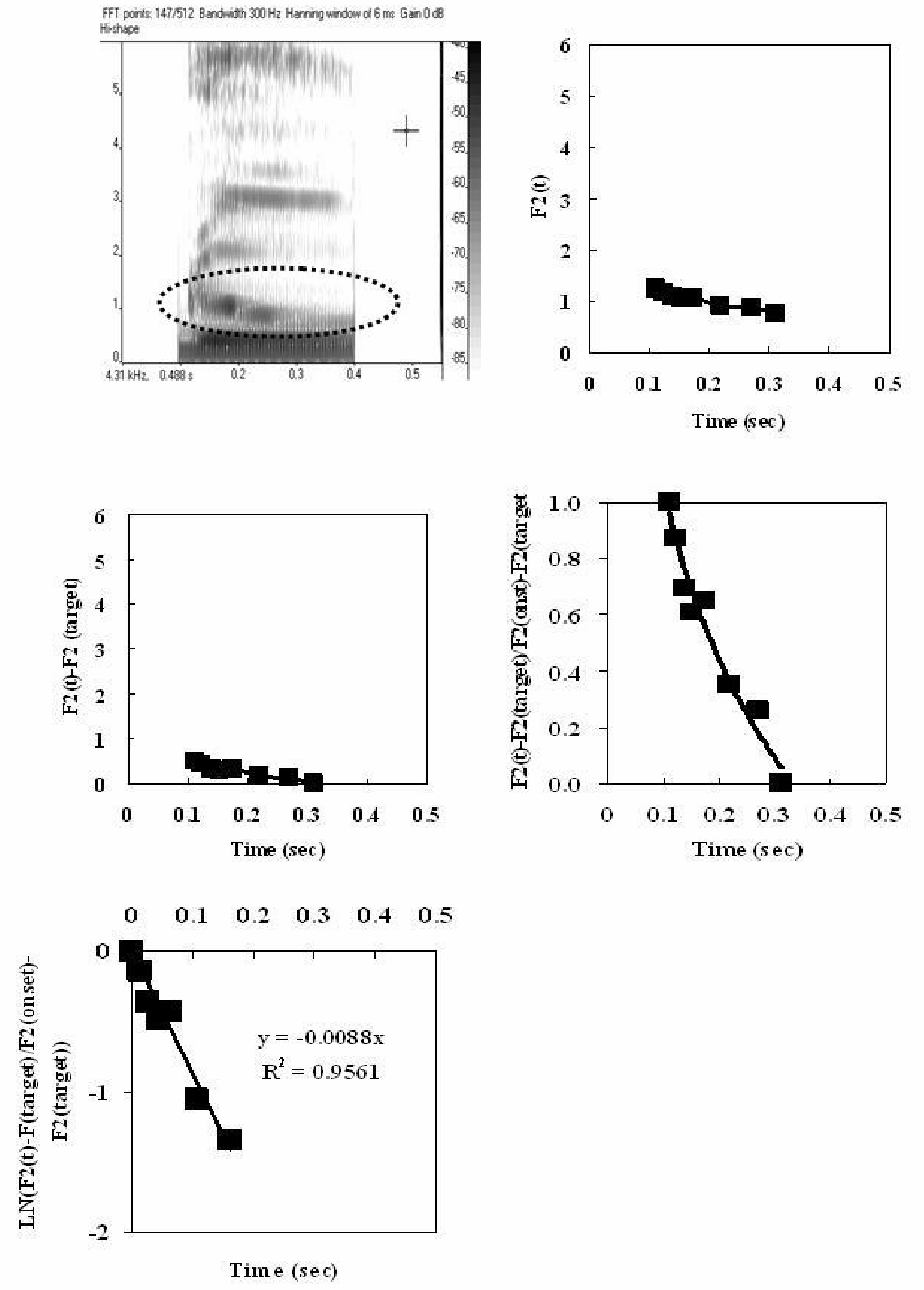


Figure 2.7 Steps involved in fitting an exponential to a formant transition

zero to one. At the CV boundary you will get 1 where F1(t) = F1(onset) and at the vowel target point you will get zero because F1(t) is F1(V) and thus F1(V)-F1(V) (=numerator of [Fn(t)-Fn(V)]/[Fn(onset)-Fn(V)]) is zero. Finally, by plotting natural logarithm of [Fn(t)-Fn(V)]/[Fn(onset)-Fn(V)] (diagram in the third row) against time and fitting a straight line the α value is given by the slope of the line. The CV boundary time point should be set at zero at this step. The R^2 value for the curve fitting indicates how well the straight line fits the observed data.

To obtain alpha values, I measured F1, F2 and F3 from the locus (1st glottal pulse after the release) and at several consecutive time points along the transition.

It is very difficult to avoid measurement errors when the difference between locus and target vowel is less than 100Hz. Therefore, to improve the accuracy of the fitting, I omitted the alpha value for the tokens with locus-target difference of 100Hz and lower. The alpha values are shown in Table 2.6.

		F1	Locus-Target	R²	F2	Locus-Target	R²	F3	Locus-Target	R²
/i/	bilabial	0.2368	86		0.0412	130	0.78	0.0492	837	0.91
	dental	0.2659	172	1	0.0544	365	0.94	0.0402	666	0.95
	alveolar	0.3352	129	1	0.0661	300	0.96	0.0360	774	0.96
	retroflex	0.0830	107	0.97	0.0635	301	1	0.0311	559	0.90
	palatal	0.0453	64		0.0129	107	0.93	0.0162	44	
	velar	0.0249	129	1	0.0112	816	0.86	0.0105	752	0.94
	uvular	0.0965	193	1	0.0158	1075	0.88	0.0053	494	0.49
/e/	bilabial	0.1349	301	1	0.0488	473	0.97	0.0317	215	0.99
	dental	0.0440	215	0.90	0.0261	347	0.95	0.0162	301	0.92
	alveolar	0.0872	236	1	0.0157	365	0.74	0.0420	240	0.09
	retroflex	0.0836	172	0.73	0.0158	322	0.93	0.0106	322	0.91
	palatal	0.0247	365	0.90	0.0467	172	0.96	0.0481	473	0.90
	velar	0.0260	322	0.71	0.0145	638	0.94	0.0106	365	0.91
	uvular	0.0937	344	0.98	0.0191	731	0.96	0.0135	236	0.95
/a/	bilabial	0.0928	430	0.97	0.1057	64		0.0094	130	0.95
	dental	0.0247	408	0.84	0.0233	236	1	0.0243	64	
	alveolar	0.0294	408	0.78	0.0296	365	0.87	0.0327	215	0.99
	retroflex	0.0512	464	0.92	0.0204	344	1.00	0.0296	258	0.94
	palatal	0.0198	473	0.95	0.0232	838	0.90	0.0238	709	0.99
	velar	0.0341	580	0.97	0.0225	215	0.82	0.0534	172	1
	uvular	0.0348	559	0.97	0.0809	129	0.67	0.0162	172	0.63
/o/	bilabial	0.0550	236	0.53	0.0202	129	0.91	0.0206	107	0.60
	dental	0.0229	64	0.01	0.0339	344	0.98	0.0000	21	
	alveolar	0.0166	193		0.0294	472	0.83	0.0284	193	0.60
	retroflex	0.0591	150	0.66	0.0205	688	0.92	0.0263	903	0.99
	palatal	0.0505	193	0.98	0.0279	1009	0.97	0.0177	344	0.88
	velar	0.0344	193	0.24	0.0327	150	0.70	0.0418	236	0.95
	uvular	0.0440	193	0.03	0.0108	43		0.0279	301	0.86
/u/	bilabial	0.0906	86		0.0126	86		0.0816	107	1
	dental	0.0768	86		0.0183	516	0.98	0.0201	107	0.48
	alveolar	0.0719	86		0.0310	666	0.95	0.0072	129	0.61

retroflex	0.0718	107	0.98	0.0088	494	0.96	0.0337	537	0.98
palatal	0.0199	86		0.0139	1483	0.93	0.0058	43	
velar	0.0237	129	1	0.0319	129	0.43	0.0849	108	0.72
uvular	0.0165	107	0.57	0.0000	21		0.0315	129	0.97

Table 2.6 Formant time constants (F1, F2 and F3) for all CVs

Alpha values for the blank slots were removed because the Locus-target distance is less than 100 Hz.

Table 2.6 indicates that α values vary considerably. An analysis of variance was performed to investigate if their pattern is lawfully related to place, vowel context or formant numbers. This analysis failed to reveal any significant systematic effects (See Appendix B.1 and B.2 for the ANOVA results).

Stevens (1998) calculated the formant transition after the stop release based on his tube model suggesting that F1 transitions for dorsal stops are slower than for other places. In my alpha measurement, the alpha value (F1) for the velar in the /u/ vowel context (palatal: 0.0199, velar: 0.0237, uvular: 0.0165) is significantly lower than the alpha values (F1) for the other cases (labial: 0.0906, dental: 0.0768, alveolar: 0.0719, retroflex: 0.0718). The same pattern is observed in the /i/ context where alphas for the dorsals (palatal: 0.0453, velar: 0.0249, uvular: 0.0965) are lower than alphas for other consonants (labial: 0.2368, dental: 0.2659, alveolar: 0.3352, retroflex: 0.0830) confirming Stevens (1998). However, alpha values (F1) in other vowel environments do not show any clear pattern.

Despite the fact that no orderly pattern could be established and linked to the phonetic dimension of place, vowel or formant number, the observed α variations cannot be dismissed as arising from measurement error.

This is evident from the fact that high R^2 values were obtained in the majority of cases. It can also be demonstrated by considerations such as the following.

The α of F2 in [ɖu] was found to be 0.0088. Its locus-target value was 494. What does this value "mean" in terms of formant transition rate? Suppose this F2 transition was described using slightly different values. What would be the effect of the goodness of fit of the curve fitting?

Figure 2.8 shows that the magnitude of an error (within the range of values observed in Table 2.6) can be considerable when α is varied.

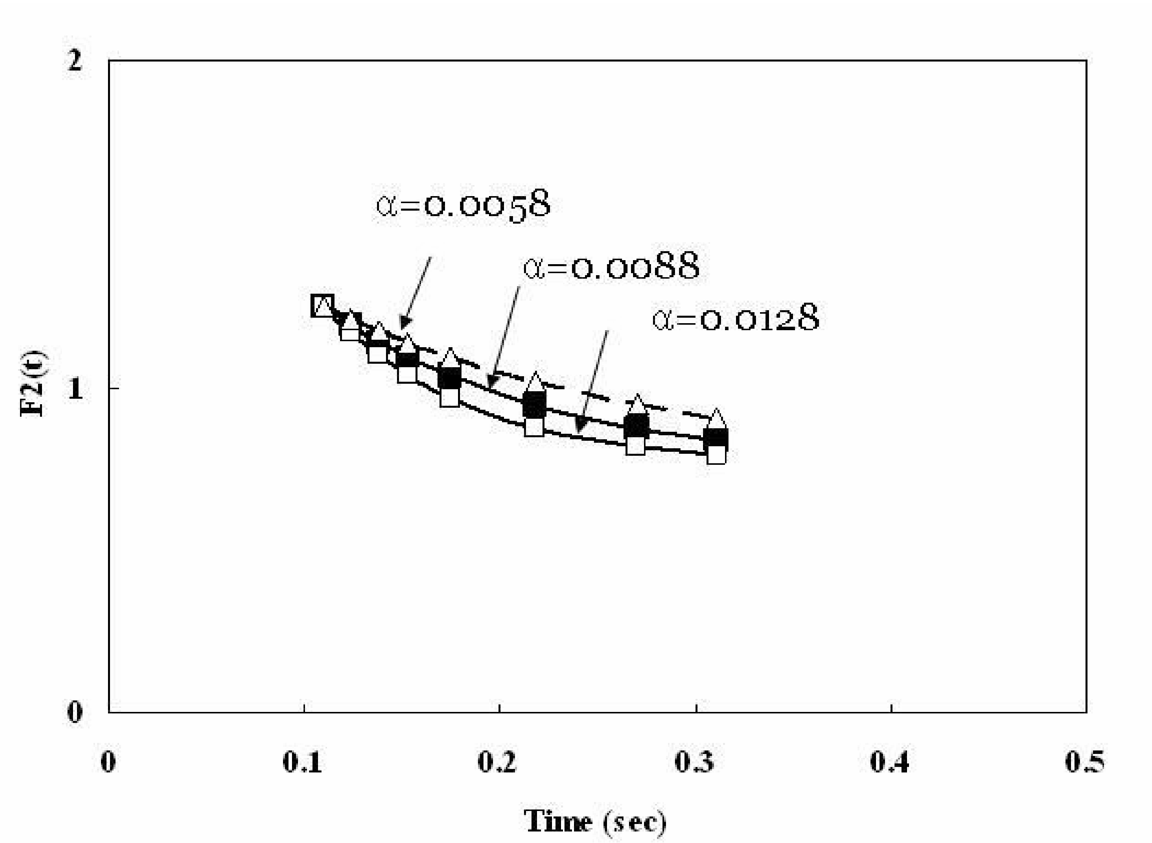


Figure 2.8 The best fitting exponential curve (i.e. simulation of formant transition) for F2 of [ɖu] compared with the curves with different alpha values

I take this observation to mean that the high R^2 values of α in Table 2.6 do in fact do a good job describing formant dynamics. Moreover, the lack of a pattern correlating strongly with phonetic categories raises intriguing questions about the role of rate of formant change as a perceptual cue. Although interesting, this topic will not be further considered here.

2.2.3 Burst spectra

As a general rule of thumb, the main spectral peak of a stop burst is determined by the length of the cavity in front of the source at the place of articulation (Stevens, 1998). This rule makes us expect to find systematic spectral effects in the present set of speech samples including seven places of articulation at five vowel environments.

The burst spectra were obtained using Soundswell 4.00 (Hitech Development AB 2000) with bandwidth of 300 Hz and Hanning window of 6 msec. To find the time point for the burst measurement, I obtained the FFT spectra at several consecutive time points in the spectrogram from well before the release of the stop occlusion. The time point for the burst spectra was defined as the time point where the FFT spectrum showed a significant amplitude change from the previous spectrum.

The spectra obtained showed the effect of vowel coarticulation probably because the tongue is more or less in position for the vowel target at the burst. This made it very hard to find the spectral pattern characteristic of the place of articulation. To minimize this vowel coarticulation effect, the burst spectra for the each place of articulation were averaged over the vowel contexts. For example, one spectrum of the alveolar consonant was obtained by averaging the spectra for [di], [dɛ], [da], [dɔ] and [du].

For the labial consonant, the frication source at the lips tends to excite a very short cavity producing a more or less flat or falling spectrum in the frequency range of interest here (Stevens and Blumstein, 1981). The labial spectrum in Figure 2.9 shows this falling pattern without discernable bumps because the vowel effect is reduced by averaging over the vowel context. The labial spectrum was used as s reference point for comparison with the spectral patterns for the other places of articulation (Figure 2.9).

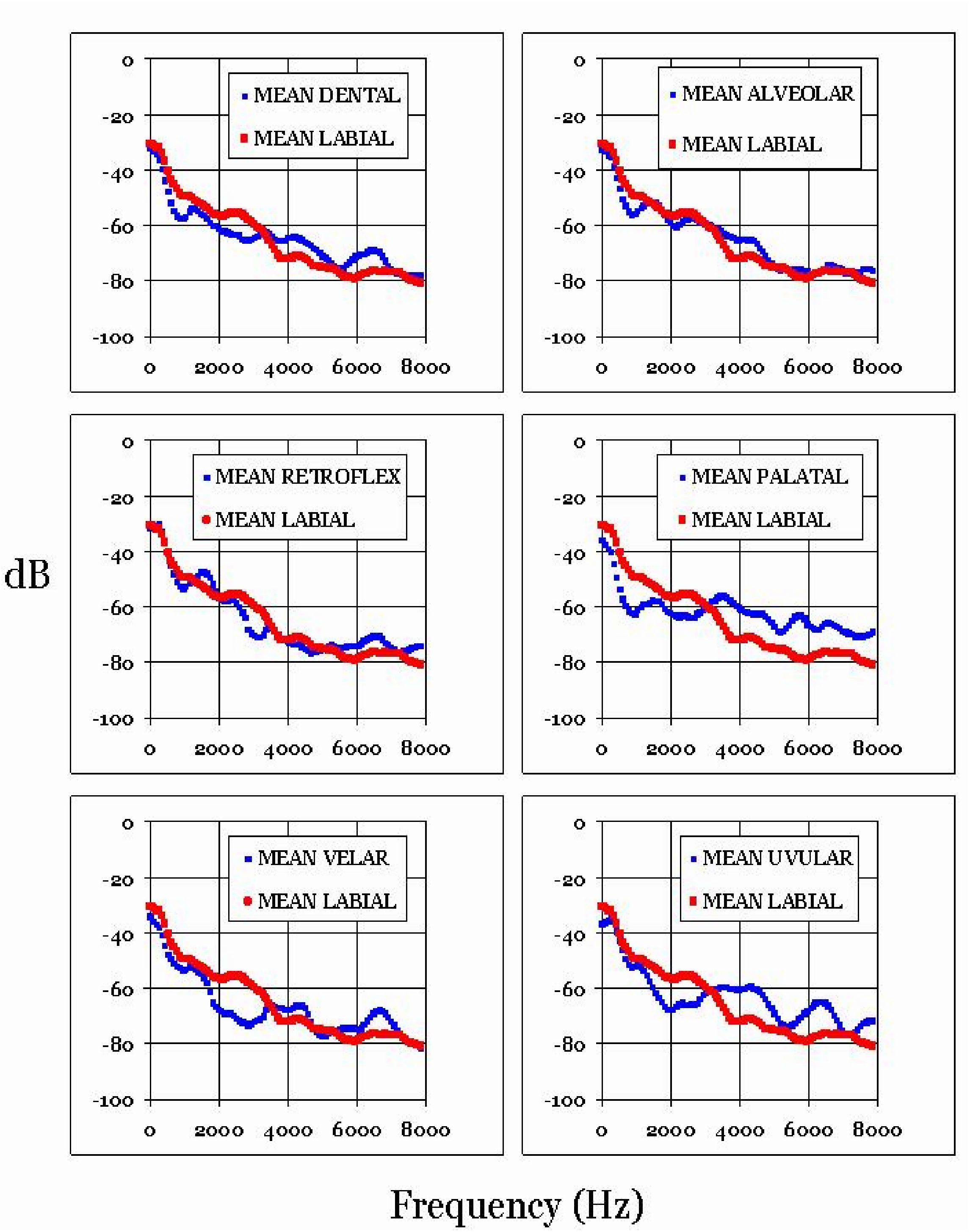


Figure 2.9 Burst spectra for the dental, alveolar, retroflex, palatal, velar and uvular consonants (in blue) compared with the burst spectrum for the labial consonant

For a coronal consonant, the frication source tends to excite higher formants producing a hump at ca. 4000 Hz for the dental consonant, at ca. 4500 Hz for the alveolar consonant and at ca. 1800 Hz for the retroflex consonant.

The palatal spectrum showed a major hump at ca. 3500 Hz. This makes sense because the shape of the oral cavity is very similar to the articulatory shape of an /i/ pronounced with strong excitation of F3 and F4 regions.

In producing the velar consonant, the length of the front and back cavity is approximately the same resulting in the well-known F2 and F3 convergence ("velar pinch") (Stevens, 1998). The velar spectrum in Figure 2.9 shows a main spectral peak at this region (i.e. ca.1500 Hz). For the uvular consonant, a convergence would be expected between F3 and F4 (Klatt and Stevens, 1969; Stevens, 1998). This is also what is observed in the current speech sample. The main peak for the uvular consonant was found at ca. 3000 Hz.

Thus, the spectral pattern for each place of articulation observed in Figure 2.9 can be said to be in reasonable agreement with well established theoretical predictions.

2.3 Perceptual data

2.3.1 Subjects and language groups

As part of the attempt to measure the perceptual differences between the CV syllables, I conducted a set of perception tests. The data for calibrating a measure of perceptual distance should *ideally* be independent of the linguistic experience of the subjects. The work exemplified by Liljencrants & Lindblom (1972) and Lindblom (1986) is based on the assumption that the vowel qualities of the world's languages have arisen within one and the same vowel space whose shape is fixed and determined by universal factors such as:

(i) the constraints imposed by the vocal tract on the range of possible vowels;

(ii) the constraints imposed by the mapping from articulatory configurations to acoustics (formant frequencies); and

(iii) the properties of the transformation from acoustic signal to an auditory representation.

In this approach, ‘similarity’ is a unique function of these constraints. The possibility of the vowel space being shaped, at least to some extent, by the experience of learning it, is neglected.

Lindblom (1986) discusses this fact. Using dissimilarity data on nine Swedish vowels (Hanson, 1967), he presents the result of plotting perceptual dissimilarity against spectral distance for each vowel individually. Overall, there is a tendency for dissimilarity to increase fairly linearly as a function of spectral distance. However, the slopes of these relationships are vowel-dependent. Front vowels show a steeper slope than back vowels. For instance, for a given distance - say that between [y:] and [ø:] which is comparable to that for [u:] and [o:] - the front pair is judged as more dissimilar. The effect is not large but it is there. “It is as if listeners make their space more spacious at the point where the universal perceptual space seems most crowded” (Lindblom 1986: 38).

However, the fact that “acquired similarity” may be a factor determining the shape of the native speaker’s vowel space does not invalidate attempts to assess the role of universal constraints in the patterning of the world’s consonant and vowel systems. But it does complicate the task of empirically observing the phonetic space in its “language-innocent” form: that is the shape determined solely by universal articulatory, acoustic and auditory mechanisms.

Since the language dependency problem cannot be avoided, the choice of subjects and experimental tasks has to be made with great care.

To address this concern, I selected subjects representing several different language backgrounds so as to allow me to estimate the magnitude of the native language effect on the results. This was done by correlating perceptual judgments with calculated distances for each language group separately and then comparing the weights that the multiple regression analysis assigns to the several independent variables. If these weights pattern in a reasonably uniform way then the language background plays only a minor role.

In the experiment, total of 20 subjects with four language groups were selected (5 subjects each for Korean, English, Hindi and Spanish language groups). The subjects were all students of the University of Texas at Austin except two subjects who are faculty members. They all reported normal hearing.

In selecting the language groups the consonant inventory of their native language was considered. Spanish and English have bilabial, dental/alveolar and velar stops with a voiceless/ voiced distinction. Korean also has the three places but a unique three-way distinction between lenis, fortis and aspirated rather than a voiceless/voiced distinction. Hindi is the most interesting case since it has retroflex stops and voiceless uvular stops in addition to the three places (bilabial, dental/alveolar, and velar). No language group has a dental/alveolar distinction (Maddieson, 1984).

Since English is known to have a substantial variety of dialects, subjects were restricted to natives of Austin and the vicinity area (including San Antonio and Houston).

Korean is said to have little dialect variation. Hence the dialect factor is not considered. Korean has some dialect variation with regard to tone, but it is not within the scope of the current research.

The Hindi subject group included three Urdu/Hindi bilingual speakers. However, this did not seem to pose a concern because Hindi and Urdu are said to have little difference in consonant inventory and sometimes treated as the same language in terms of consonant and vowel inventory (Maddieson 1984: 270).

Spanish is one of the languages which have a wide range of dialect variation. In this research, four Mexican Spanish speakers and one Panamanian Spanish speaker participated.

2.3.2 Procedures

2.3.2.1 Training sessions

First, the subject listened to the whole set of 35 syllables twice in a warm-up session. Then, the subject was asked to learn the symbol that should be assigned to each consonant. It was decided to assign arbitrary symbols to each consonant place rather than phonetic symbols or ordered numbers. Since there are seven different stops, seven symbols (!: bilabial, ~: dental, ;: alveolar, @: retroflex , %: palatal, +: velar, ^: uvular) were used. The symbols were selected so that they would be completely distinguishable from one anther and so that they should be completely arbitrary. Subjects were asked to take as much time as they wanted to learn the symbol assigned to each consonant.

After they were comfortable with the symbols, the subjects were asked to begin the trial run by simply guessing the symbol given above. The computer indicated right or wrong. If the subject got it wrong, (s)he made another response and the computer would give her/him more feedback. The procedure would be repeated for each trial stimulus until (s)he got each one right. There was no limit on response time. There were five blocks in the training session each of which corresponded to a given vowel environment. Each block was repeated twice making the trial run consist of 10 subtests. Since response time was not limited and the computer program repeated the stimuli until the subject chose the correct answer, the overall length of time of the trial session differed from subject to subject. On average, the whole trial session took about 40 minutes.

The results of subjects whose total error score (defined as a percentage-score, the number of incorrect responses divided by the total number of stimulus presentations) was more than 30 were not included in the analysis. There were 5 such subjects in the experiment.

2.3.2.2 Identification test

The subjects were then asked to identify each CV syllable they heard by selecting the symbol assigned to the consonant. There were 5 repetitions for each CV. Thus the total number of stimuli was 175 (35 CVs by 5 repetitions). There were 5 groups where each group contained 35 stimuli. The interval between the stimuli was set to 3 seconds while between groups was 10 seconds. Average duration per subject was about 15 minutes (divided into five 3-minute runs).

2.3.2.3 Dissimilarity judgment task

The subjects were asked to judge how dissimilar two CV syllables sounded using a scale of 0 (same) to 6 (maximally different). The interval between the CVs in a pair was set at 200 ms. The subjects were asked to respond as quickly as possible after they heard the stimuli. They needed to mark a corresponding number (representing distance) on given answer sheets. Each subject had five sessions. The response time was limited to 3 seconds.

These discrimination data were collected for each vowel context separately. Since there were 5 vowels, there were 5 subtests in a session. Each subtest was composed of 49 pairs of CV stimuli (7 consonants by 7 consonants with V constant = 49 pairs). In each subtest the vowel was kept constant which means that the subject's judgments could be expected to be related to the differences between the acoustic information in the stop burst and the formant transitions. Each subtest took 16 minutes. Since there were five sessions per subject, the whole dissimilarity task took 80 minutes.

Each subject was asked to come on three different calendar days (not necessarily consecutive). On the first day, the subject finished training session, identification tests and one session of the dissimilarity judgment task.

On the second and third day, the subject completed two sessions of the dissimilarity judgment task.

2.3.3 Results

From the identification results, confusion matrices were obtained. Since the identification test in this experiment is relatively easy (e.g. the test does not include stimuli in noise), the main contribution of the identification results is to provide a reference point for interpreting the dissimilarity data and evaluating the native language effect discussed in sections 2.3.3.2 and 2.3.3.4.

The results of the dissimilarity task were used to construct perceptual distance matrices to be compared with the acoustic distance.

2.3.3.1 Response bias

The response bias refers to the tendency of a subject to favor one CV syllable over another CV syllable.

To exemplify, in the identification test, the subjects may select [bi] more frequently than they select [d̪i]. We need to decide if this discrepancy is small enough to be ignored.

The response bias was calculated using the method proposed by Sidwell and Summerfield (1986) based on Luce (1963).

$$b = LN\sqrt{(R_a / R_b)} \qquad \text{(Equation 2-4)}$$

Where R_a is the product of the two cells corresponding to one response and R_b was the product of the two cells corresponding to the other response.

Now let’s consider the following matrix in Table 2.7.

	bi	d̪i
bi	92	2
d̪i	6	53

Table 2.7 Example of similarity matrix for response bias calculation

In this CV pair the head of each row represents the first stimulus in the pair and the head of each column indicates the second CV. The numbers in the cells are the averaged distance scores given by the subjects.

In this two by two matrix, the response bias b is:

$b = LN\sqrt{(92*6/2*53)}$ = 0.83.

A positive value indicates that the subjects have a tendency to choose the CV syllable at the head of the column. A negative value indicates the opposite. In the example above the subjects tend to favor [bi] over [d̪i].

The maximum possible absolute bias value can be calculated by providing each cell of one column with maximum score and each cell of the other column with minimum score. The largest possible value for each cell is 100 (20 subjects by 5 repetitions) and the minimum value is zero. However, if one of the four cells is filled with zero, the response bias cannot be calculated. Following Sidwell and Summerfield (1986: 286) I replaced zero by one as an approximation.

By (Equation 2-4),

$b = LN\sqrt{(100*100/1*1)}$ = 4.61.

The maximum response bias possible is *4.61* when all of the responses favor one

CV as shown in Table 2.8.

	bi	d̪i
bi	100	1
d̪i	100	1

Table 2.8 Example of maximum response bias (dissimilarity judgment task)

The response bias values for the identification test are shown in Table 2.9.

/i/ vowel context

	bi	d̪i	di	ɖi	ɟi	gi
d̪i	0.83					
di	1.37	-0.19				
ɖi	2.68	1.16	1.32			
ɟi	0.42	0.60	-0.21	-0.80		
gi	0.44	-0.38	0.17	-0.03	-0.30	
ɢi	-0.39	-0.81	-1.01	-1.32	-0.50	-0.52

/e/ vowel context

	bɛ	dɛ	dɛ	ɖɛ	ɟɛ	gɛ
dɛ	-0.49					
dɛ	-0.22	-0.47				
ɖɛ	1.73	1.15	1.29			
ɟɛ	0.33	0.90	-0.14	0.12		
gɛ	1.14	0.55	0.32	0.23	-0.49	
ɢɛ	0.73	0.44	1.07	-0.08	-0.31	-0.74

/a/ vowel context

	ba	d̪a	da	ɖa	ɟa	ga
d̪a	0.25					
da	0.21	-0.25				
ɖa	0.01	0.71	0.75			
ɟa	0.28	0.52	-0.13	-0.28		
ga	0.20	0.75	-0.01	-0.36	0.54	
ɢa	0.89	0.52	0.22	0.62	0.86	0.66

/o/ vowel context

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ
d̪ɔ	-0.20					
dɔ	0.30	-0.30				
ɖɔ	0.20	0.30	-0.24			
ɟɔ	0.06	0.26	-0.79	-0.13		
gɔ	0.23	0.88	-0.07	-0.31	0.47	
ɢɔ	-0.23	0.16	0.36	0.67	0.45	0.25

/u/ vowel context

	bu	d̪u	du	ɖu	ɟu	gu
d̪u	0.44					
du	0.31	-0.67				
ɖu	0.57	-0.52	-0.27			
ɟu	0.15	-0.64	-0.17	0.27		
gu	0.21	0.46	-0.10	-0.02	0.12	
ɢu	1.11	0.67	0.45	0.19	-0.01	0.66

Table 2.9 Response bias for identification test

In general, the response bias value is very low (absolute values are less than zero for most of the cases). The largest absolute value is *2.68* for [bi] and [ɖi].

The response bias values were also calculated for the dissimilarity data. Here, the largest possible response bias is *1.79* when all of the responses favor one CV as shown in Table 2.10.

	bi	d̪i
bi	6	1
d̪i	6	1

Table 2.10 Example of maximum response bias (identification test)

The maximum number in each cell is 6 because the maximum distance value for each CV pair is 6 and the response was averaged over the subjects. Again, 0 is replaced with 1.

/i/ vowel context

	bi	d̪i	di	ɖi	ɟi	gi
d̪i	-0.14					
di	-0.10	0.01				
ɖi	0.05	0.11	-0.16			
ɟi	0.75	0.90	0.79	0.68		
gi	0.32	0.49	0.33	0.24	-0.43	
ɢi	0.31	0.44	0.33	0.24	-0.43	-0.09

/e/ vowel context

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ
dɛ	-0.27					
dɛ	-0.25	-0.06				
ɖɛ	-0.28	0.15	-0.10			
ɟɛ	-0.17	0.19	0.11	0.17		
gɛ	-0.08	0.23	0.13	0.19	0.11	
ɢɛ	-0.09	0.17	0.13	0.23	0.01	-0.10

/a/ vowel context

	ba	d̪a	da	ɖa	ɟa	ga
d̪a	-0.54					
da	-0.54	0.03				
ɖa	-0.35	0.04	0.06			
ɟa	-0.04	0.44	0.43	0.31		
ga	-0.19	0.35	0.39	0.18	-0.13	
ɢa	-0.59	-0.07	-0.16	-0.20	-0.58	-0.59

/o/ vowel context

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ
d̪ɔ	-0.38					
dɔ	-0.59	-0.20				
ɖɔ	-0.29	0.12	0.01			
ɟɔ	-0.11	0.29	0.52	0.29		
gɔ	-0.01	0.32	0.60	0.30	0.07	
ɢɔ	0.13	0.48	0.74	0.45	0.28	0.14

/u/ vowel context

	bu	d̪u	du	ɖu	ɟu	gu
d̪u	-0.29					
du	-0.55	-0.07				
ɖu	-0.58	-0.20	-0.04			
ɟu	-0.18	0.21	0.40	0.43		
gu	-0.30	0.13	0.23	0.29	-0.09	
ɢu	-0.46	-0.07	0.10	0.10	-0.22	-0.28

Table 2.11 Response bias for the dissimilarity judgment task

As shown in Table 2.11, most of the response bias scores are small except a few pairs in the /a/ context. The observed maximum of the absolute response bias score is 0.79 in [di]-[ɟi] pair.

2.3.3.2 Identification test results

Confusion matrices were computed based on the responses of the subjects with head of each row representing the given stimulus and the head of each column indicating the subject's response (Appendix C.1). Each language group's responses were considered separately as well as pooled.

Since there was little response bias in the identification test (as discussed in 2.3.3.1), the confusion matrices were symmetrized using the method in Klein, Plomp, and Pols (1970). Shepard (1972) also argues for the idea of symmetrizing matrices suggesting that the asymmetries in confusion data are the result of two imperfect measures of the same thing (i.e. the confusability of the two items) and for the purpose of making a perceptual distance map, it can be justifiable to leave out the asymmetries.

The mean of the confusions between the two CVs are calculated to obtain symmetrized matrices.

$C = (C_{ij}+C_{ji})/2$ (Equation 2-5)

Where C represents confusions and i and j are the stimuli.

The symmetrized confusion matrices calculated by Equation 2-5 were also constructed by language group and pooled. Note that the correct responses are removed from the matrices.

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi							
d̪i	2						
di	0	14					
ɖi	8	15	17				
ɟi	0	0	0	0			
gi	0	0	0	0	14		
ɢi	0	0	0	0	8	15	

	bɛ	dɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ							
d̪ɛ	3						
dɛ	1	21					
ɖɛ	2	22	22				
ɟɛ	0	0	0	0			
gɛ	4	0	0	0	10		
ɢɛ	5	2	2	2	6	21	

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba							
d̪a	0						
da	0	18					
ɖa	0	11	13				
ɟa	0	0	0	0			
ga	0	0	0	0	5		
ɢa	0	0	0	0	2	14	

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ							
d̪ɔ	1						
dɔ	0	13					
ɖɔ	0	8	13				
ɟɔ	0	2	2	1			
gɔ	0	0	0	0	1		
ɢɔ	0	0	0	0	1	19	

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu							
d̪u	1						
du	0	16					
ɖu	0	9	12				
ɟu	0	0	0	0			
gu	0	0	0	0	6		
ɢu	0	0	0	0	6	20	

Table 2.12 Symmetrized responses for the identification task (Korean subjects)

In the Korean group (Table 2.12), most confusions are found among dentals, alveolars and retroflexes. Retroflexes are confused most frequently with alveolars (17 instances before /i/, 22 before /e/, 13 before /a/ and /o/ respectively, and 12 before /u/) and dentals (15 instances). There are also a lot of confusions between dental and alveolar (14 before /1/, 21 before /e/, 18 before /a/, 13 before /o/ and 16 before /u/). Confusions between dentals and alveolars are not unexpected because of their proximity in place of articulation. Maddieson (1984) also reports that distinctions between these two places are typologically unusual (1984: 32).

What is surprising are the numerous confusions between velars and uvulars (15 before /i/, 21 before /e/, 14 before /a/, 19 before /o/ and 20 before /u/) because they could be expected to sound very different (at least before front vowels).

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi							
d̪i	4						
di	4	13					
ɖi	10	13	17				
ɟi	0	2	1	2			
gi	0	2	2	2	21		
ɢi	0	2	2	2	6	10	

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ							
d̪ɛ	5						
dɛ	4	14					
ɖɛ	6	15	22				
ɟɛ	2	1	1	1			
gɛ	1	0	0	0	12		
ɢɛ	3	2	2	2	6	16	

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba							
d̪a	0						
da	0	23					
ɖa	0	12	12				
ɟa	0	0	0	0			
ga	0	0	0	0	1		
ɢa	0	0	0	0	1	11	

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ							
d̪ɔ	0						
dɔ	0	9					
ɖɔ	0	3	18				
ɟɔ	0	0	1	0			
gɔ	0	1	1	1	0		
ɢɔ	0	0	0	0	0	11	

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu							
d̪u	0						
du	0	17					
ɖu	2	7	14				
ɟu	0	0	0	0			
gu	0	1	1	2	1		
ɢu	0	2	2	3	0	16	

Table 2.13 Symmetrized responses for the identification task (English subjects)

The tendency that dentals, alveolars and retroflexes cause most confusions are also found in the English group (Table 2.13). For this group retroflexes are the most confusable stop consonant with alveolars, (17 instances before /i/, 22 before /e/, 12 before /a/, 18 before /o/, and 12 before /u/) and with dentals (13 before /i/, 15 before /e/, 12 before /a/, 3 before /o/, and 7 before /u/).

The palatal stop is confused with the velar stop especially before /i/ (21 instances) and before /e/ (12 instances) for this group. This may due to the fact that English has two allophones of /ɡ/. Before front vowels this segment is palatalized. Thus the palatal stop in this experiment corresponds to the velar phoneme before a front vowel for English speakers.

	bi	d̪i	di	ɖi	ɟi	ɡi	ɢi
bi							
d̪i	8						
di	6	12					
ɖi	8	9	20				
ɟi	4	6	2	3			
ɡi	7	10	6	8	19		
ɢi	5	9	6	5	15	17	

	bɛ	dɛ	dɛ	ɖɛ	ɟɛ	ɡɛ	ɢɛ
bɛ							
d̪ɛ	8						
dɛ	6	13					
ɖɛ	7	15	22				
ɟɛ	2	2	4	2			
ɡɛ	4	3	4	3	12		
ɢɛ	7	9	11	10	10	18	

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba							
d̪a	1						
da	0	1					
ɖa	0	3	17				
ɟa	4	4	2	2			
ga	2	2	4	3	7		
ɢa	5	3	5	4	11	18	

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ							
d̪ɔ	1						
dɔ	0	4					
ɖɔ	0	6	19				
ɟɔ	4	1	0	0			
gɔ	3	1	2	2	7		
ɢɔ	4	2	3	3	9	21	

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu							
d̪u	1						
du	0	22					
ɖu	1	14	13				
ɟu	1	4	3	6			
gu	1	4	3	5	7		
ɢu	2	1	0	2	1	15	

Table 2.14 Symmetrized responses for the identification task (Hindi subjects)

It is very interesting to note that Hindi speakers who have retroflexes in their consonant inventory also show the same pattern (Table 2.14). They tend to confuse retroflexes with alveolars (20 instances before /i/, 22 before /e/, 17 before /a/, 19 before /o/, and 13 before /u/) and dentals (9 instances before /i/, 15 before /e/, 3 before /a/, 6 before /o/, and 14 before /u/) just like those of other language groups.

It seems to suggest that, for the present stimuli, the retroflex is a difficult sound for them, too. In other words, they have trouble in distinguishing the retroflex when it does not appear in Hindi words.

	bi	d̪i	di	ɖi	ɟi	ɡi	ɢi
bi							
d̪i	0						
di	5	14					
ɖi	8	12	17				
ɟi	0	5	11	10			
ɡi	0	1	5	5	14		
ɢi	0	1	5	5	15	23	

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ							
dɛ	6						
dɛ	7	21					
ɖɛ	9	19	21				
ɟɛ	5	6	7	7			
gɛ	5	5	6	6	21		
ɢɛ	10	5	7	10	14	16	

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba							
d̪a	0						
da	2	20					
ɖa	3	16	18				
ɟa	4	3	4	8			
ga	3	5	5	8	13		
ɢa	4	4	4	8	16	21	

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ							
d̪ɔ	3						
dɔ	4	15					
ɖɔ	2	9	18				
ɟɔ	2	2	6	9			
gɔ	3	5	8	6	12		
ɢɔ	4	3	6	8	18	18	

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu							
d̪u	0						
du	0	16					
ɖu	0	14	21				
ɟu	0	8	9	10			
gu	0	7	7	6	12		
ɢu	3	8	8	7	15	17	

Table 2.15 Symmetrized responses for the identification task (Spanish subjects)

In general, the Spanish group (Table 2.15) shows the same pattern as the other groups but the confusion rate is generally higher.

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi							
d̪i	14						
di	17	63					
ɖi	36	61	79				
ɟi	7	16	17	16			
gi	7	15	16	15	74		
ɢi	6	12	13	12	46	67	

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ							
d̪ɛ	22						
dɛ	20	69					
ɖɛ	26	73	89				
ɟɛ	12	11	12	11			
gɛ	14	12	11	13	59		
ɢɛ	25	23	22	27	39	77	

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba							
d̪a	4						
da	2	66					
ɖa	3	51	63				
ɟa	8	8	7	10			
ga	7	10	9	12	32		
ɢa	11	10	9	12	34	69	

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ							
d̪ɔ	5						
dɔ	5	41					
ɖɔ	5	35	70				
ɟɔ	8	6	9	11			
gɔ	8	9	12	14	30		
ɢɔ	8	8	10	12	30	70	

	bu	d̪u	du	ɖu	ɟu	ɡu	ɢu
bu							
d̪u	2						
du	0	71					
ɖu	3	45	60				
ɟu	1	14	14	18			
ɡu	1	12	12	16	31		
ɢu	5	12	11	17	24	68	

Table 2.16 Symmetrized responses for the identification task (All subjects)

It is worth noting that the language groups show very similar patterns in the identification test. Especially the finding that Hindi speakers could not distinguish the retroflex very well is unexpected. It is striking because it appears justified to argue that a native language effect is somewhat inevitable when people respond to speech sounds. It is possible that the presence of highly confusable sounds like the dental and the alveolar and very unusual sound like the uvular makes them consider the stimuli as foreign rather than natural. The results would seem to suggest that pooling the data might be justified.

2.3.3.3 Dissimilarity judgment task results

The results of the dissimilarity judgment tasks are presented below in triangular matrices obtained from the seven by seven matrices by symmetrization. The maximum distance that any cell can contain is six.

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	0.24						
d̪i	2.34	0.48					
di	3.28	1.18	0.24				
ɖi	2.32	1.46	0.42	0.2			
ɟi	4.28	3.98	3.92	3.7	0.04		
gi	5.32	5.3	4.6	4.84	4.58	0.08	
ɢi	5.78	5.5	5.06	5.14	5.3	2.96	0.04

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	0.16						
d̪ɛ	2.3	0.28					
dɛ	2.18	0.4	0.4				
ɖɛ	2.56	0.76	0.98	0.32			
ɟɛ	4.86	4.36	4.2	4.06	0.12		
gɛ	4.38	3.36	3.54	3.38	4.24	0.12	
ɢɛ	3.96	3.82	3.44	3.7	4.5	1.58	0.2

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	0						
d̪a	3.38	0.6					
da	3.46	0.66	0.44				
ɖa	3.76	1.44	0.5	0.32			
ɟa	4.62	4.1	4.14	4.1	0.04		
ga	5.08	4.48	4.28	4.4	3.5	0.08	
ɢa	5.06	4.66	4.5	4.08	3.68	0.22	0.08

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	0.2						
d̪ɔ	3.7	0.32					
dɔ	3.9	1.66	0.48				
ɖɔ	3.66	1.84	0.98	0.24			
ɟɔ	4.56	4.2	3.8	3.8	0		
gɔ	4.92	4.16	4.08	3.84	4.08	0.08	
ɢɔ	4.62	4.2	4.16	4.14	4.18	0.46	0.04

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	0.04						
d̪u	3.42	0.2					
du	3.6	1.26	0.4				
ɖu	3.84	2.94	1.36	0.24			
ɟu	4.66	4.1	4.28	3.74	0.04		
gu	4.48	4.32	3.86	3.7	4.58	0.08	
ɢu	4.38	4.16	4.1	3.9	4.46	1.82	0.2

Table 2.17 Perceptual dissimilarity matrix for Korean subjects

The CV pairs showing maximum distance for the Korean group (Table 2.17) are labial and uvular before /i/ (5.78), labial and palatal before /e/ (4.86), labial and velar before /a/ (5.08), labial and velar before /o/ (4.92) and velar and palatal before /u/ (4.58).

The uvular and velar pairs show a very small distance in every vowel context (0.04 before /i/, 0.2 before /e/, 0.08 before /a/, 0.04 before /o/ and 0.2 before /u/) which is similar to the identification results(section 2.3.3.2).

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	0.16						
d̪i	2.4	0.28					
di	2.28	2.02	0.04				
ɖi	1.86	2.08	0.28	0.04			
ɟi	4.24	3.66	4.28	4.52	0.04		
gi	5.12	4.7	5.1	5	4.04	0.04	
ɢi	5.6	5.06	5.18	5.22	4.18	2.54	0.08

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	0.2						
d̪ɛ	2	0.16					
dɛ	2.78	0.38	0.04				
ɖɛ	2.68	1.26	0.68	0.12			
ɟɛ	4.68	4.2	4.38	4.26	0.2		
gɛ	4.16	3.8	4.02	4.08	4.12	0.12	
ɢɛ	3.22	2.82	2.76	3.46	4.38	3.14	0.04

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	0.04						
d̪a	3.42	0.08					
da	3.9	0.96	0.12				
ɖa	3.86	1.3	1.34	0.04			
ɟa	5.06	4.64	4.48	4.82	0		
ga	4.62	4.68	4.24	4.48	3.76	0.08	
ɢa	4.74	4.72	4.54	4.54	3.84	0.96	0.08

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	0						
d̪ɔ	4.12	0					
dɔ	4	2.38	0.48				
ɖɔ	4.38	2.62	1.14	0.12			
ɟɔ	4.88	4.24	4.02	4.18	0.2		
gɔ	4.76	4.12	4.48	4.32	3.96	0	
ɢɔ	4.86	4.16	4.68	4.7	4.22	1.38	0.04

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	0.08						
d̪u	3.62	0.04					
du	3.7	0.84	0.2				
ɖu	3.66	2.28	1.7	0.16			
ɟu	4.72	4.28	4.54	4.24	0.04		
gu	3.78	4.5	4.26	4.2	4.08	0.12	
ɢu	3.84	4.12	4.24	4.14	4.14	1.74	0.08

Table 2.18 Perceptual dissimilarity matrix for English subjects

The CV pairs showing maximum distances for the English group (Table 2.18) are uvular and retroflex before /i/ (5.22), labial and palatal before /e/ (4.68), labial and palatal before /a/ (5.06), labial and palatal before /o/ (4.88) and labial and palatal before /u/ (4.72).

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	0.28						
d̪i	2.1	0.56					
di	2.68	2.82	0.92				
ɖi	2.2	2.82	1.02	0.52			
ɟi	3.5	3.6	4.08	4.36	0.08		
gi	4.24	4.2	4.44	4.22	4.38	0.36	
ɢi	4.84	4.98	5.1	4.96	4.8	3.46	0.52

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	0.36						
d̪ɛ	1.7	0.64					
dɛ	2.16	1.4	0.68				
ɖɛ	3.14	2.12	1.08	0.92			
ɟɛ	4.42	4.34	4.12	4.24	0.88		
gɛ	3.32	3.22	3.76	3.54	4.16	0.4	
ɢɛ	3.24	3.2	3.28	3.66	4.8	2.82	0.76

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	0.2						
d̪a	3.92	0.36					
da	4.14	2.38	0.28				
ɖa	4.34	2.88	1	0.16			
ɟa	4.32	4.68	4.68	4.44	0.32		
ga	4.06	4.52	4.52	4.44	4.56	0.2	
ɢa	4.5	4.06	4.4	4.2	4.74	1.1	0.84

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	0.24						
d̪ɔ	4.22	0.68					
dɔ	4.62	3.04	0.56				
ɖɔ	4.06	3.3	1.62	0.64			
ɟɔ	4.46	4.22	4.02	4.48	0.36		
gɔ	4.5	3.68	4.12	3.8	4.88	0.44	
ɢɔ	3.78	4.04	4.24	4.58	4.62	0.8	0.24

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	0.2						
d̪u	3.72	0.44					
du	4.04	1.62	0.36				
ɖu	4.06	3.18	1.82	0.64			
ɟu	4.48	4.5	4.72	4.32	0.32		
gu	3.82	4.4	4.44	3.86	5.1	0.32	
ɢu	3.26	3.86	4.4	4.02	4.58	1.86	0.36

Table 2.19 Perceptual dissimilarity matrix for Hindi subjects

The CV pairs showing maximum distances for the Hindi group (Table 2.19) are uvular and dental before /i/ (4.98), labial and palatal before /e/ (4.42), palatal and uvular before /a/ (4.74), labial and alveolar before /o/ (4.62) and palatal and uvular before /u/ (4.58).

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	0.68						
d̪i	3.14	0.48					
di	2.82	1.38	0.24				
ɖi	2.26	1.72	0.72	0.48			
ɟi	4.28	4.54	3.96	4.42	0.16		
gi	5.06	4.62	4.08	4.82	4.26	0.28	
ɢi	5.26	5.12	5.4	5.04	5.04	2.68	0.12

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	0.24						
d̪ɛ	2.52	0.68					
dɛ	2.24	0.54	0.48				
ɖɛ	2.64	1.36	1.1	0.48			
ɟɛ	4.62	4.38	4.3	4.12	0.32		
gɛ	3.6	3.06	3.36	3.5	3.88	0.56	
ɢɛ	2.46	3.32	2.9	3.76	4.5	2.56	0.36

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	0.24						
d̪a	3.46	0.28					
da	3.3	0.66	0.48				
ɖa	3.78	1.36	0.98	0.48			
ɟa	4.86	4.58	4.28	4.28	0.2		
ga	4.18	4.02	4.02	3.82	3.96	0.32	
ɢa	4.18	4.42	4.36	4.26	4.32	0.46	0.68

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	0.28						
d̪ɔ	3.52	0.4					
dɔ	3.94	1.22	0.76				
ɖɔ	3.64	2.82	1.3	0.36			
ɟɔ	4.36	4.38	4	4.26	0.28		
gɔ	4.46	3.96	4.04	3.76	4.54	0.2	
ɢɔ	4.24	4.04	4.56	4.58	4.3	0.78	0.24

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	0.16						
d̪u	3.56	0.28					
du	3.9	1.22	0.48				
ɖu	4.1	3.42	1.58	0.44			
ɟu	4.94	3.98	4.46	4.4	0.32		
gu	4.32	4.3	4.48	4.22	4.52	0.32	
ɢu	3.7	4.3	4.46	4.2	4.34	1.9	0.48

Table 2.20 Perceptual dissimilarity matrix for Spanish subjects

The CV pairs of maximum distances for the Spanish group (Table 2.20) are labial and uvular before /i/ (5.26), labial and palatal before /e/ (4.62), labial and palatal before /a/ (4.86), retroflex and uvular before /o/ (4.58) and labial and palatal before /u/ (4.94).

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	0.34						
d̪i	2.495	0.45					
di	2.765	1.85	0.36				
ɖi	2.16	2.02	0.61	0.31			
ɟi	4.075	3.945	4.06	4.25	0.08		
gi	4.935	4.705	4.555	4.72	4.315	0.19	
ɢi	5.37	5.165	5.185	5.09	4.83	2.91	0.19

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	0.24						
d̪ɛ	2.13	0.44					
dɛ	2.34	0.68	0.4				
ɖɛ	2.755	1.375	0.96	0.46			
ɟɛ	4.645	4.32	4.25	4.17	0.38		
gɛ	3.865	3.36	3.67	3.625	4.1	0.3	
ɢɛ	3.22	3.29	3.095	3.645	4.545	2.525	0.34

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	0.12						
d̪a	3.545	0.33					
da	3.7	1.165	0.33				
ɖa	3.935	1.745	0.955	0.25			
ɟa	4.715	4.5	4.395	4.41	0.14		
ga	4.485	4.425	4.265	4.285	3.945	0.17	
ɢa	4.62	4.465	4.45	4.27	4.145	0.685	0.42

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	0.18						
d̪ɔ	3.89	0.35					
dɔ	4.115	2.075	0.57				
ɖɔ	3.935	2.645	1.26	0.34			
ɟɔ	4.565	4.26	3.96	4.18	0.21		
gɔ	4.66	3.98	4.18	3.93	4.365	0.18	
ɢɔ	4.375	4.11	4.41	4.5	4.33	0.855	0.14

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	0.12						
d̪u	3.58	0.24					
du	3.81	1.235	0.36				
ɖu	3.915	2.955	1.615	0.37			
ɟu	4.7	4.215	4.5	4.175	0.18		
gu	4.1	4.38	4.26	3.995	4.57	0.21	
ɢu	3.795	4.11	4.3	4.065	4.38	1.83	0.28

Table 2.21 Perceptual dissimilarity matrix averaged over language groups

In the pooled data, The CV pairs showing maximum distances (Table 2.21) are labial and uvular before /i/ (5.37), labial and palatal before /e/ (4.65), labial and velar before /a/ (4.49), labial and palatal before /o/ (4.57) and labial and palatal before /u/ (4.7).

The results of the pooled data indicate that the subjects tend to give larger numbers to pairs of labial-uvular, labial-palatal, and labial-velar whereas the velar-uvular pair obtains a small number in the pooled data. So does the dental-alveolar pair.

2.3.3.4 Effect of language groups

In the dissimilarity judgment task as well as the identification test, the language groups show strikingly similar response patterns. As discussed in section 2.3.3.2, even Hindi speakers share this pattern with other language groups also with regard to the retroflex stop. The correlation coefficients between language groups for the dissimilarity judgment task (Table 2.22) varied from 0. 93 (English vs. Hindi) to 0.97 (Korean vs. Spanish and English vs. Spanish). This high correlation leads some support to the assumption of previous research (e.g. Lindblom, 1986) which was based on assuming that the vowel qualities of the world's languages have arisen within one and the same vowel space whose shape is fixed and determined only by universal factors.

	Korean	English	Hindi	Spanish
Korean	1			
English	0.96	1		
Hindi	0.95	0.93	1	
Spanish	0.97	0.97	0.94	1

Table 2.22 Correlation coefficients among language groups in the dissimilarity judgment task

The finding is also in agreement with Iverson and Evans (2007: 2842) who suggest that “there is a surprising degree of uniformity in the ways that individuals with different language backgrounds perceive second language vowels.”

Chapter 3: Acoustic Distances as a Measure of Perceptual Contrast

3.1 Development of a measure of acoustic distance

3.1.1 Formant distances

In the introduction I made some informal observations noting that syllables like [ɢi] and [ɟu] sound very different. In other words they score high on perceptual contrast whereas [ɟi] and [gu] appear perceptually more similar. How can such judgments be tested and anchored in acoustic attributes? How can the perceptual contrasts among CV syllables be defined and quantified?

Figure 3.1 compares two hypothetical F2 transitions associated with two CV syllables sharing the same vowel but differing in the initial consonant.

It does not seem unrealistic to assume that the information on these formant transitions would be available in the auditory system and that their

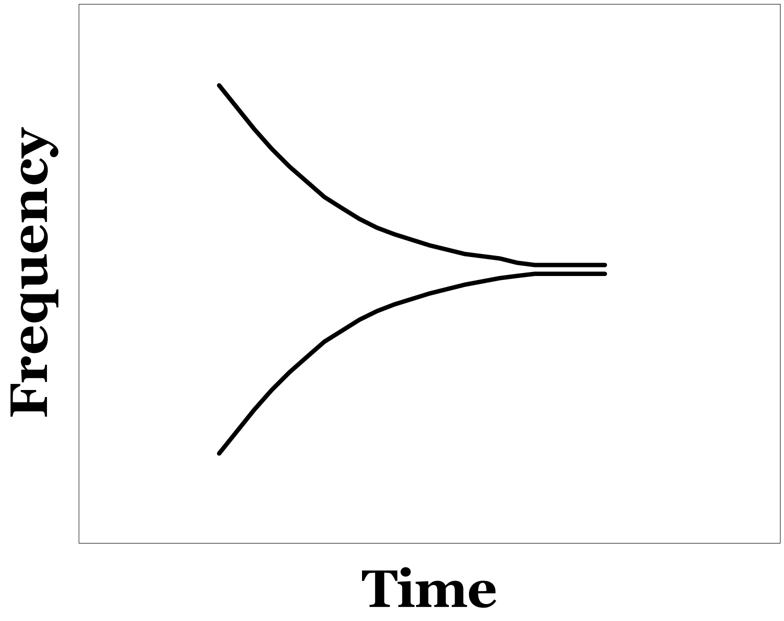


Figure 3.1 Hypothetical F2 transitions associated with two CV syllables

perceptual difference is related to some geometric difference between the two formant curves. In analogy with the distance measure used in quantifying Vowel contrast (Lindblom, 1986), it seems possible to consider the frequency-time area that the two curves of Figure 3.1 enclose (see also Klatt, 1979). Roughly speaking, the probability that the auditory system would confuse the two might be inversely related to the frequency-time area between them.

Pursuing this reasoning I decided to simplify the task of deriving that area. I shall assume that the two curves have identical exponential decay. This assumption is at variance with my empirical findings (see section 2.2.1) but appears justified as a first-order approximation of a more detailed measurement.

Given this simplifying assumption I find that the distance between the two curves becomes proportional to the difference between their locus-target distance, the contour between locus and target being predictable.

The measure I shall explore can thus be written as

$[\text{LOCUS}\ (C_i) - \text{TARGET}\ (V_k)] - [\text{LOCUS}\ (C_j) - \text{TARGET}\ (V_l)]$

Since $V_k = V_l$ the expression reduces to

$\text{LOCUS}\ (C_i) - \text{LOCUS}\ (C_j)$.

In other words, for same-vowel comparisons, it would be sufficient to limit the comparison to the formant differences at vowel onsets disregarding the formants of the vowel mid-point. The procedure is the following:

(a) Specification of formant frequencies

(b) Convert the formant into Mel values using the formula (Fant, 1960)

$$M_n = (1000/LN(2))*LN(1+F_n/1000)$$

(c) Find the perceptual distance using the formula

$$D_{ij} = [(M1_i-M1_j)^2+(M2_i-M2_j)^{2+}(M3_i-M3_j)^2]^{1/2}$$

Where $M1_i$ and $M1_j$ represent the mel values of the formant onsets (loci) of syllable *i* and syllable *j*.

The formant-based distance matrices obtained by the procedure in (a)-(c) and calibrated in mel units are shown in Table 3.1.

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	0						
d̪i	143	0					
di	163	83	0				
ɖi	163	52	52	0			
ɟi	342	345	355	324	0		
gi	422	315	278	301	597	0	
ɢi	600	462	447	441	641	641	0

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	0						
d̪ɛ	78	0					
dɛ	64	41	0				

ɖɛ	110	37	59	0			
ɟɛ	481	473	442	456	0		
gɛ	120	185	181	214	557	0	
ɢɛ	208	261	268	295	675	130	0

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	0						
d̪a	236	0					
da	313	96	0				
ɖa	302	188	236	0			
ɟa	596	405	310	480	0		
ga	271	242	285	131	505	0	
ɢa	223	156	227	80	501	119	0

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	0						
d̪ɔ	321	0					
dɔ	383	119	0				
ɖɔ	601	456	500	0			
ɟɔ	655	389	335	386	0		
gɔ	172	237	319	430	527	0	
ɢɔ	138	254	346	485	581	71	0

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	0						
d̪u	396	0					
du	445	96	0				
ɖu	497	328	306	0			
ɟu	884	502	442	612	0		
gu	113	301	342	390	783	0	
ɢu	24	401	445	500	883	115	0

Table 3.1 Distance matrices based on formant 1-3 (in Mel)

In the [i] context, [ɢi] is maximally different from other Ci syllables (average distance of [ɢi] from other Ci is 431.8 which far exceeds those of other Ci's: for [bi] it's 305.5, for [d̪i] it's 215.8). Another front vowel [ɛ] shows a similar result. [ɟɛ] and [ɢɛ] are the two Cɛ's that are maximally different from other Cɛ's.

In the [a] context, [ɟa] and [ba] are particularly salient with the average distance of 466.17 for [ɟa] and 323.5 for [ba]. In the [u] context, [ɟu] is the most salient syllable.

In the perceptual dissimilarity matrices numbers range between 0 and 6 but they show a similar pattern. In the [i] context, [ɢi] stands out with the average distance of 3.5 from other Ci's. In the [u] context, [ɟu] is most different with the average distance of 2.4 from other Cu's.

3.1.2 Time constant differences

The time constant differences matrices were compiled for each formant separately. They were obtained by calculating Equation 3-1:

$$\alpha \text{ distance} = \alpha(F_{ni}) - \alpha(F_{nj}) \qquad \text{(Equation 3-1)}$$

where *Fn* stands for the *n*th formant and *i* and *j* represent different CV syllables.

As discussed in section 2.2.2, it is extremely hard to measure the formant at several time points when the difference between the formant at CV boundary and the formant at mid-vowel is small (less than 100 Hz). In view of that consideration I replaced the time constant values of such cases by averaging the values over vowels. For example, if the F1 time constant for [bi] had to be removed, I would calculate a mean F1 time constant for [bɛ], [ba], [bɔ] and [bu]. Time constant differences were obtained after I had filled the empty slots with the average number in question.

	F1						
	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi							
d̪i	0.1952						
di	0.2645	0.0693					
ɖi	0.0123	0.1829	0.2522				
ɟi	0.0469	0.2422	0.3115	0.0593			
gi	0.0458	0.2410	0.3103	0.0581	0.0012		
ɢi	0.0258	0.1694	0.2387	0.0135	0.0728	0.0716	

	F2						
	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi							
d̪i	0.0132						
di	0.0249	0.0117					
ɖi	0.0223	0.0091	0.0026				
ɟi	0.0283	0.0415	0.0532	0.0506			
gi	0.0300	0.0432	0.0549	0.0523	0.0017		
ɢi	0.0254	0.0386	0.0503	0.0477	0.0029	0.0046	

	F3						
	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi							
d̪i	0.0090						
di	0.0132	0.0042					
ɖi	0.0181	0.0091	0.0049				
ɟi	0.0193	0.0103	0.0061	0.0012			
gi	0.0387	0.0297	0.0255	0.0206	0.0194		
ɢi	0.0439	0.0349	0.0307	0.0258	0.0246	0.0052	

	F1						
	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ							
d̪ɛ	0.0909						
dɛ	0.0477	0.0432					
ɖɛ	0.0513	0.0396	0.0036				
ɟɛ	0.1102	0.0193	0.0625	0.0589			
gɛ	0.1089	0.0180	0.0612	0.0576	0.0013		
ɢɛ	0.0412	0.0497	0.0065	0.0101	0.0690	0.0677	

	F2						
	bε	d̪ε	dε	ɖε	ɟε	gε	ɢε
bε							
d̪ε	0.0227						
dε	0.0331	0.0104					
ɖε	0.0330	0.0103	0.0001				
ɟε	0.0021	0.0206	0.0310	0.0309			
gε	0.0343	0.0116	0.0012	0.0013	0.0322		
ɢε	0.0297	0.0070	0.0034	0.0033	0.0276	0.0029	

	F3						
	bε	d̪ε	dε	ɖε	ɟε	gε	ɢε
bε							
d̪ε	0.0155						
dε	0.0103	0.0258					
ɖε	0.0211	0.0056	0.0314				
ɟε	0.0164	0.0319	0.0061	0.0375			
gε	0.0211	0.0056	0.0314	0	0.0375		
ɢε	0.0182	0.0027	0.0285	0.0029	0.0346	0.0029	

	F1						
	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba							
d̪a	0.0681						
da	0.0634	0.0047					
ɖa	0.0416	0.0265	0.0218				
ɟa	0.0730	0.0049	0.0096	0.0314			
ga	0.0587	0.0094	0.0047	0.0171	0.0143		
ɢa	0.0580	0.0101	0.0054	0.0164	0.0150	0.0007	

	F2						
	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba							
d̪a	0.0043						
da	0.0021	0.0063					
ɖa	0.0072	0.0029	0.0092				
ɟa	0.0044	0.0001	0.0064	0.0028			
ga	0.0051	0.0008	0.0071	0.0021	0.0007		
ɢa	0.0534	0.0576	0.0513	0.0605	0.0577	0.0372	

	F3						
	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba							
d̪a	0.0161						
da	0.0233	0.0072					
ɖa	0.0202	0.0041	0.0031				
ɟa	0.0144	0.0017	0.0089	0.0058			
ga	0.0440	0.0279	0.0207	0.0238	0.0296		
ɢa	0.0068	0.0093	0.0165	0.0134	0.0372	0.0372	

	F1						
	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ							
d̪ɔ	0.0287						
dɔ	0.0384	0.0671					
ɖɔ	0.0041	0.0246	0.0425				
ɟɔ	0.0045	0.0332	0.0339	0.0086			
gɔ	0.0206	0.0493	0.0178	0.0247	0.0161		
ɢɔ	0.0110	0.0397	0.0274	0.0151	0.0065	0.0096	

	F2						
	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ							
d̪ɔ	0.0137						
dɔ	0.0092	0.0045					
ɖɔ	0.0003	0.0134	0.0089				
ɟɔ	0.0077	0.0060	0.0015	0.0074			
gɔ	0.0125	0.0012	0.0033	0.0122	0.0048		
ɢɔ	0.0088	0.0050	0.0004	0.0085	0.0011	0.0139	

	F3						
	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ							
d̪ɔ	0.0049						
dɔ	0.0078	0.0029					
ɖɔ	0.0057	0.0008	0.0021				
ɟɔ	0.0029	0.0078	0.0107	0.0086			
gɔ	0.0212	0.0163	0.0134	0.0155	0.0241		
ɢɔ	0.0073	0.0024	0.0005	0.0016	0.0139	0.0139	

	F1						
	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu							
d̪u	0.0130						
du	0.0464	0.0335					
ɖu	0.0011	0.0119	0.0453				
ɟu	0.0469	0.0599	0.0934	0.0481			
gu	0.0470	0.0600	0.0934	0.0481	0		
ɢu	0.0542	0.0672	0.1006	0.0553	0	0.0072	

	F2						
	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu							
d̪u	0.0093						
du	0.0108	0.0127					
ɖu	0.0114	0.0095	0.0222				
ɟu	0.0063	0.0044	0.0171	0.0051			
gu	0.0117	0.0136	0.0009	0.0231	0.0180		
ɢu	0.0088	0.0107	0.0021	0.0202	0.0151	0.0534	

	F3						
	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu							
d̪u	0.0615						
du	0.0744	0.0129					
ɖu	0.0479	0.0136	0.0265				
ɟu	0.0517	0.0098	0.0227	0.0038			
gu	0.0033	0.0648	0.0777	0.0512	0.0550		
ɢu	0.0501	0.0114	0.0243	0.0022	0.0534	0.0534	

Table 3.2 Distance matrices for formant time constants

The time constants can be assumed to capture the dynamics of formant transitions from burst to vowel target. If indeed formant dynamics contributes to the listener's percept of a syllable, it might influence her/his judgment of distance. Conceivably, time constant distances might thus be a component in explaining the listener's dissimilarity judgment responses.

3.1.3 Burst spectra

Since burst spectra depend on the place of articulation of the consonant, they contribute another potential cue in the perception of that consonant.

Differences between observed burst spectra were calculated as follows:

$$D_{ij} = \sqrt{\sum (S_{in} - S_{jn})^2} \qquad \text{(Equation 3-2)}$$

where S_{in} and S_{jn} represent the spectral levels at frequency n in syllables i and j in analogy with the distance measure proposed by Plomp (1970).

	bi	d̪i	di	ɖi	ɟi	ɡi	ɢi
bi	0.0						
d̪i	10.1	0.0					
di	11.8	9.7	0.0				
ɖi	11.2	9.7	8.7	0.0			
ɟi	13.8	7.7	10.1	8.6	0.0		
ɡi	13.9	8.7	12.2	12.0	7.2	0.0	
ɢi	15.2	13.6	12.6	10.8	10.7	12.0	0.0

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	ɡɛ	ɢɛ
bɛ	0.0						
d̪ɛ	12.2	0.0					
dɛ	12.9	8.3	0.0				
ɖɛ	5.3	10.0	10.1	0.0			
ɟɛ	13.2	16.6	12.0	14.6	0.0		
ɡɛ	11.7	10.0	12.0	10.9	11.6	0.0	
ɢɛ	18.3	16.3	12.9	17.9	11.7	10.8	0.0

	ba	d̪a	da	ɖa	ɟa	ɡa	ɢa
ba	0.0						
d̪a	9.6	0.0					
da	10.7	11.9	0.0				
ɖa	10.8	12.7	11.9	0.0			
ɟa	13.7	10.8	11.5	15.1	0.0		
ɡa	10.8	10.5	11.8	13.4	12.0	0.0	
ɢa	14.1	10.8	9.6	12.1	8.1	10.1	0.0

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	ɡɔ	ɢɔ
bɔ	0.0						
d̪ɔ	7.9	0.0					
dɔ	12.2	11.6	0.0				
ɖɔ	12.9	12.2	12.0	0.0			
ɟɔ	16.8	18.5	15.6	11.8	0.0		
ɡɔ	10.1	6.8	10.3	10.0	17.0	0.0	
ɢɔ	11.8	9.4	13.8	12.1	19.7	8.7	0.0

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	0.0						
d̪u	9.8	0.0					
du	14.0	9.8	0.0				
ɖu	14.1	10.9	9.4	0.0			
ɟu	13.0	12.3	17.7	16.7	0.0		
gu	19.0	13.4	10.9	10.9	18.0	0.0	
ɢu	20.2	14.2	11.4	12.2	18.8	8.5	0.0

Table 3.3 Distance matrices for bursts

3.2 Results

Multiple regression analyses were performed with averaged dissimilarity results as the dependent variable and three independent variables as defined above: (i) differences in locus patterns, (ii) differences in transition rates and (iii) differences between burst spectra.

A primary question addressed by such an analysis is whether there is a correlation at all between acoustic dimensions and the perceptual scores. It is not a priori obvious that our choice of simplified difference measures is relevant. Then, if a correlation does exist, I proceed to examining the relative roles of the three independent variables in determining the listener's distance responses.

Each cell of the dependent variable contained an average dissimilarity score which took on a value ranging between zero to six (Recall that subjects were instructed to respond by a number between zero and six). Plots of averaged dissimilarity scores against formant-based distances were first prepared as an exploratory step in the data processing. This exercise revealed a pattern which is illustrated in Figure 3.2 with results from the [i] context.

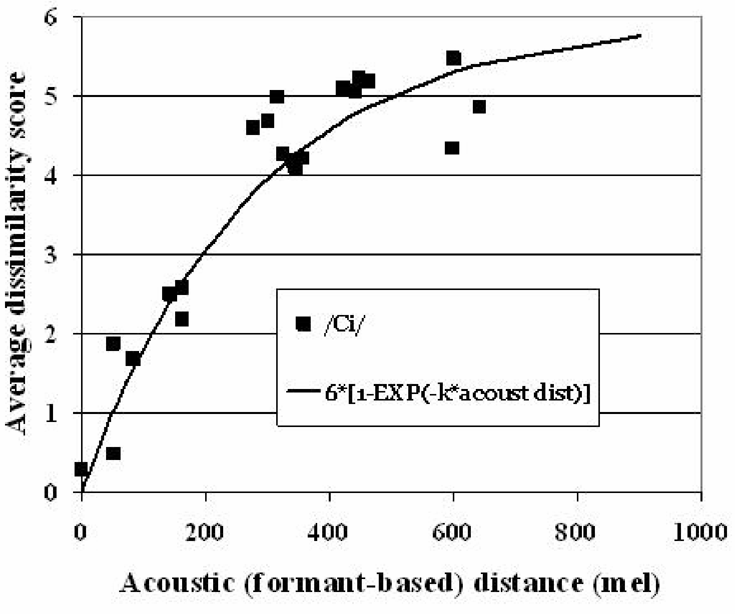


Figure 3.2 Plots of averaged dissimilarity scores against formant-based distances compared with upside down exponential curve (6*[**1-e**$^{\text{(-k* acoust dist)}}$])

Acoustic distance is here defined in terms of formant-based distance. The dissimilarity data points appear as a cluster whose general form resembles an upside down exponential curve, *viz* ($\mathbf{1\text{-}e^{(-k^{*}\ acoust\ dist)}}$).

To fit this curve shape to the data I divided the number in each cell by 6. I then subtracted the result from 1 so as to obtain a number equal to [1-Perceptual Dissimilarity/6]. Next, I took the natural logarithm of each [1-Perceptual Dissimilarity/6] number and plotted the result against acoustic distance. I fitted a straight line to the data points to check out the goodness of fit and determine the value of *k* (which measures the slope of the straight line, or equivalently the curvature of the $\mathbf{(1\text{-}e^{(-k^{*}\ acoust\ dist)})}$ line**.**

Having done this for all the data, I concluded that the upside-down curve was a satisfactory approximation of the perceptual data and that transforming the raw scores into LN[1-Perceptual Dissimilarity/6] would be helpful in running the Multi-Regressions analyses.

Predictor variables were formant-based distances (in Mel), spectral distances at the burst, the time constant distances for F1, the time constant distance of F2 and the time constant distance for F3.

3.2.1 Formants

Multiple regression analysis with formant distance alone as a factor (Table 3.4) showed that 52% of the dissimilarity judgments (coded as LN[1-Perceptual Dissimilarity/6]) could be explained in terms of formant distance alone. R^2=0.52 for pooled data.

Regression Statistics	
Multiple R	0.72
R Square	**0.52**
Adjusted R Square	0.52
Standard Error	0.38
Observations	245

ANOVA

	df	SS	MS	F	Significance F
Regression	1	38.38	38.38	265.20	0.00
Residual	243	35.17	0.14		
Total	244	73.54			

Table 3.4 Results of multiple regression analysis with formant distance as a predictor (vowels pooled)

The R^2 varies according to the following vowel. Before the front vowels, R^2 values are higher than before back vowels with R^2 of 0.65 before /i/ and 0.81 before /e/ (Tables 3.5 and 3.6).

Regression Statistics	
Multiple R	0.80
R Square	**0.65**
Adjusted R Square	0.64
Standard Error	0.42
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	1	15.13	15.13	85.56	0.00
Residual	47	8.31	0.18		
Total	48	23.44			

Table 3.5 Results of multiple regression analysis with formant distance as a predictor (/i/ vowel)

Regression Statistics	
Multiple R	0.90
R Square	**0.81**
Adjusted R Square	0.81
Standard Error	0.20
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	1	8.11	8.11	204.86	0.00
Residual	47	1.86	0.04		
Total	48	9.98			

Table 3.6 Results of multiple regression analysis with formant distance as a predictor (/e/ vowel)

The R^2 is markedly low in back vowel contexts (0.42 for /a/, 0.39 for /o/, and 0.59 for /u/).

Regression Statistics	
Multiple R	0.65
R Square	**0.42**
Adjusted R Square	0.41
Standard Error	0.42
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	1	6.039229	6.039229	33.97276	4.89E-07
Residual	47	8.35504	0.177767		
Total	48	14.39427			

Table 3.7 Results of multiple regression analysis with formant distance as a predictor (/a/ vowel)

Regression Statistics	
Multiple R	0.62
R Square	**0.39**
Adjusted R Square	0.37
Standard Error	0.40
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	1	4.61	4.61	29.43	0.00
Residual	47	7.36	0.16		
Total	48	11.96			

Table 3.8 Results of multiple regression analysis with formant distance as a predictor (/o/ vowel)

Regression Statistics	
Multiple R	0.77
R Square	**0.59**
Adjusted R Square	0.58
Standard Error	0.32
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	1	6.72	6.72	67.56	0.00
Residual	47	4.67	0.10		
Total	48	11.39			

Table 3.9 Results of multiple regression analysis with formant distance as a predictor (/u/ vowel)

3.2.2 Formants and time constants

Including time constants for F1, F2 and F3 was found to improve the prediction of perceptual distance results. In vowel pooled data, R^2 improves to 0.62 (Table 3.10).

Regression Statistics	
Multiple R	0.79
R Square	**0.62**
Adjusted R Square	0.62
Standard Error	0.34
Observations	245

ANOVA

	df	SS	MS	F	Significance F
Regression	4	45.76	11.44	98.84	0.00
Residual	240	27.78	0.12		
Total	244	73.54			

Table 3.10 Results of multiple regression analysis with formant distance and time constant as predictors (vowels pooled)

The time constant's prediction power is especially strong in the /i/ vowel context. Adding the time constant as a predictor increases the R^2 as high as 0.96 (Table 3.11). However, the asymmetry in R^2 between front and back vowel contexts is maintained. While R^2 before /e/ vowel reaches 0.89 (Table 3.12), R^2 values before backs vowels are comparatively lower.

Regression Statistics	
Multiple R	0.98
R Square	**0.96**
Adjusted R Square	0.95
Standard Error	0.15
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	4	22.39	5.60	234.37	0.00
Residual	44	1.05	0.02		
Total	48	23.44			

Table 3.11 Results of multiple regression analysis with formant distance and time constant as predictors (/i/ vowel)

Regression Statistics	
Multiple R	0.95
R Square	**0.89**
Adjusted R Square	0.89
Standard Error	0.15
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	4	8.93	2.23	93.49	0.00
Residual	44	1.05	0.02		
Total	48	9.98			

Table 3.12 Results of multiple regression analysis with formant distance and time constant as predictors (/e/ vowel)

Regression Statistics	
Multiple R	0.74
R Square	**0.55**
Adjusted R Square	0.51
Standard Error	0.38
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	4	7.88	1.97	13.30	0
Residual	44	6.58	0.15		
Total	48	14.398			

Table 3.13 Results of multiple regression analysis with formant distance and time constant as predictors (/a/ vowel)

Regression Statistics	
Multiple R	0.71
R Square	**0.51**
Adjusted R Square	0.46
Standard Error	0.37
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	4	6.08	1.52	11.36	0.00
Residual	44	5.89	0.13		
Total	48	11.96			

Table 3.14 Results of multiple regression analysis with formant distance and time constant as predictors (/o/ vowel)

The effect of adding time constants as predictors is significant in the /u/ vowel context (Table 3.15) where R^2 increases by 0.17 (from 0.59 to 0.76).

Regression Statistics	
Multiple R	0.87
R Square	**0.76**
Adjusted R Square	0.73
Standard Error	0.25
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	4	8.61	2.15	34.16	0.00
Residual	44	2.77	0.06		
Total	48	11.39			

Table 3.15 Results of multiple regression analysis with formant distance and time constant as predictors (/u/ vowel)

3.2.3 Formants, time constants and bursts

Adding burst spectra to the predictors of the regression analyses shows little improvement in front vowel contexts (Tables 3.17 and 3.18), mainly because they are already high and there is a ceiling effect. In other words, R²'s are so high that there is no room for improvement.

Regression Statistics	
Multiple R	0.81
R Square	**0.66**
Adjusted R Square	0.66
Standard Error	0.32
Observations	245

ANOVA

	df	SS	MS	F	Significance F
Regression	5	48.83	9.77	94.42	0.00
Residual	239	24.72	0.10		
Total	244	73.54			

Table 3.16 Results of multiple regression analysis with formant distance, time constant and burst spectrum distance as predictors (vowels pooled)

Regression Statistics	
Multiple R	0.98
R Square	**0.96**
Adjusted R Square	0.96
Standard Error	0.14
Observations	50.00

ANOVA

	df	SS	MS	F	Significance F
Regression	5	23.49	4.70	227.39	0.00
Residual	44	0.91	0.02		
Total	49	24.40			

Table 3.17 Results of multiple regression analysis with formant distance, time constant and burst spectrum distance as predictors (/i/ vowel)

Regression Statistics	
Multiple R	0.93
R Square	**0.86**
Adjusted R Square	0.84
Standard Error	0.58
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	5	87.39	17.48	52.84	0.00
Residual	43	14.22	0.33		
Total	48	101.61			

Table 3.18 Results of multiple regression analysis with formant distance, time constant and burst spectrum distance as predictors (/e/ vowel)

We note that there are some improvements in R^2 values for the back vowel environments (Tables 3.19-21).

Regression Statistics	
Multiple R	0.80
R Square	**0.64**
Adjusted R Square	0.59
Standard Error	1.09
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	5	88.36	17.67	14.99	0.00
Residual	43	50.68	1.18		
Total	48	139.04			

Table 3.19 Results of multiple regression analysis with formant distance, time constant and burst spectrum distance as predictors (/a/ vowel)

Regression Statistics	
Multiple R	0.78
R Square	**0.61**
Adjusted R Square	0.56
Standard Error	1.05
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	5	72.31	14.46	13.18	0.00
Residual	43	47.19	1.10		
Total	48	119.50			

Table 3.20 Results of multiple regression analysis with formant distance, time constant and burst spectrum distance as predictors (/o/ vowel)

Regression Statistics	
Multiple R	0.93
R Square	**0.86**
Adjusted R Square	0.85
Standard Error	0.60
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	5	97.87	19.57	55.04	0.00
Residual	43	15.29	0.36		
Total	48	113.16			

Table 3.21 Results of multiple regression analysis with formant distance, time constant and burst spectrum distance as predictors (/u/ vowel)

3.2.4 Summary of the results

Multiple regression analysis with formant distance alone as a predictor variable shows that 52% of the perceptual distance (i.e. results of the dissimilarity judgment task) can be explained with formant distance alone (R^2=0.52) in pooled data. Adding the time constant improves the predictions somewhat (R^2 rises to 0.62) whereas burst spectra differences contribute very little in explaining the perceptual distance results (Figure 3.3).

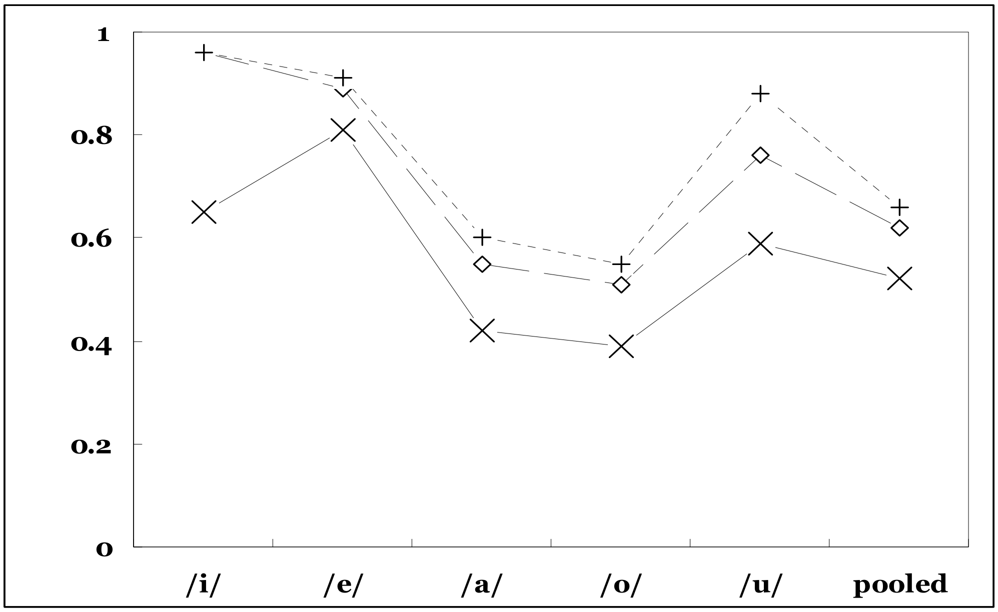


Figure 3.3 R^2 values with different predictor variables for the different vowel contexts

Solid line with crosses: formant distance

Broken line: formant distance and the time constant

Dotted lines: formant distance, the time constant and the burst distance

In the front vowel contexts, formant-based distance and the time constants can explain the perceptual distance rather successfully (R^2 = 0.96 for

/i/ context and $R^2 = 0.89$ for /e/ context) while in the back vowel contexts adding the dimension of burst spectrum differences helps improving the R^2 values (especially for /a/ context where $R^2 = 0.55$ with formant and the time constant as predictor variables while $R^2 =0.64$ with formant, the time constant and burst spectra distance as predictor variables).

Chapter 4: Articulatory Distances Combined with Acoustic Distances as an Extended Measure of Perceptual Contrast

4.1 Mirror neurons

A mirror neuron, discovered by Vittorio Gallese, Luciano Fadiga, Leonardo Fogassi and Giacomo Rizzolatti (1996) in experiments with macaque monkeys, is a neuron that fires both when animals act and when they see others conduct the action. The original observation concerned neurons in the ventral premotor cortex that were activated both when the macaque monkey grasped a nut and when the macaque saw a human do the same.

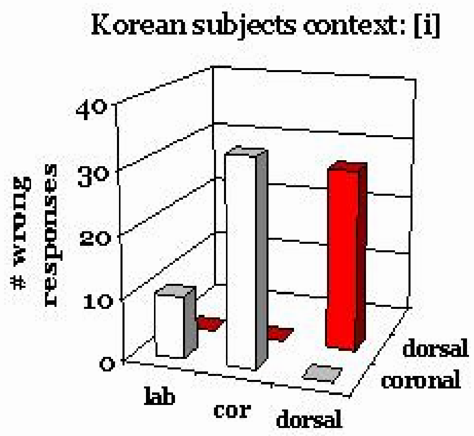
The existence of mirror neurons has been claimed to be important for understanding the intentions (Fogassi et al., 2005), and feelings of others (Wicker et al., 2003; Singer et al., 2004) and gestural communication (Skoyles, 2000).

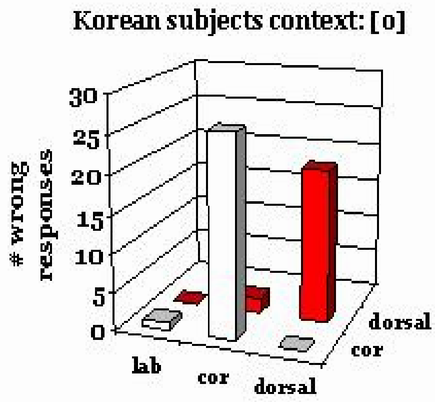
It has also been proposed that mirror neurons are responsible for language learning by imitation (Gallese et al., 1996; Rizzolatti and Arbib, 1998). Even though the question of the actual existence and location of neurons responsible for language is yet to be answered, it is interesting to note that such neurons do indeed exist in non-human primates and function as a link between the sender and the receiver in communication.

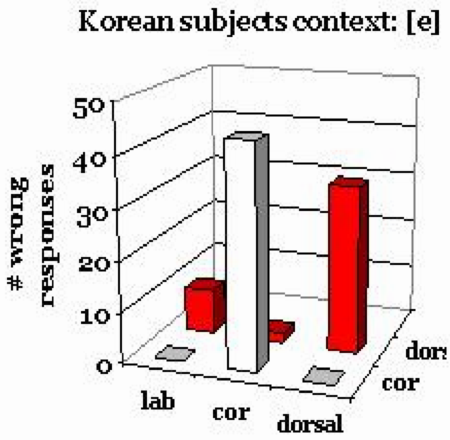
While performing my own experiments, it occurred to me that the subjects might have tried to imitate the CV syllables when responding to the stimuli. A preliminary examination of the data indicated that this seemingly unconscious action was indeed a possibility and was not limited to only a few subjects during the experiments. It was observed for almost all subjects during the entire course (learning session, identification and dissimilarity judgment task) of the experiments. The question raised in mind was thus: In identifying and judging the dissimilarity between two arbitrary stimuli were listener responses a mixed product of the acoustic differences between the syllables on the one hand and their articulatory differences on the other?

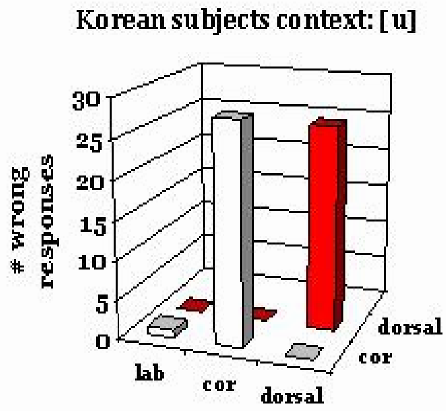
This possibility made me consider ways of quantifying a hypothetical articulatory effect on the results of the experiments. I decided to develop a measure of “articulatory distance” to be used along with the previously derived estimate of how different they sound (see previous chapter).

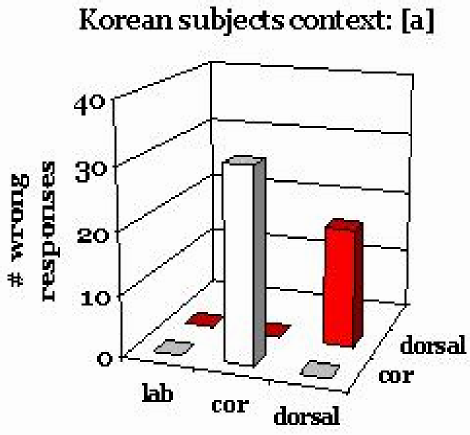
The results of the identification tests (Figure 4.1) show that most of the confusions were made within a given articulatory category (i.e. coronal with coronal, dorsal with dorsal). For example, when a subject heard a dental consonant, he confused it mostly with an alveolar or retroflex but rarely with a palatal, velar, or uvular consonant. This tendency was found in every language group and every vowel context. An exception was the treatment of the /i/ vowel context by the English language group. These subjects confused coronals with bilabials as often as with other coronals.

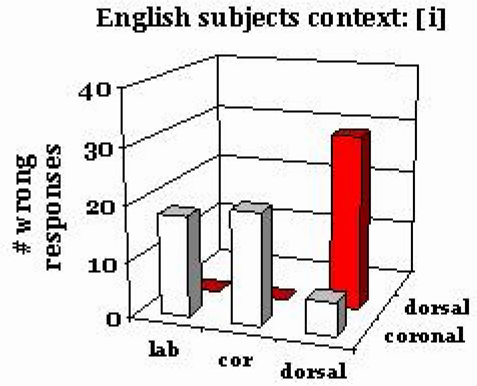

Korean subjects context: [i]

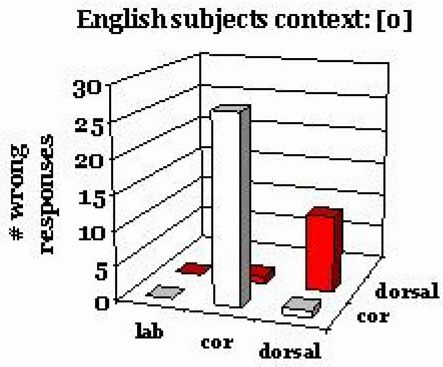

Korean subjects context: [o]

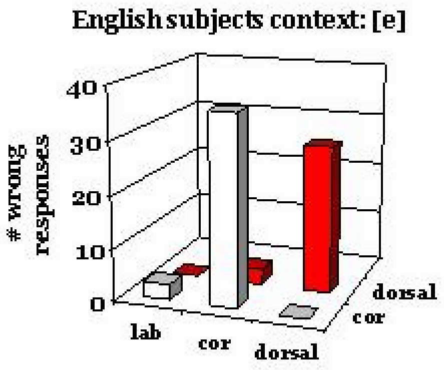

Korean subjects context: [e]

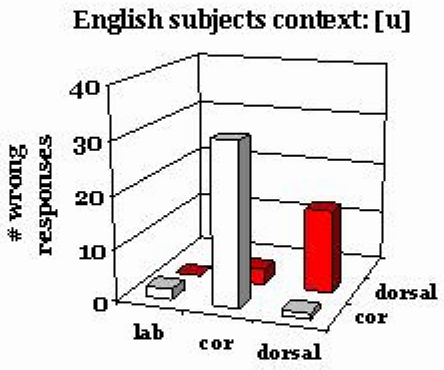

Korean subjects context: [u]

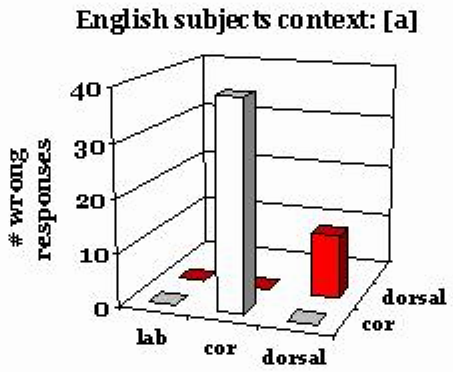

Korean subjects context: [a]

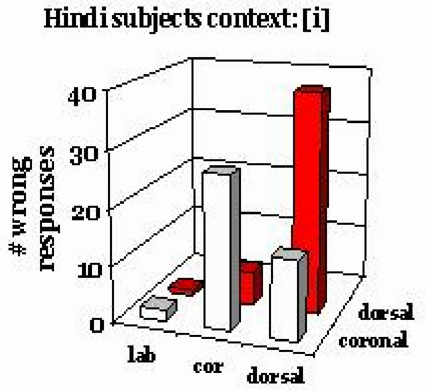

English subjects context: [i]
# wrong responses
40
30
20
10
0
lab
cor
dorsal
dorsal
coronal

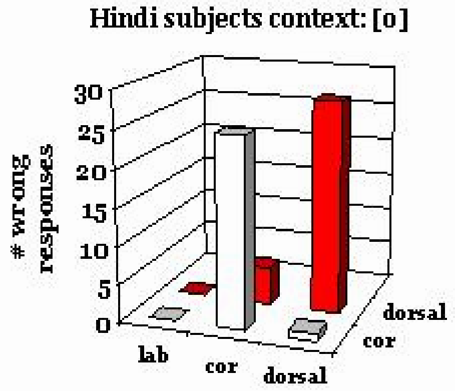

English subjects context: [o]
# wrong responses
30
25
20
15
10
5
0
lab
cor
dorsal
dorsal
cor

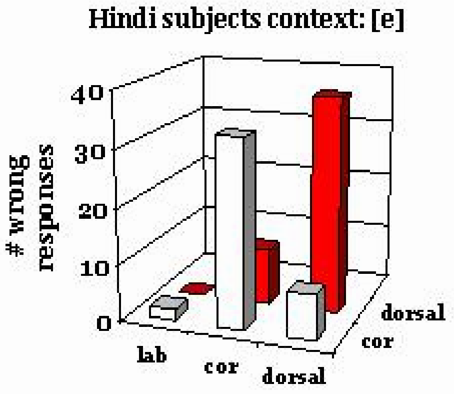

English subjects context: [e]
# wrong responses
40
30
20
10
0
lab
cor
dorsal
dorsal
cor

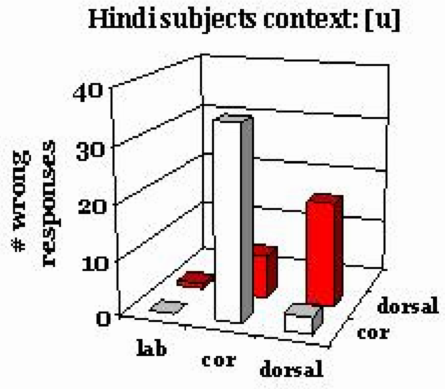

English subjects context: [u]
# wrong responses
40
30
20
10
0
lab
cor
dorsal
dorsal
cor

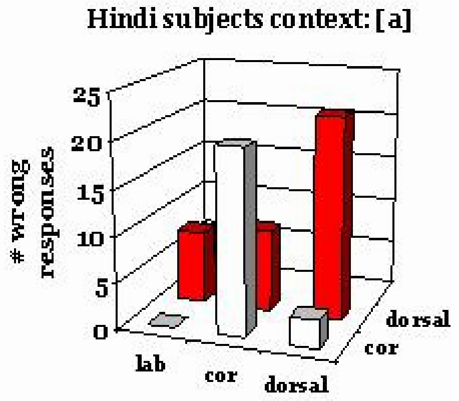

English subjects context: [a]
# wrong responses
40
30
20
10
0
lab
cor
dorsal
dorsal
cor

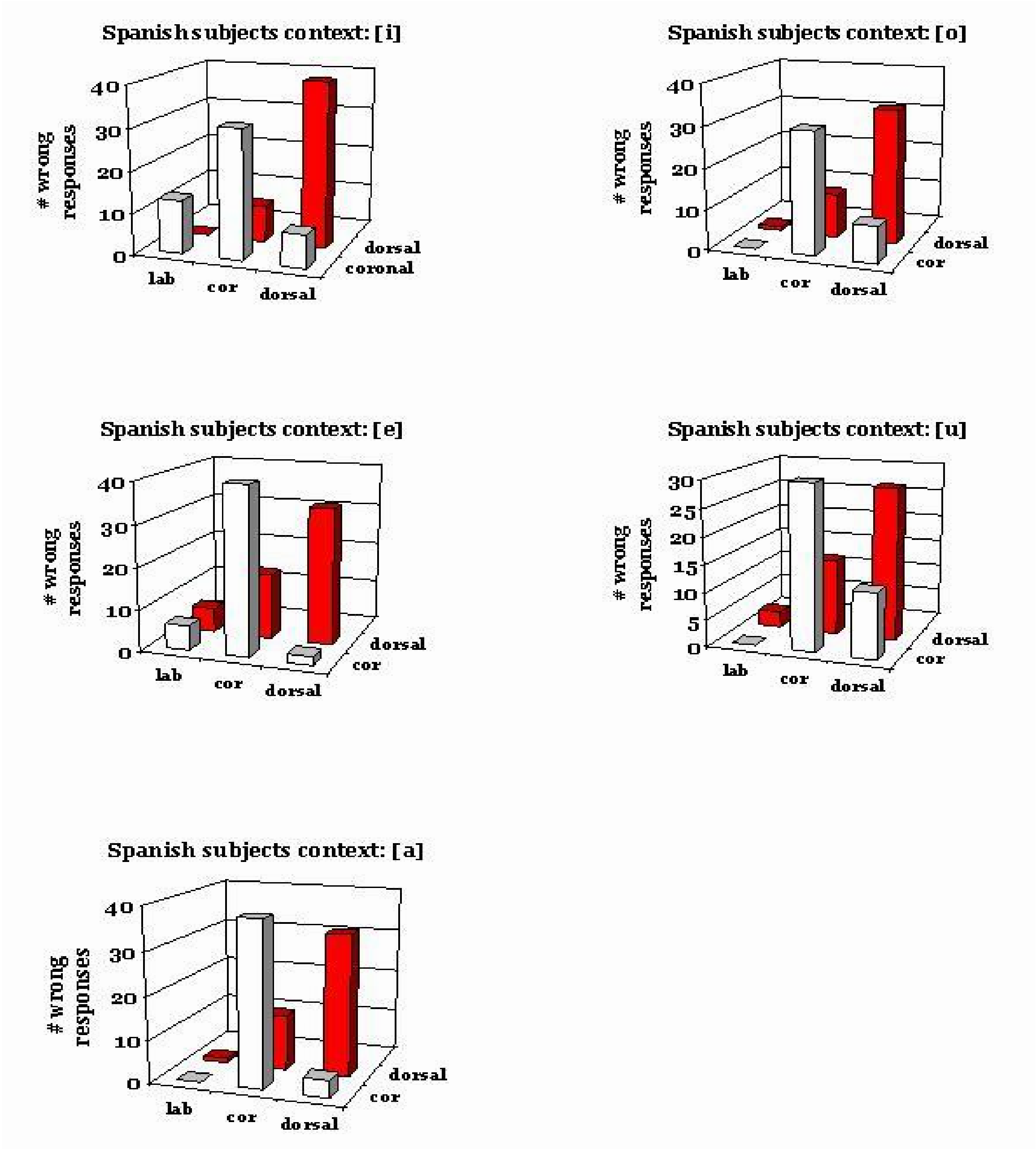

Hindi subjects context: [i]
# wrong responses
0
10
20
30
40
lab
cor
dorsal
dorsal
coronal


Hindi subjects context: [o]
# wrong responses
0
5
10
15
20
25
30
lab
cor
dorsal
dorsal
cor


Hindi subjects context: [e]
# wrong responses
0
10
20
30
40
lab
cor
dorsal
dorsal
cor


Hindi subjects context: [u]
# wrong responses
0
10
20
30
40
lab
cor
dorsal
dorsal
cor


Hindi subjects context: [a]
# wrong responses
0
5
10
15
20
25
lab
cor
dorsal
dorsal
cor


Spanish subjects context: [i]

Spanish subjects context: [o]

Spanish subjects context: [e]

Spanish subjects context: [u]

Spanish subjects context: [a]

Figure 4.1 3-dimensional confusion matrices obtained for identification task

A given panel refers to a given vowel context and a specific language group.

Y-axis = number of confusions (wrong responses) summed and grouped into three (rather than seven) places of articulation: labial, coronal and dorsal;

X-axis = response category;

Z-axis = stimulus category.

Figure 4.1 suggests that there is a strong tendency for coronals to be confused with coronals and for dorsals to be confused with dorsals. This finding provides a clear indication that articulatory factors should be considered in predicting the perceptual performance of the subjects in this experiment; in the following I try to capture this articulatory effect in terms of the notion of articulatory distance.

4.2 Development of a measure of articulatory distance

4.2.1 X-ray data on CV:s

X-ray data for a set of CV:s comparable to the present stimuli were based on experiments and analyses done in connection with the previously mentioned APEX project[1]. The present attempt is based on the corpus that I used earlier in deriving F2 and F3 onsets (see section 2.2.1.4), namely the tracings of 400 tongue

[1] X-ray films were made in a collaborative effort between the Phonetics Laboratory at Stockholm University and the Dept of Radiology at Danderyd Hospital, Stockholm (Branderud et al. 1998).

shapes obtained from X-ray lateral profiles and the numerical specification of those shapes in terms of Principal Components (Lindblom, 2003).

Also relevant is some work on quantifying "articulatory cost" in terms of "deviation from neutral" (Lindblom, 2007). In Lindblom's specificational scheme (2003, 2007) each individual tongue contour is described in terms of the x and y coordinates of 25 'fleshpoints'. In deriving his "articulatory cost" measure Lindblom first drew straight lines between identically numbered fleshpoints. Then the "articulatory cost" of moving the tongue between its rest position and a specific target articulation was taken to be proportional to the articulatory distance between these two configurations. Articulatory distance between two arbitrary contours a(x) and b(x) was defined as a 'root mean square' distance (calibrated in mm) between them:

$$\text{dist (a,b)} = \sqrt{(\sum_{1}^{25}[a(x)-b(x)]^2)/25} \qquad \text{(Equation 4-1)}$$

4.2.2 Articulatory distance defined as an RMS difference between articulatory contours

The RMS metric was used to calculate the distances between seven articulatory configurations corresponding to the present places of articulation, one for each stop. Using the numerical specifications of bilabial, dental, alveolar, retroflex, palatal, velar and uvular constrictions in Lindblom (2007), I derived the distance matrix shown in Table 4.1.

	Bilabial	Dental	Alveolar	Retroflex	Palatal	Velar	Uvular
Bilabial	0						
Dental	12.6	0					
Alveolar	22.5	9.9	0				
Retroflex	33.9	21.2	11.4	0			
Palatal	52.1	39.5	29.6	18.2	0		
Velar	71.4	58.8	49.0	37.6	19.4	0	
Uvular	110.8	98.2	88.3	76.9	58.7	39.4	0

Table 4.1 Articulatory distances among seven stop articulations

The numbers in the matrix represent an RMS differences between articulatory contours (in mm).

A maximum difference of 110.8 mm is observed between bilabial and uvular stops. The minimum distance occurs between the dental and alveolar stops with 9.9 mm.

4.3 Results

Multiple regression analyses were next performed with one dependent variable (perceptual dissimilarity) and six independent variables: formant-based distance (in MEL), spectral distance at the burst, the time constants of F1, F2 and F3 and articulatory distance.

Regression Statistics	
Multiple R	0.88
R Square	**0.77**
Adjusted R Square	0.77
Standard Error	0.26
Observations	245

ANOVA

	df	SS	MS	F	Significance F
Regression	6	56.99	9.50	136.60	0.00
Residual	238	16.55	0.07		
Total	244	73.54			

Table 4.2 Results of multiple regression analysis with formant distance, time constant, burst spectrum distance and articulatory distance as predictors (vowels pooled)

The results of multiple regression analyses indicate a score of R=.88, and R^2=.77 for the pooled data (Table 4.2). This suggests that articulatory distance does make a significant contribution in predicting the perceptual distance judgments.

Comparing the regression results between the LN[1-(Perceptual dissimilarity/6)] and the formant-based (Mel) distance (Table 4.2) , we find that the multiple regression with six independent variables increased the predictability of perceptual judgments from R^2=0.52 to 0.77.

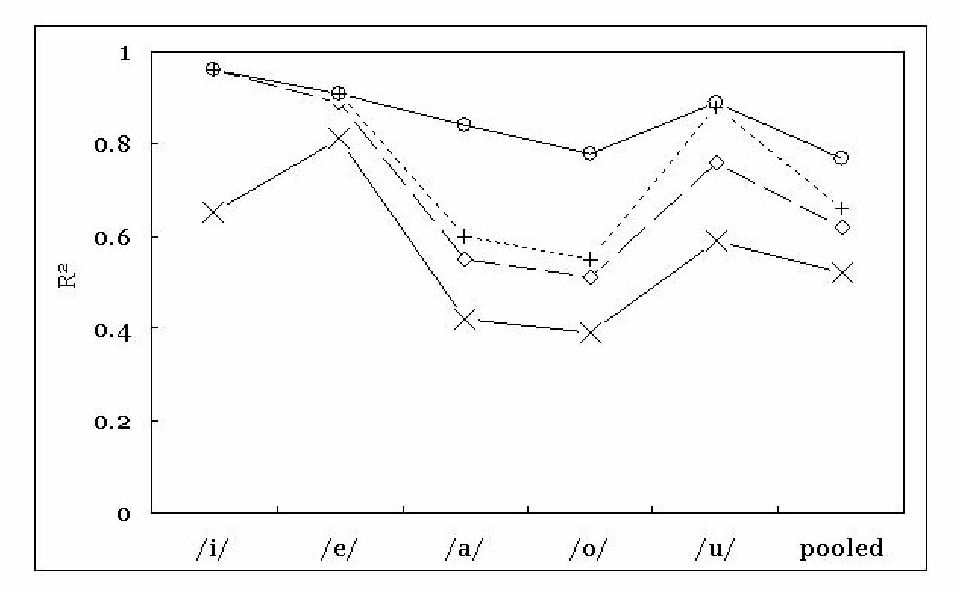


Figure 4.2 R² values with different predictor variables for the different vowel contexts

Solid line with crosses: formant distance

Broken line: formant distance and the time constant

Dotted lines: formant distance, the time constant and the burst distance

Solid line with circles: formant distance, the time constant, the burst distance and the articulatory distance

It is noteworthy that in the front vowel contexts (i.e. before /i/ and /e/), formant based distances and the time constant distances can predict most of the perceptual responses (Figure 4.2). Before the /i/ vowel, the above-mentioned two predictor variables explain 96% of the perceptual responses and before /e/, they predict 89% of the perceptual responses.

It is also striking that the articulatory distance improves the predictability for the pooled data, especially in the back vowel contexts. In the pooled data the inclusion of articulatory distance improves the R^2 by 0.11, before /a/ by 0.24 and before /o/ by 0.23. This finding provides strong positive evidence for an imitative component in the listener responses. We might speculate that the mirror neurons discussed in section 4.1 form part of the mechanism underlying this behavior. In the identification test and the dissimilarity task the response time was limited to 3 seconds which is a very short time for subjects to intentionally or consciously imitate the stimuli. However, after the experiment the subjects reported that they felt the response time was too short and that they were in a hurry to make responses, which suggests that if they did indeed try to imitate the CV's, that process must have taken place subconsciously. Confirming the direct relationship of the present results and the function of mirror neurons is beyond the scope of this research and it must be left to further study.

Chapter 5: Preferred Consonant Place in Each Vowel Context

5.1 Simulating preferred vowel systems: the Liljencrants and Lindblom approach

It has been shown in the current study (Chapter 3) that with formant distance, the time constants for F1, F2 and F3, and burst spectra distance as the predictors of the regression analyses, the multiple regression analysis yielded reasonable results (Figure 5.1).

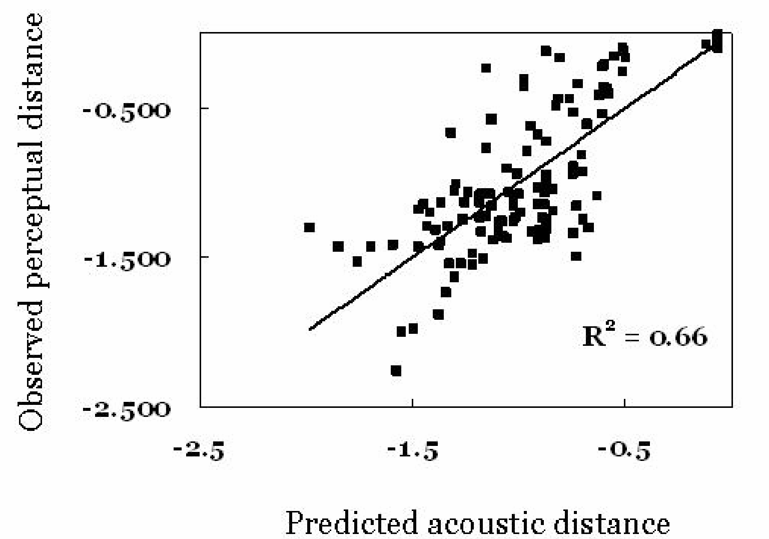


Figure 5.1 Predicted acoustic distance (derived from Multiple regression analyses using locus distances, time constant differences, burst spectra differences) plotted against the perceptual distance judgements coded as LN(1-perceptual distance/6).

R^2 is 0.66 for pooled data (0.96 for /i/, 0 .91 for /e/ 0.6 for /a/, 0.55 for /o/ and 0.88 for /u/), which means that 66% of the data can be explained by the independent variables (Figure 3.3 repeated here as Figure 5.2).

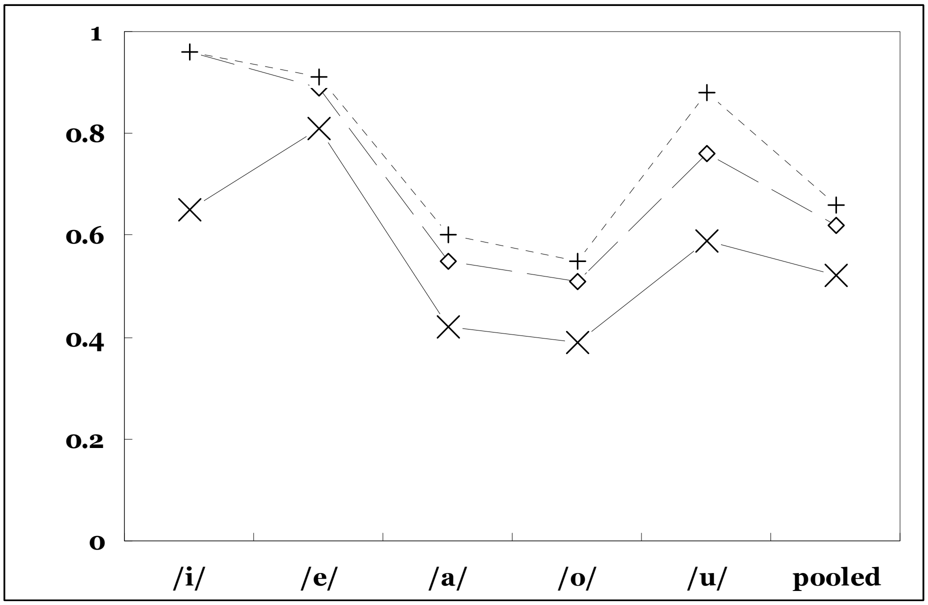


Figure 5.2 R^2 values with different predictor variables for the different vowel contexts

Solid line with crosses: formant distance

Broken line: formant distance and the time constant

Dotted lines: formant distance, the time constant and the burst distance

It has also been shown that the time constants for F1, F2 and F3 successfully described formant dynamics (section 2.2.2) and the burst spectra show somewhat systematic spectral effects in the present set of speech samples including seven places of articulation at five vowel environments (section 2.2.3). Thus it is safe to say that perceptual distance can be successfully predicted by the acoustic distances including F-patterns, the time constants for F1, F2, and F3, and the burst spectra distances.

Now I need to answer the second question:

> *If consonant (CV) systems were seen as adaptations to a demand for perceptual contrast, what would these systems be like?*

For vowel systems, Liljencrants and Lindblom (1972) performed a numerical simulation by using the criterion of maximizing perceptual contrast. Many researchers thereafter explored the role of speech perception in phonology (Flemming (1995), Hume and Johnson (2001) and Ten Bosch (1991) among others).

The simulation of Liljencrants and Lindblom (1972) was quite successful in predicting vowel system when the number of vowels in the system is less than seven. When the number of vowels is seven or more, the problem of "too many high vowels” occurred. It revealed that the open-close (sonority) dimension is preferred than the front-back (chromaticity) dimension in selecting vowels. To remedy this problem Lindblom (1986) adopted a measure of auditory distance based on whole spectra rather than formant frequencies. Diehl, Lindblom and Creeger (2003) and Lindblom, Diehl and Creeger (2006) took more step to improve auditory realism by introducing the notion of Dominant Frequency, a measure derives from the zero-crossing frequencies observed at the output of auditory filters.

However, the limitation of the researches is that they considered only steady-state vowels which are very rare in natural speech.

In the present work an attempt is made to expand the Liljencrants and Lindblom (1972) to CV syllables since it is desirable to have a more general measure which could be applied also to time-varying patterns CV syllables.

The criterion that Liljencrants and Lindblom (1972) used in their approach to find vowels in the perceptual vowel space can be written as follows:

$$\sum_{i}^{m} 1/r_i^2 \rightarrow \text{minimized} \qquad \text{(Equation 5-1)}$$

Where r refers to the distance between the ith pair of vowels, and the number of pairs per system is $m = n(n\text{-}1)/2$ where n is equal to the number of vowels in the system.

I will address the issue of preferred consonant place in two ways. First, I will calculate the distinctiveness in terms of the sum of dissimilarity judgment responses for each CV syllable and derive predictions about favored place inventories by rank ordering those sums. Secondly, I will replicate the Liljencrants and Lindblom simulation method and apply it to the present perceptual dissimilarity results.

5.2 Preferred consonant places based on maximal distinctiveness

The simplest way to predict the preferred consonant place in a given consonant space would be to calculate the sum of the distances between a given CV and those of the other CVs, and then compare this sum with the corresponding sums calculated for the other CVs. For example, for [bi] there are six distance numbers because it can be compared with six other Ci syllables. The sum of those six numbers can be taken to represent a measure of how different [bi] sounded from all the other Cis. Such distance sums were derived for each of the seven consonants in each vowel context. Compared with the method of Liljencrants and Lindblom this method is focused more on the degree of perceptual distinctiveness of each individual CV rather than on finding an optimal 'system' of places. The results of this simulation will then be compared with the simulation method of Liljencrants and Lindblom in the next section.

The matrices that I obtained as a result of the dissimilarity judgment task are triangular because I performed symmetrization of the matrices. For example, the responses for the [bi]-[di] pair and the [di]-[bi] pair were averaged (2.77) and placed in a single cell (i.e. the fourth cell in the first column) while the other cell (i.e. the second cell in third column) was left unfilled as illustrated in Table 5.1.

	bi	d̪i	di
bi	0.34		
d̪i	2.50	0.45	
di	2.77	1.85	0.36

Table 5.1 Example of triangular perceptual distance matrix

My first step was to turn the triangular matrix into a quadratic one by means of the copying procedure demonstrated in Tabl2 5.2.

	bi	d̪i	di
bi	0.34	2.50	2.77
d̪i	2.50	0.45	1.85
di	2.77	1.85	0.36

Table 5.2 Example of square perceptual distance matrix

Next, I calculated the sum of each row (Table 5.3). The sum scores are then compared: a large sum score means that the CV is highly distinctive, a small sum score indicates the opposite.

	bi	d̪i	di	ɖi	ɟi	gi	ɢi	Sum
bi	0.34	2.50	2.77	2.16	4.08	4.94	5.37	22.14
d̪i	2.50	0.45	1.85	2.02	3.95	4.71	5.17	20.63
di	2.77	1.85	0.36	0.61	4.06	4.56	5.19	19.39
ɖi	2.16	2.02	0.61	0.31	4.25	4.72	5.09	19.16
ɟi	4.08	3.95	4.06	4.25	0.08	4.32	4.83	25.56
gi	4.94	4.71	4.56	4.72	4.32	0.19	2.91	26.33
ɢi	5.37	5.17	5.19	5.09	4.83	2.91	0.19	28.74

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ	Sum
bɛ	0.24	2.13	2.34	2.76	4.65	3.87	3.22	19.20
dɛ	2.13	0.44	0.68	1.38	4.32	3.36	3.29	15.60
dɛ	2.34	0.68	0.40	0.96	4.25	3.67	3.10	15.40
ɖɛ	2.76	1.38	0.96	0.46	4.17	3.63	3.65	16.99
ɟɛ	4.65	4.32	4.25	4.17	0.38	4.10	4.55	26.41
gɛ	3.87	3.36	3.67	3.63	4.10	0.30	2.53	21.45
ɢɛ	3.22	3.29	3.10	3.65	4.55	2.53	0.34	20.66

	ba	d̪a	da	ɖa	ɟa	ga	ɢa	Sum
ba	0.12	3.55	3.70	3.94	4.72	4.49	4.62	25.12
d̪a	3.55	0.33	1.17	1.75	4.50	4.43	4.47	20.18
da	3.70	1.17	0.33	0.96	4.40	4.27	4.45	19.26
ɖa	3.94	1.75	0.96	0.25	4.41	4.29	4.27	19.85
ɟa	4.72	4.50	4.40	4.41	0.14	3.95	4.15	26.25
ga	4.49	4.43	4.27	4.29	3.95	0.17	0.69	22.26
ɢa	4.62	4.47	4.45	4.27	4.15	0.69	0.42	23.06

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ	Sum
bɔ	0.18	3.89	4.12	3.94	4.57	4.66	4.38	25.72
d̪ɔ	3.89	0.35	2.08	2.65	4.26	3.98	4.11	21.31
dɔ	4.12	2.08	0.57	1.26	3.96	4.18	4.41	20.57
ɖɔ	3.94	2.65	1.26	0.34	4.18	3.93	4.50	20.79
ɟɔ	4.57	4.26	3.96	4.18	0.21	4.37	4.33	25.87
gɔ	4.66	3.98	4.18	3.93	4.37	0.18	0.86	22.15
ɢɔ	4.38	4.11	4.41	4.50	4.33	0.86	0.14	22.72

	bu	d̪u	du	ɖu	ɟu	gu	ɢu	Sum
bu	0.12	3.58	3.81	3.92	4.70	4.10	3.80	24.02
d̪u	3.58	0.24	1.24	2.96	4.22	4.38	4.11	20.72
du	3.81	1.24	0.36	1.62	4.50	4.26	4.30	20.08
ɖu	3.92	2.96	1.62	0.37	4.18	4.00	4.07	21.09
ɟu	4.70	4.22	4.50	4.18	0.18	4.57	4.38	26.72
gu	4.10	4.38	4.26	4.00	4.57	0.21	1.83	23.35
ɢu	3.80	4.11	4.30	4.07	4.38	1.83	0.28	22.76

Table 5.3 Rectangular matrices for the predicted acoustic distance

Finally, in each vowel context the sum scores are ranked from largest to smallest to determine the consonant places which give the greatest distinctiveness in each vowel context.

For an inventory of three consonant places, the procedure selects [ɢi, ɟi, gi] in the Ci context, [ɟɛ, gɛ, ɢɛ] in the Ce context, [ɟa, ba, ɢa] in the Ca context, [ɟɔ, bɔ, ɢɔ] in the Cɔ context, and [ɟu, bu, gu] in the Cu context. The palatal and the uvular stops are the most preferred stop consonants whereas the dental, the alveolar, the retroflex stops are never selected (Table 5.4).

Ci	ɢi	ɟi	gi
Cɛ	ɟɛ	gɛ	ɢɛ
Ca	ɟa	ba	ɢa
Cɔ	ɟɔ	bɔ	ɢɔ
Cu	ɟu	bu	gu

Table 5.4 Predicted consonant places in five vowel environments

Inventory size: 3. In other words, the diagram shows the CV syllables that are favored when three stop consonants are chosen.

In every vowel context, the palatal consonant is selected. The second most popular place is the uvular (in the [i, ɛ, a, ɔ] contexts). The bilabial are selected in the [a, ɔ, u] contexts and the velar consonant in the [i, ɛ. u] contexts.. These results show reasonable agreement with the F2 vs. F3 space for the present CV syllables in which the uvular, the retroflex and the palatal consonants are located at the boundary of the space (Figure 2.2_pooled data repeated here as

5.3) and are thus acoustically salient.

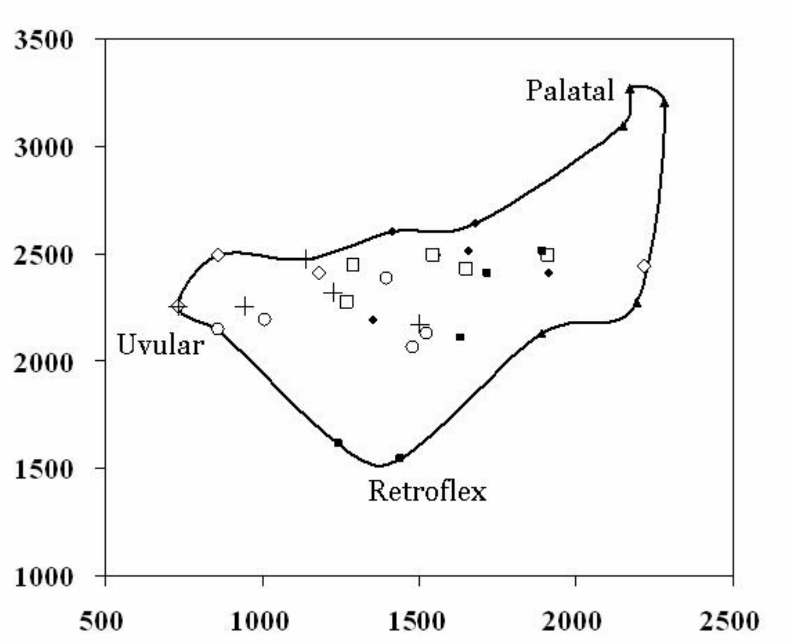


Figure 5.3 F3 vs F2 space at stop release

The outermost points were connected with smoothed curve so that it enclose all measurements

The results are quite different from observed stop place systems. UPSID reports that the most common place system has the bilabial, the dental/the alveolar and the velar consonant (Maddieson, 1984: 32). However, the present simulation predicted that bilabial, dental/alveolar and velar are never selected together when the number of consonants is three.

number of consonants	Preferred CVs				
3	ɢi	gi	ɟi		
4	ɢi	gi	ɟi	bi	
5	ɢi	gi	ɟi	bi	di

Table 5.5 Preferred consonant places predicted by the rank ordering of the individual CV salience in Ci

number of consonant	Preferred CVs				
3	ɟɛ	ɢɛ	bɛ		
4	ɟɛ	ɢɛ	bɛ	gɛ	
5	ɟɛ	ɢɛ	bɛ	gɛ	d̪ɛ

Table 5.6 Preferred consonant places predicted by the rank ordering of the individual CV salience in Cɛ

number of consonants	Preferred CVs				
3	ɟa	ɢa	ba		
4	ɟa	ɢa	ba	ɖa	
5	ɟa	ɢa	ba	ɖa	ga

Table 5.7 Preferred consonant places predicted by the rank ordering of the individual CV salience in Ca

number of consonants	Preferred CVs				
3	ɟɔ	ɖɔ	bɔ		
4	ɟɔ	ɖɔ	bɔ	dɔ	
5	ɟɔ	ɖɔ	bɔ	dɔ	ɢɔ

Table 5.8 Preferred consonant places predicted by the rank ordering of the individual CV salience in Cɔ

number of consonants	Preferred CVs				
3	ɟu	bu	gu		
4	ɟu	bu	gu	ɢu	
5	ɟu	bu	gu	ɢu	ɖu

Table 5.9 Preferred consonant places predicted by the rank ordering of the individual CV salience in Cu

Tables 5.5-5.9 show the consonant places selected when the number of stop places is three, four or five.[2] It is interesting to note that bilabial, dental/alveolar and velar are not selected together even in the four-consonant systems (in every vowel context) or five-consonant systems (in the /i/ and /ɛ/ vowel contexts).

5.3 Replication of the method of Liljencrants and Lindblom

The measure adopted by Liljencrants and Lindblom (1972) optimized systems, that is sets of CV's, rather than individual CV's. This criterion identified vowel contrasts consisting of vowel qualities maximally distant from one another in the

[2] Maddieson (1984: 31) reports that 171 languages have three stop places, 103 languages have four stop places and 35 languages have five stop places. In other words 97.4% out of 317 languages has three, four or five places.

perceptual space. In this section their method is used and the results are compared with my maximal distinctiveness method given in the previous section.

The simulation was performed using the SCHROSYS software (version 1.0) a program developed by Carl Creeger, Dept. of Psychology, The University of Texas at Austin. The input to this program is a triangular matrix with dissimilarity scores, in the present case the data obtained from the dissimilarity judgments. Since the language effect was shown to be minimal, the matrix with data averaged over the language group. The program first asks for the number of consonants in the system, then it performs the pair-wise comparisons of the Liljencrants-Lindblom distance criterion and then computes the optimal set of entities, in this case a set of CV's syllables. In other words, the output of the program is a list of predicted CVs.

Ci	bi	ɖi	ɢi
Cɛ	bɛ	ɟɛ	ɢɛ
Ca	ba	ɖa	ga
Cɔ	bɔ	ɟɔ	ɢɔ
Cu	bu	ɖu	gu

Table 5.10 Predicted consonant places in five vowel environments where the number of consonant is three (predicted by the method of Liljencrants and Lindblom, 1972)

In the three stop consonant system (Table 5.10), the simulation selects [bi, ɖi, gi] in the Ci context, [bɛ, ɟɛ, ɢɛ] in the Cɛ context, [ba, ɖa, ɡa] in the Ca context, [bɔ, ɟɔ, gɔ] in the Cɔ context, and [bu, ɖu, ɡu] in the Cu context. The bilabial consonant is the most preferred (selected in every vowel context) and the retroflex and the uvular stops are selected quite often (retroflex for Ci, Ca, Cu; uvulars for Ci, Cɛ, Cɔ). The dental and the alveolar consonants are never selected in three consonant systems.

number of consonants	Preferred CVs				
3	bi	ɖi	ɢi		
4	bi	di	ɟi	ɢi	
5	bi	d̪i	di	ɟi	ɢi

Table 5.11 Preferred consonant places predicted by the method of Liljencrants and Lindblom (1972) in Ci

number of consonant	Preferred CVs				
3	be	ɟe	ɢe		
4	be	de	ɟe	ɢe	
5	be	de	ɖe	ɟe	ɢe

Table 5.12 Preferred consonant places predicted by the method of Liljencrants and Lindblom (1972) in Cɛ

number of consonants	Preferred CVs				
3	ba	ɖa	ga		
4	ba	da	ɟa	ɢa	
5	ba	da	ɖa	ɟa	ɢa

Table 5.13 Preferred consonant places predicted by the method of Liljencrants and Lindblom (1972) in Ca

number of consonants	Preferred CVs				
3	bɔ	ɟɔ	ɢɔ		
4	bɔ	dɔ	ɟɔ	ɢɔ	
5	bɔ	dɔ	ɖɔ	ɟɔ	ɢɔ

Table 5.14 Preferred consonant places predicted by the method of Liljencrants and Lindblom (1972) in Cɔ

number of consonants	Preferred CVs				
3	bu	ɖu	gu		
4	bu	du	ɟu	ɢu	
5	bu	du	ɖu	ɟu	ɢu

Table 5.15 Preferred consonant places predicted by the method of Liljencrants and Lindblom (1972) in Cu

Tables 5.11-5.15 show the consonant place simulation in vowel contexts. Even though the palatal consonant is not selected in the three- consonant system, it occurs in every vowel context in the predicted four-consonant systems. Furthermore, every vowel context is predicted to select bilabial, alveolar, palatal and uvular consonant when the number of consonants is set at four. In five consonant systems, every vowel context (except for /i/ vowel context) chooses bilabial, alveolar, retroflex, palatal and uvular consonant. These results are in line with the maximal distinctiveness simulation in the previous section except that the retroflex consonant was found to be popular in this approach.

In general, the two simulations are similar in that they both prefer the auditorily salient CVs (e.g. palatals, uvulars) to more confusing but more frequently attested CVs (e.g. alveolars and velars). In the three-consonant system, the common combination of bilabial, dental/alveolar and velar is never predicted in this simulation, either. From these observations, it can be safely said that results of the replication of Liljencrants and Lindblom (1972) confirm the results based on rank ordering of the salience of individual syllables in the previous section.

The discrepancy between the predictions and real language data suggests that distinctiveness, as defined in the present work, cannot alone correctly predict the favored typological patterns of consonant place inventories. One possible explanation of this discrepancy may be articulatory: It may be that some gestures are easier to pronounce than others for physiological reasons. Consequently *both* ease of articulation *and* auditory distinctiveness may influence the structure of phonetic inventories (Martinet, 1964; Lindblom, 1990). However at the present moment it is not at all clear how 'ease of articulation' can be rigorously defined or represented, or how much weight each of these two aspects of speech (auditory distinctiveness and ease of articulation) should be given in determining consonant inventories. More detailed investigations on the concept of ease of articulation are called for.

Chapter 6: Summary and conclusions

The focus of this thesis is on the role of perceptual distinctiveness in consonant inventories. While distinctiveness appears to play a role in the shaping of vowel systems, a literature review indicates that its status in consonant selections remains unclear.

To address this issue I used speech materials recorded by a trained phonetician containing samples of CV syllables whose initial consonant was a voiced stop produced at seven places of articulation (bilabial, dental, alveolar, retroflex, palatal, velar and uvular) and whose vowel was drawn from five vowels: [i] [ɛ] [a] [ɔ] and [u]. 35 syllables were selected from these recordings.

Detailed acoustic measurements were performed: formant patterns at vowel onsets (loci) and vowel mid-points, transitions rates and burst spectra. To validate the speech material comparisons were made with published data and with formant frequencies derived by means of an articulatory model. Although the 35 syllables cannot be said to be representative of any one language they were found to have acoustic properties compatible both with published and theoretically expected data.

Perceptual data were collected on these 35 syllables. The role of the listeners' native language was examined by using four groups of subjects with Korean, English, Hindi and Spanish as their mother tongue. The listening tasks included a training session, an identification test and a discrimination task (judging how dissimilar two syllables sounded). Comparisons of the language groups indicated small differences between response patterns. Pooling the data therefore seemed justified.

A third part of the work consisted of developing a measure of 'acoustic distance' to be used in an attempt to predict the listener responses. Three types of acoustic distance were calculated for each CV pair in each vowel context: (i) a formant-based distance (based on the differences in formant onsets (loci); (ii) differences in formant transition rates (quantified in terms of the time constants of decaying exponentials fitted to the transitions); (iii) differences between burst spectra (computed using the spectrum-based similarity measure proposed by Plomp (1970)).

When the raw dissimilarity scores of individual contrasts (i.e. the average perceptual dissimilarity assigned by all listeners to syllable *i* and syllable *j*) were plotted against formant-based distances a fairly regular pattern emerged. The data points approximated an upside-down exponential, (1-e$^{(-k^{*}\text{ acoust dist})}$), an observation suggesting coding dissimilarity results as LN[1-Perceptual Dissimilarity/Maximum Score] in regression analyses aimed at investigating how dissimilarities depended on the acoustic variables.

Multiple Regression analyses were performed with the coded dissimilarities as the dependent variable and with (combinations of) formant-based distances, time constant differences and burst differences as the independent variables. A reasonably high correlation was observed for formant-based distances as the only independent variable both for individual vowel contexts and the pooled data (R^2=0.52 for pooled data, 0.65 before /i/, 0.81 before /e/, 0.42 before /a/, 0.39 before /o/, and 0.59 before /u/). As second (transition rate differences) and third (burst differences) were added as independent variables, R^2 values increased reaching 0.66 for the pooled data and three independent variables. Conclusion: The acoustic measurements could be successfully used to help explain listener responses.

A fourth step in the analysis of the dissimilarity data was taken as a result of identification scores sometimes deviating systematically from the pattern expected on the basis of acoustic distances: There was a tendency for coronals to be confused with coronals and for dorsals to be confused with dorsals. The question arose whether listener responses had to some extent been influenced by their (subconscious) attempt to imitate the stimuli. This suspicion led to the development of still another independent variable. On the basis of articulatory data on the seven consonant places of articulation a measure of ‘articulatory distance’ was defined. This metric assigned the largest score to the labial-uvular contrast and the smallest value to the dental-alveolar pairs. Adding this variable to the Multi-Regression analyses increased predictability further raising the R^2 score for pooled data to 0.77.

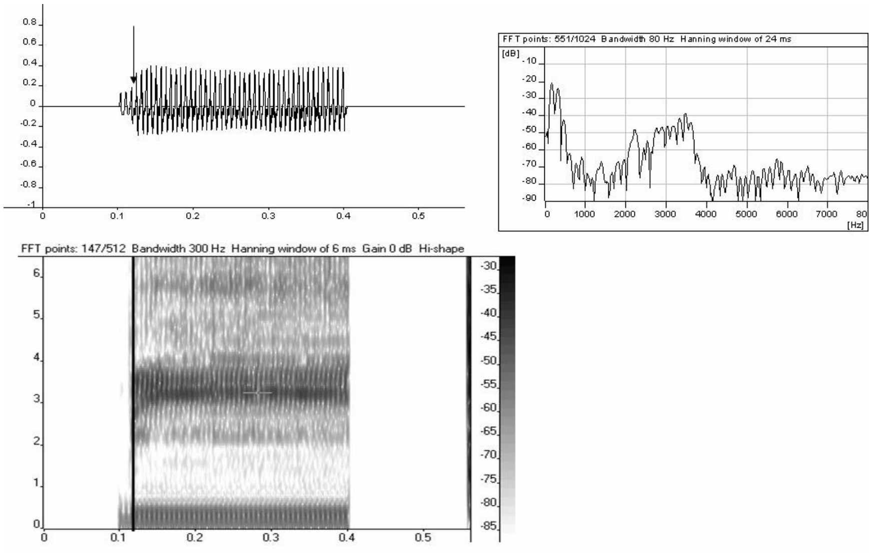
The fifth and final chapter of the thesis returns to the question raised in the introduction: Does an empirically motivated measure of contrast succeed in predicting favored patterns of stop consonant places as well as it has been reported to work for vowel systems? In other words, if distinctiveness explains the widespread preference for [i a u] as the core of vowel inventories, does it also account for the strong typological preference for labial-dental/alveolar-velar as the corner stones in consonant place systems? My answer is based on two methods. The first selected optimal place sets from a rank ordering of the CV syllables with respect to ‘individual salience’ (defined as the sum of a syllable’s perceptual distance to other places in the same vowel context). The second was a replication of the Liljencrants & Lindblom systemic criterion of maximizing distances within all vowel pairs. Both methods failed to produce the typologically prevalent pattern of [b d g]. Instead of these places predictions were found to be vowel-dependent and to often favor CV:s located at the ‘corners’ of the acoustic

F3-F2 space, viz., uvular, palatal and retroflex.

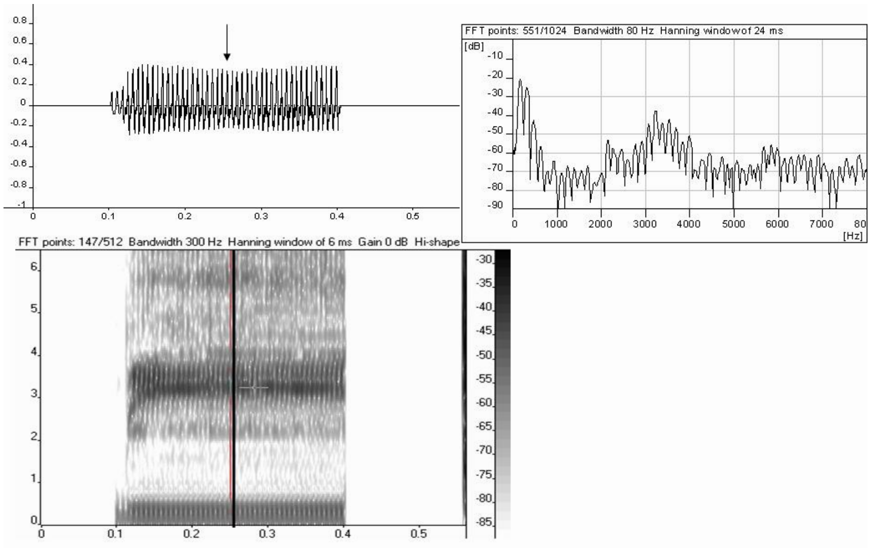
My conclusion is that distinctiveness alone is unlikely to account for how languages use place of articulation in voiced stops. For more successful attempts, future work should be directed towards defining and incorporating production constraints such as ‘articulatory ease’.

Appendix A Formant patterns (i) at CV boundary and (ii) at vowel mid-point for each CV

[bi]

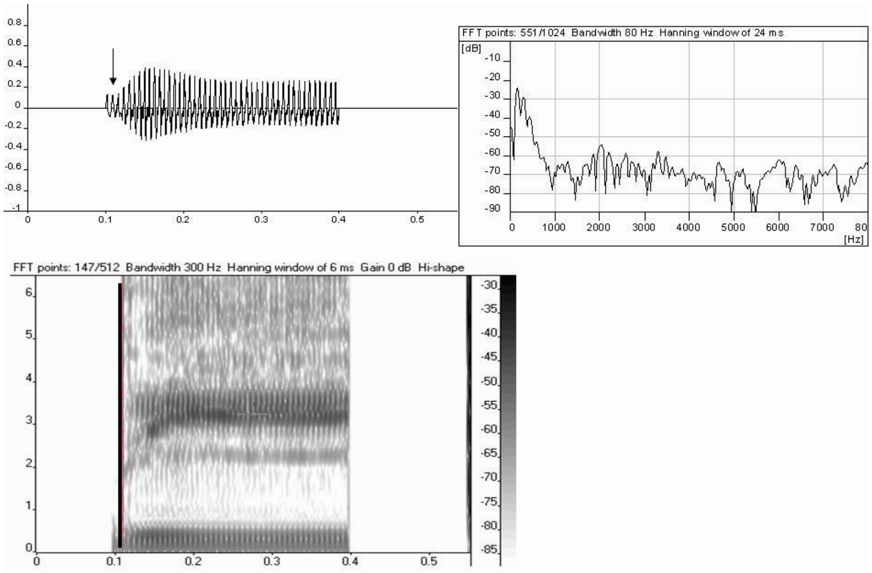


(i)



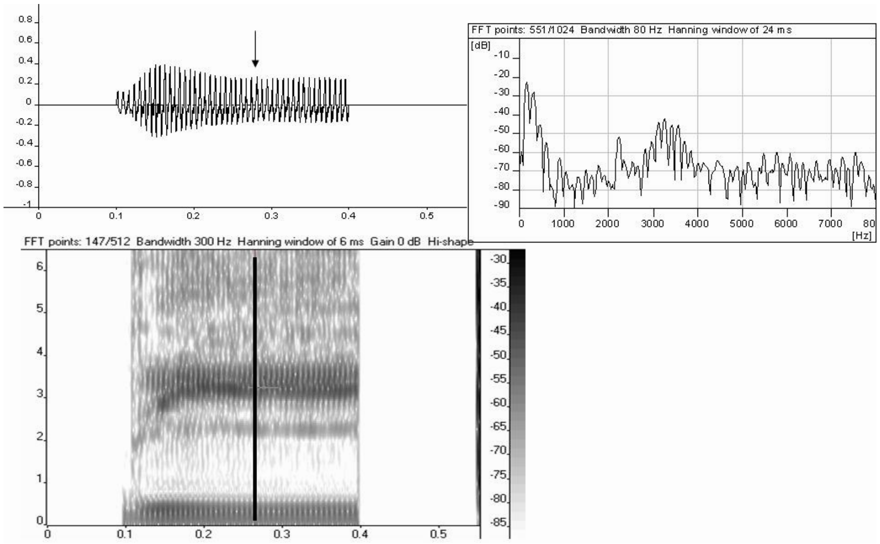
(ii)

[d̪i]



FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

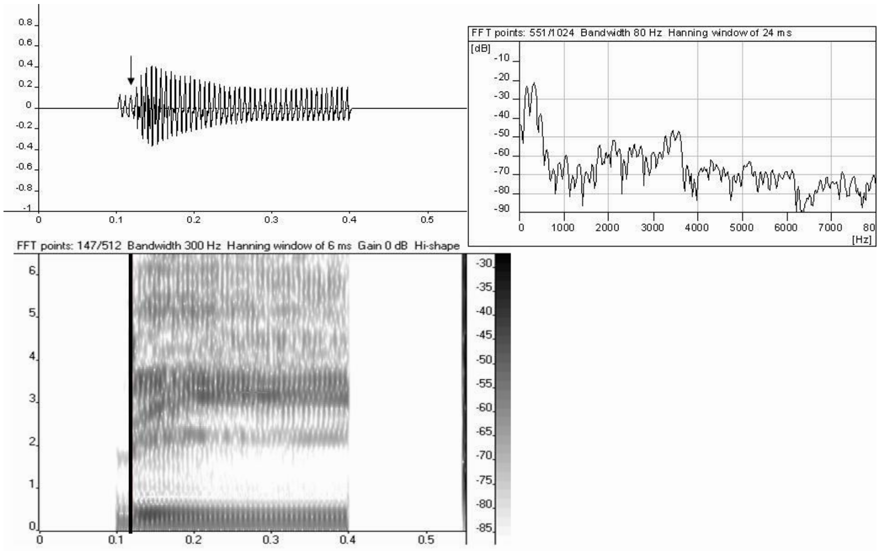
(i)



FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

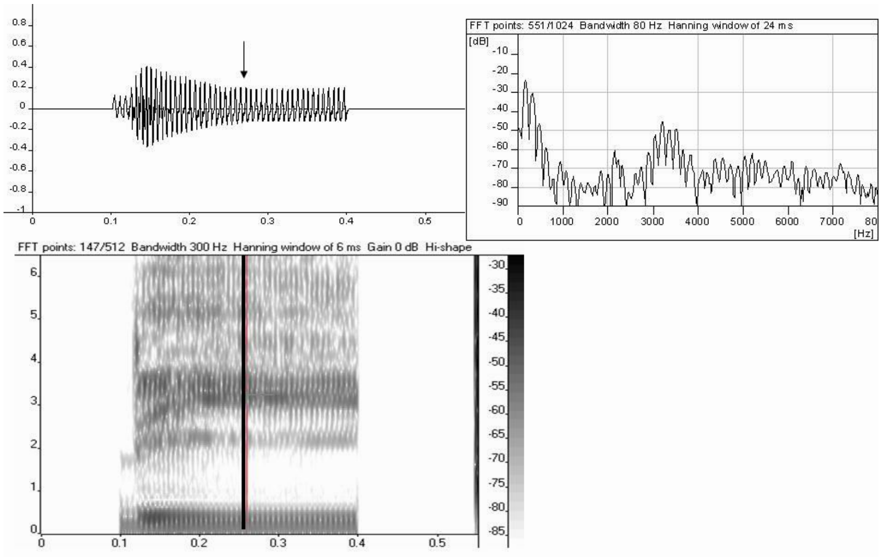
(ii)

[di]



0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

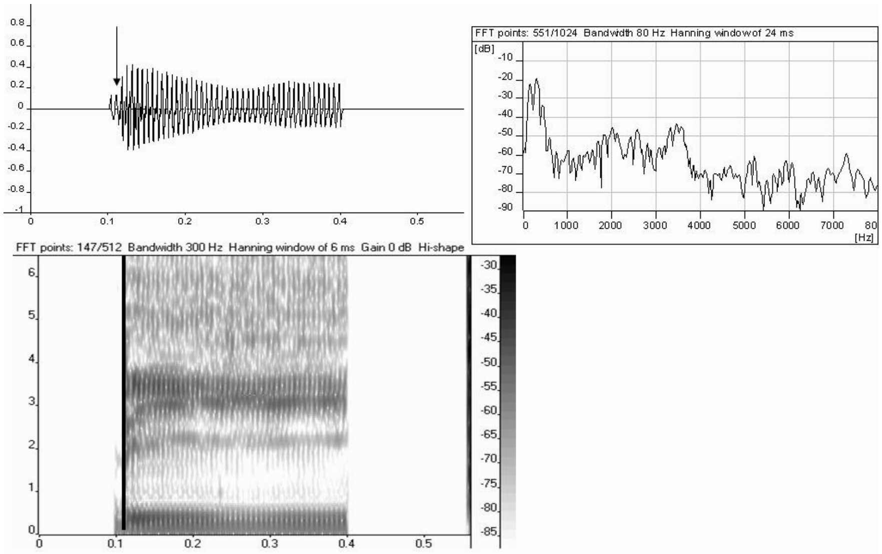
(i)



0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

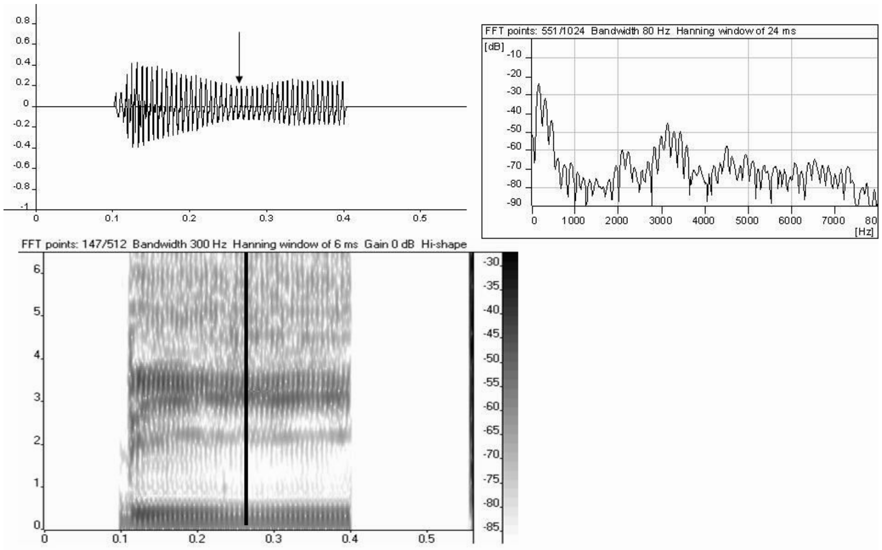
(ii)

[ɖi]



FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0 1000 2000 3000 4000 5000 6000 7000 80
[Hz]
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0 0.1 0.2 0.3 0.4 0.5
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
0 0.1 0.2 0.3 0.4 0.5
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

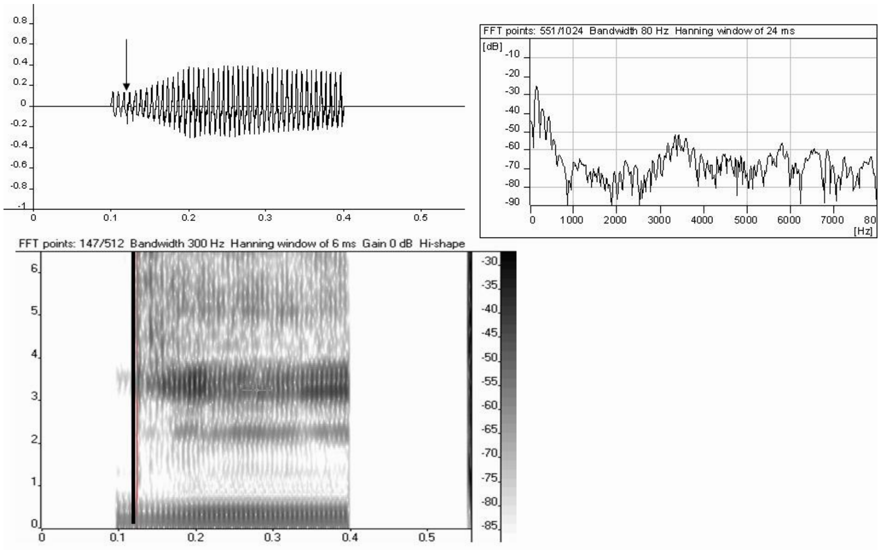
(i)



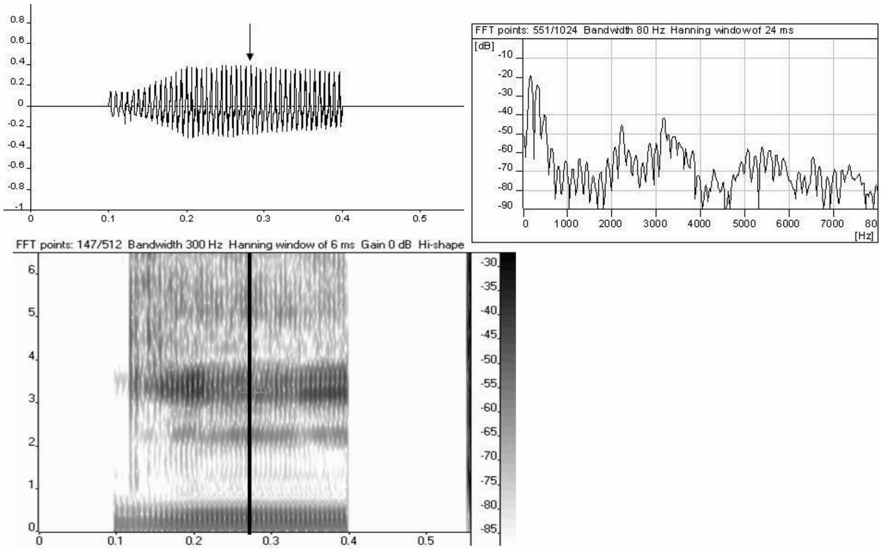
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0 1000 2000 3000 4000 5000 6000 7000 80
[Hz]
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0 0.1 0.2 0.3 0.4 0.5
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
0 0.1 0.2 0.3 0.4 0.5
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(ii)

[ɟi]

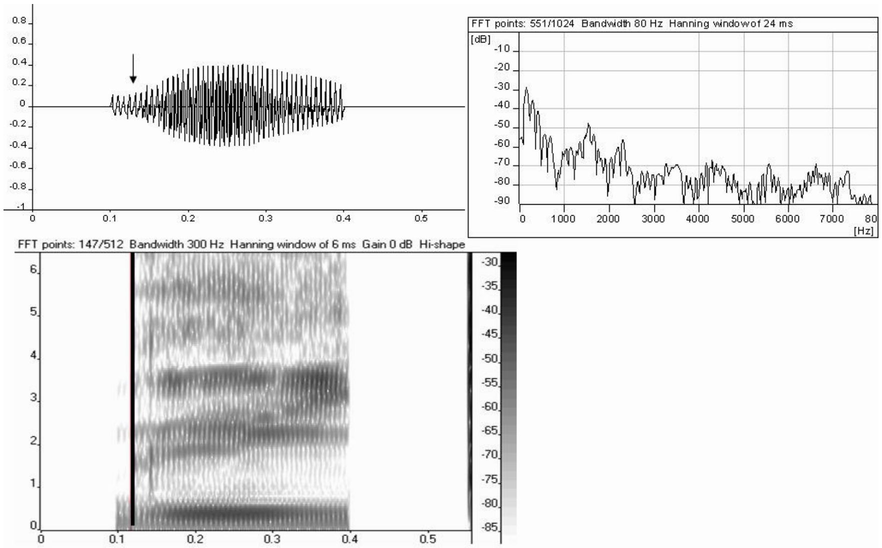

FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0.1
0.2
0.3
0.4
0.5
-10
-20
-30
-40
-50
-60
-70
-80
-90
1000
2000
3000
4000
5000
6000
7000
80
-35
-45
-55
-65
-75
-85

(i)

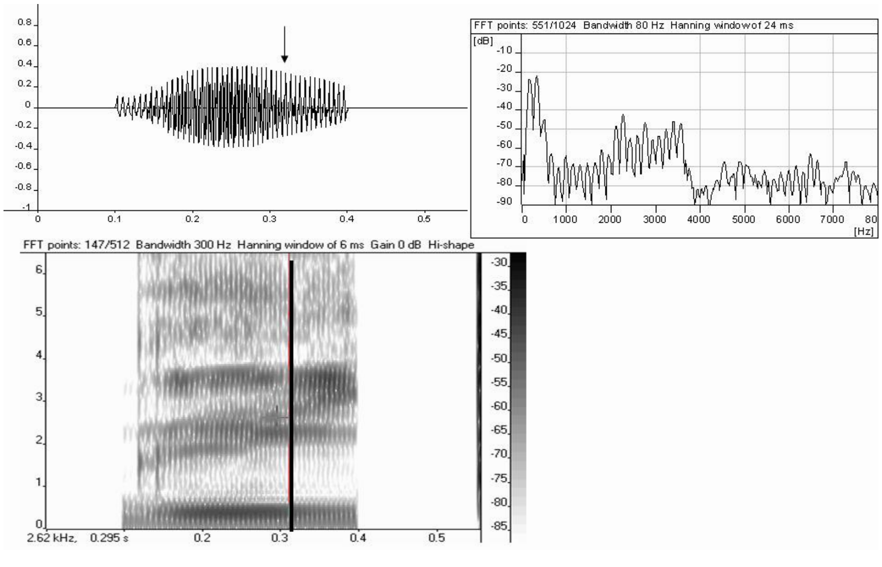

FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0.1
0.2
0.3
0.4
0.5
-10
-20
-30
-40
-50
-60
-70
-80
-90
1000
2000
3000
4000
5000
6000
7000
80
-35
-45
-55
-65
-75
-85

(ii)

[gi]

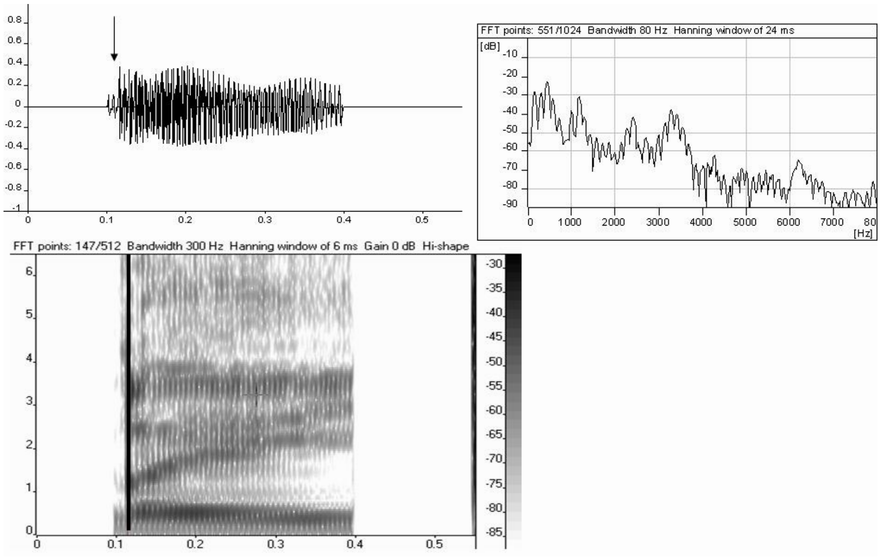

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
0
0.1
0.2
0.3
0.4
0.5
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(i)

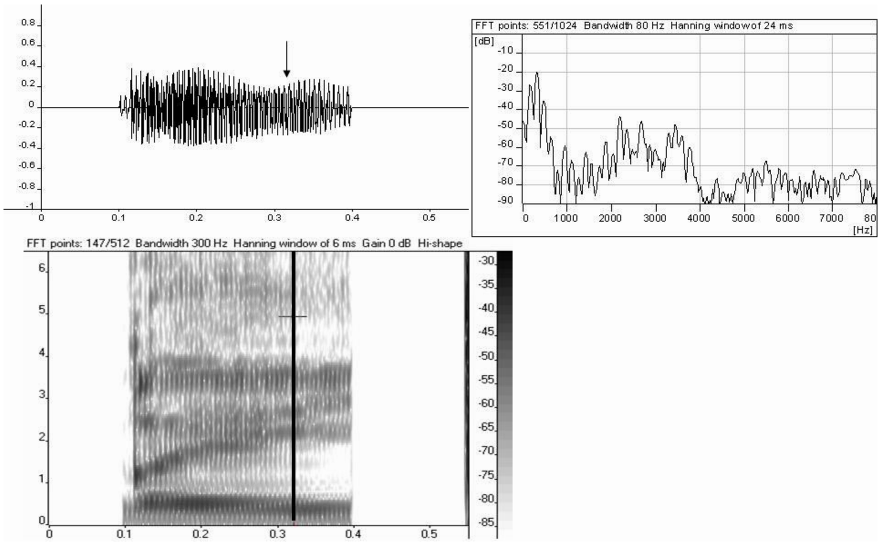

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
2.62 kHz, 0.295 s
0.2
0.3
0.4
0.5
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(ii)

[ɢi]

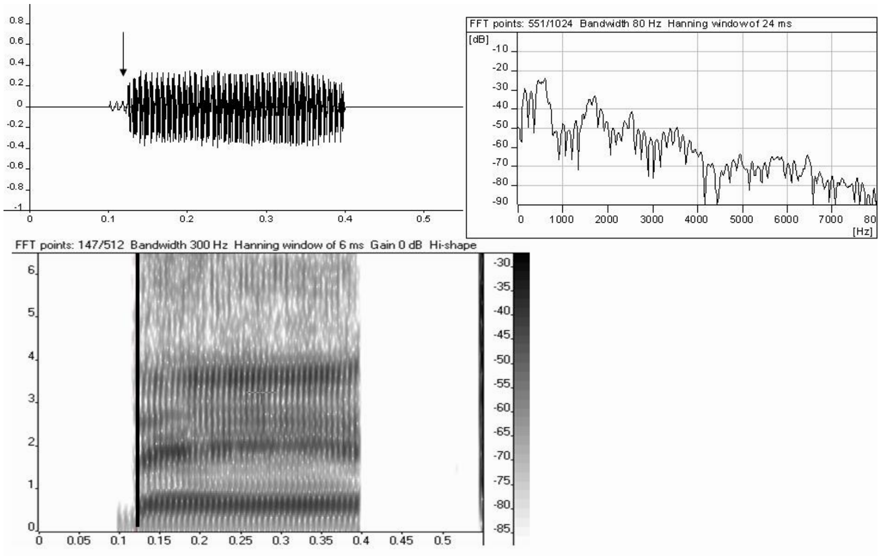

FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(i)

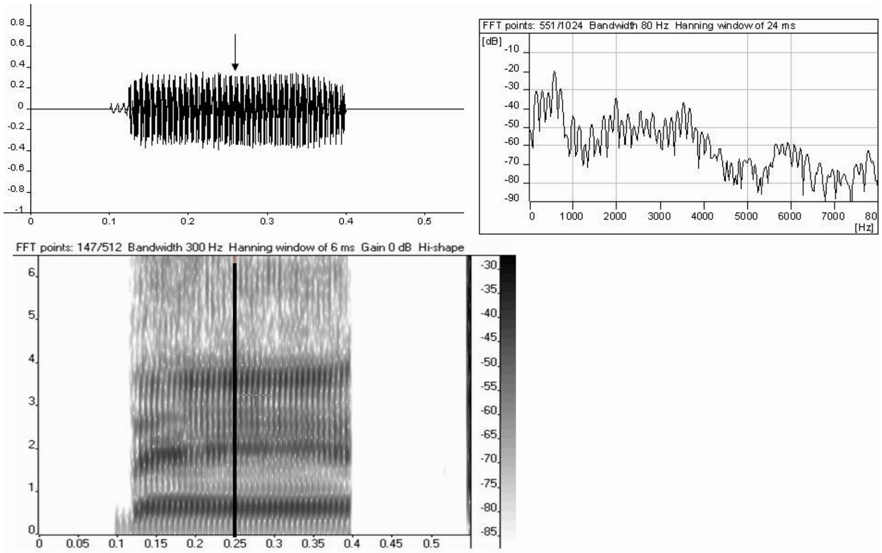

FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(ii)

[be]

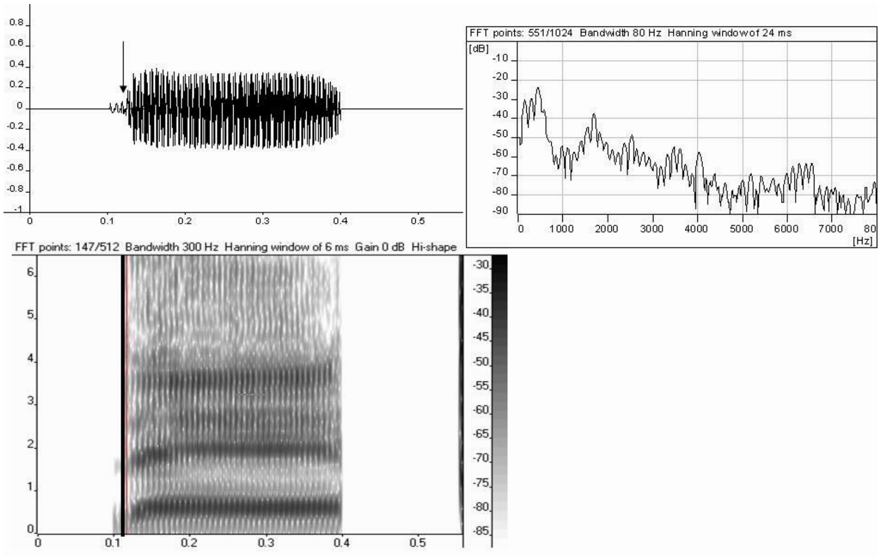

FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(i)

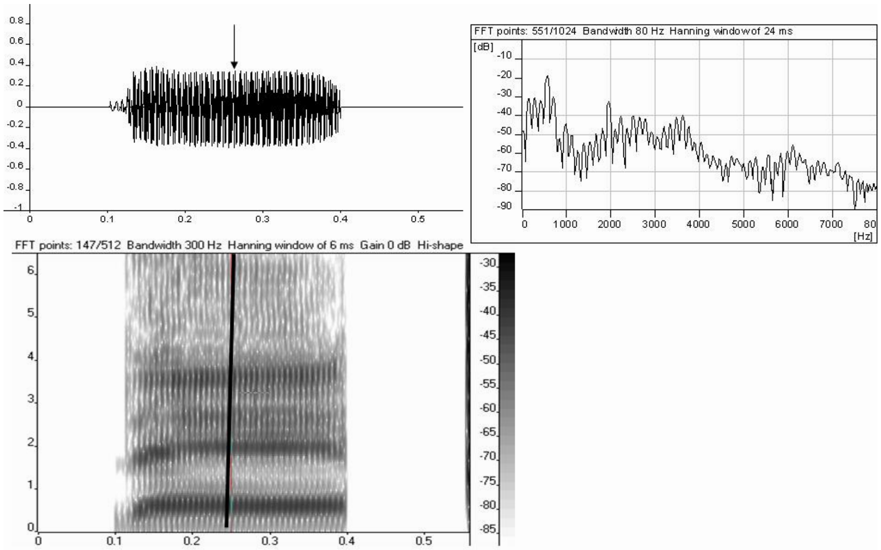

FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(ii)

[d̪e]

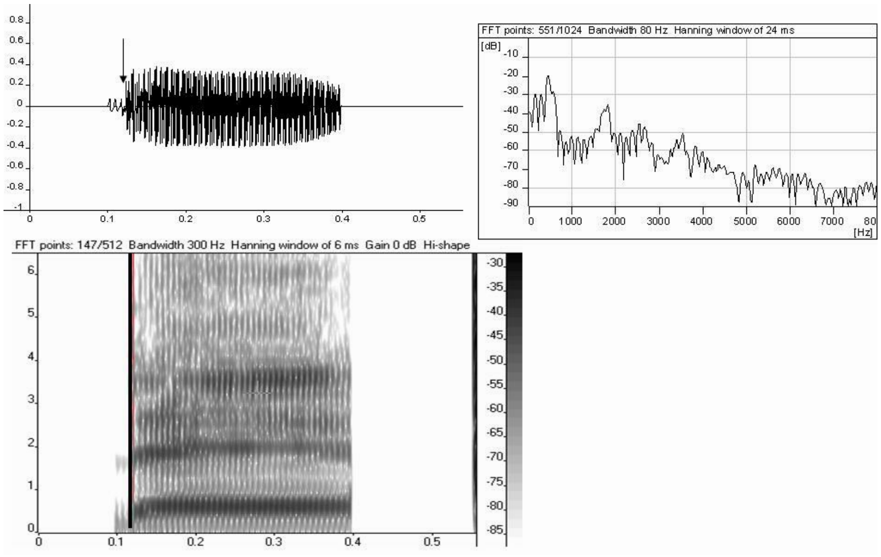

FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(i)


FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

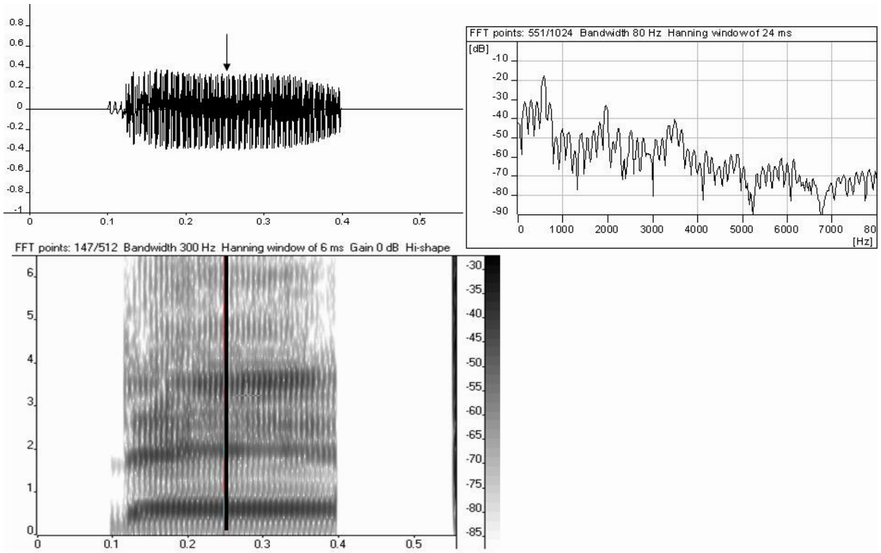
(ii)

[de]


FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms

FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

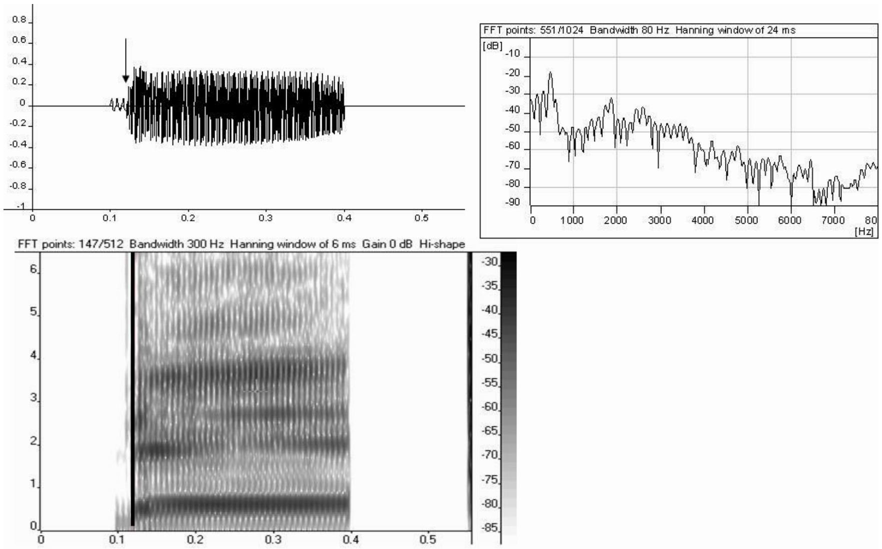
(i)


FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms

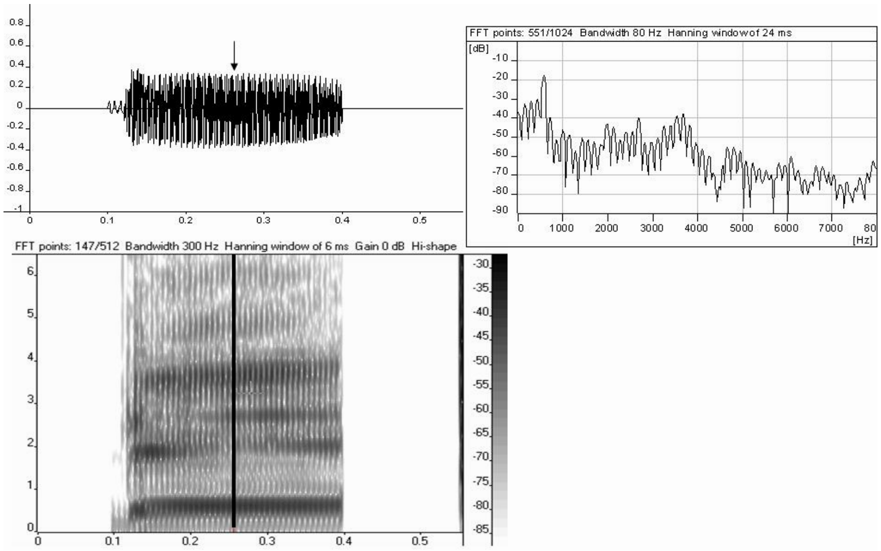
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(ii)

[ɖe]

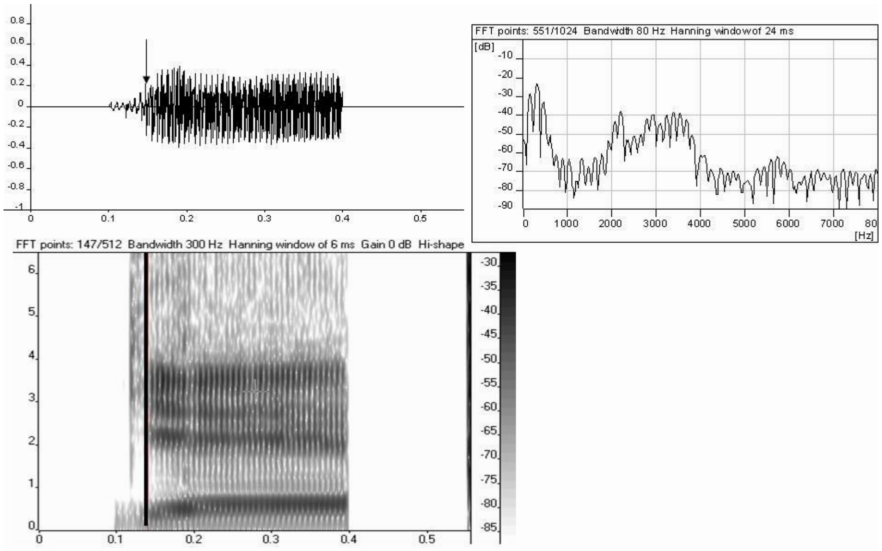

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(i)


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(ii)

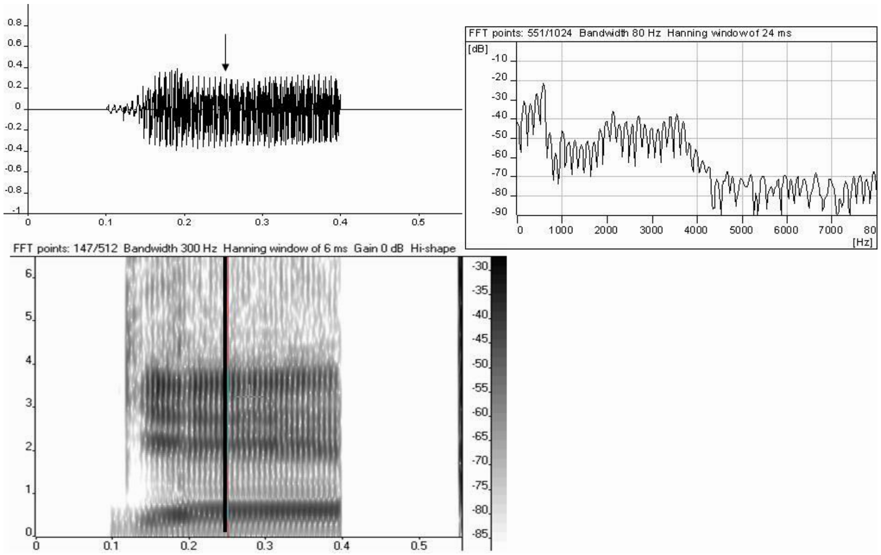
[ɟe]



FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms

FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(i)

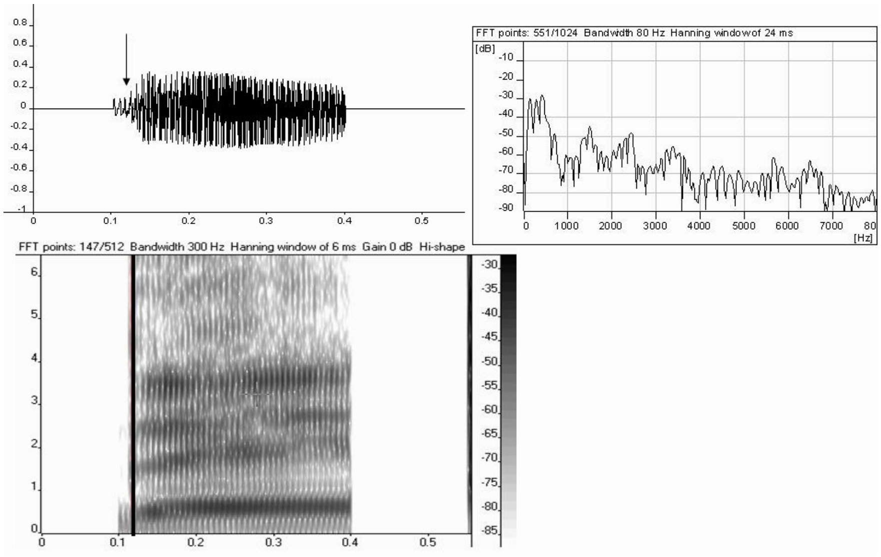


FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms

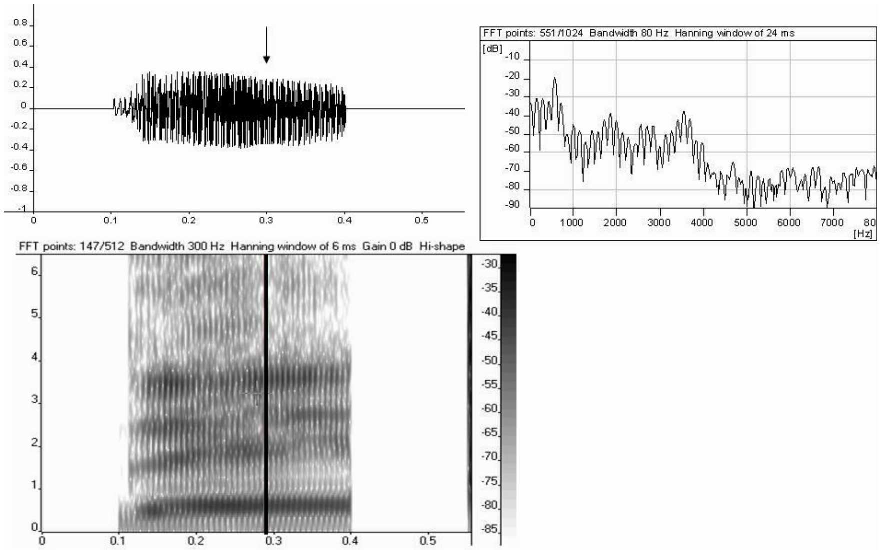
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(ii)

[ge]

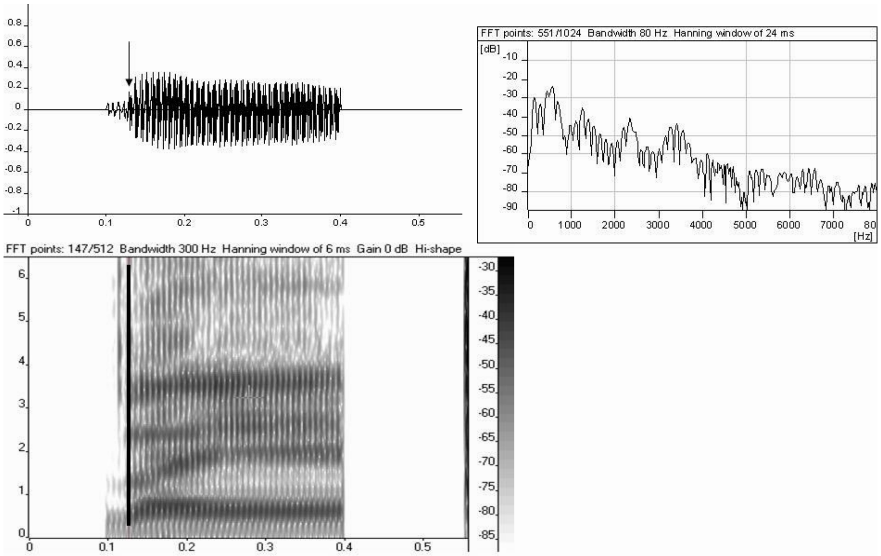

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0 0.1 0.2 0.3 0.4 0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10 -20 -30 -40 -50 -60 -70 -80 -90
0 1000 2000 3000 4000 5000 6000 7000 80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6 5 4 3 2 1 0
0 0.1 0.2 0.3 0.4 0.5
-30 -35 -40 -45 -50 -55 -60 -65 -70 -75 -80 -85

(i)


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0 0.1 0.2 0.3 0.4 0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10 -20 -30 -40 -50 -60 -70 -80 -90
0 1000 2000 3000 4000 5000 6000 7000 80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6 5 4 3 2 1 0
0 0.1 0.2 0.3 0.4 0.5
-30 -35 -40 -45 -50 -55 -60 -65 -70 -75 -80 -85

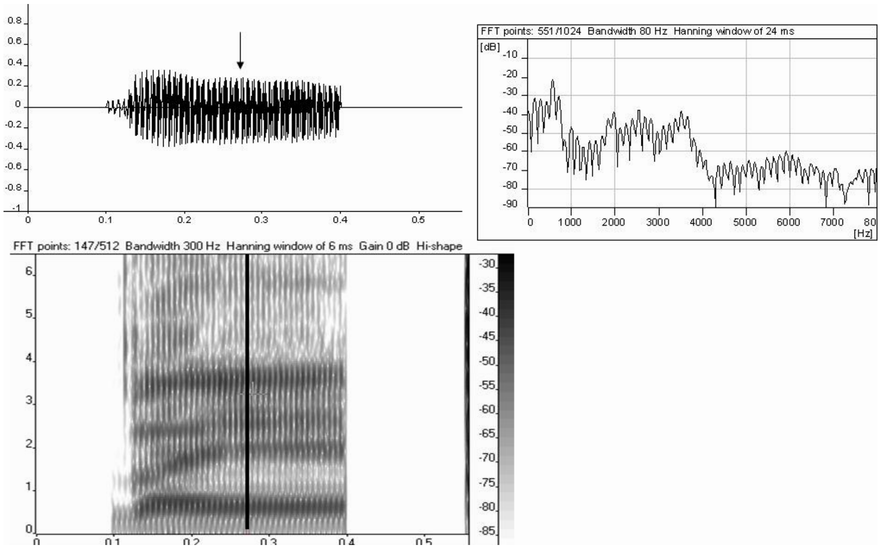
(ii)

[ɢe]



0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
0
0.1
0.2
0.3
0.4
0.5
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

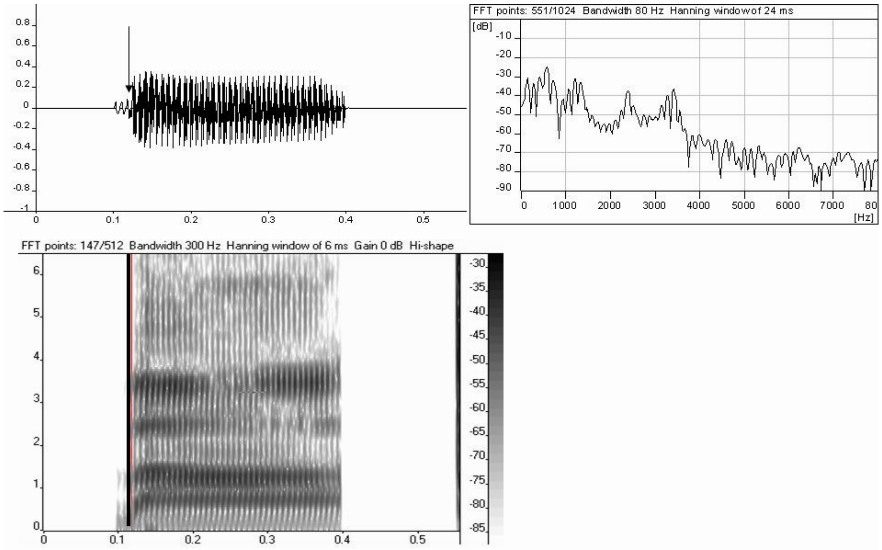
(i)



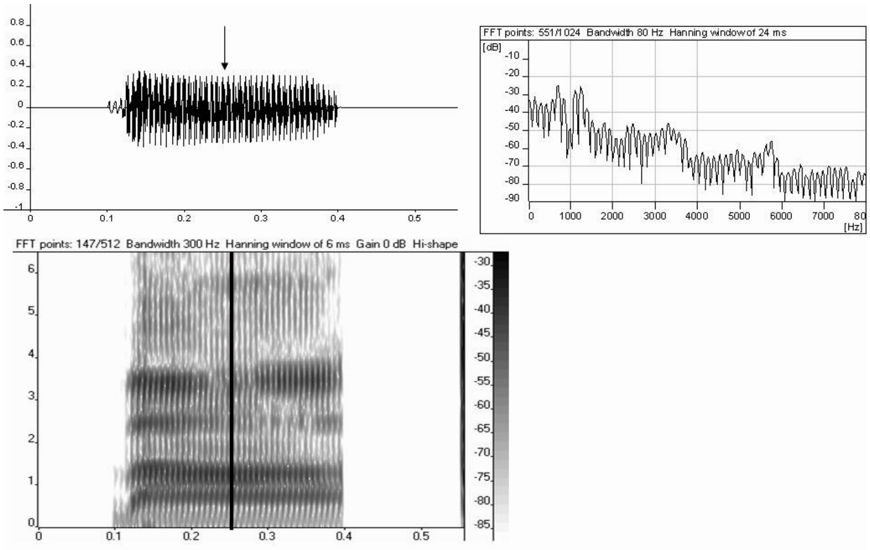
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
0
0.1
0.2
0.3
0.4
0.5
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(ii)

［ba］

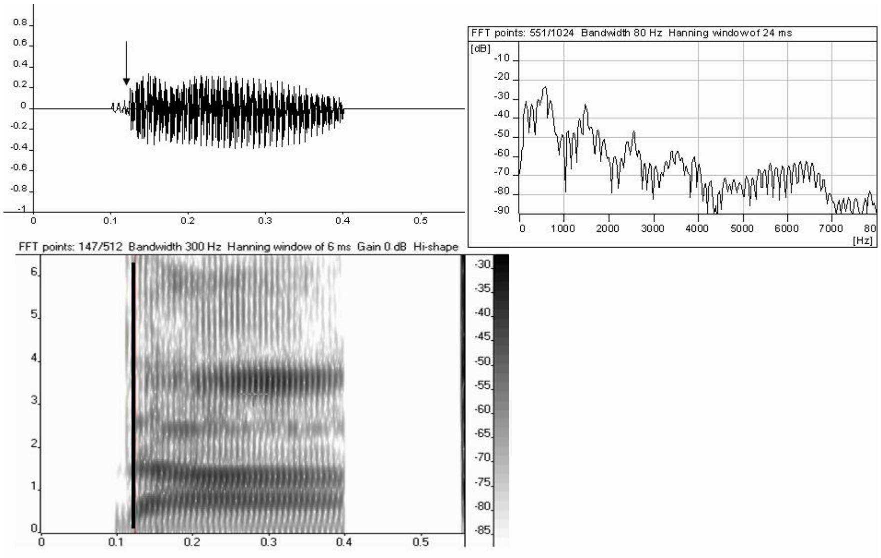

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(i)


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

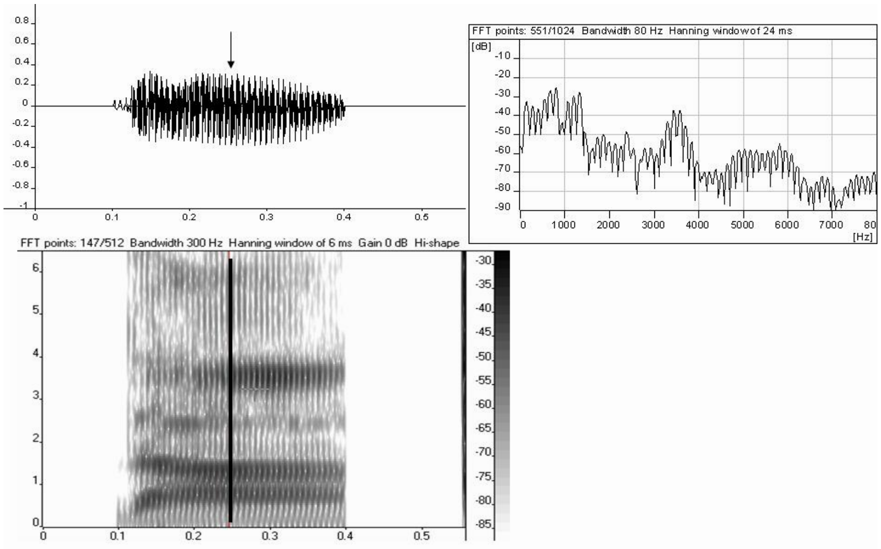
(ii)

[d̪a]



FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

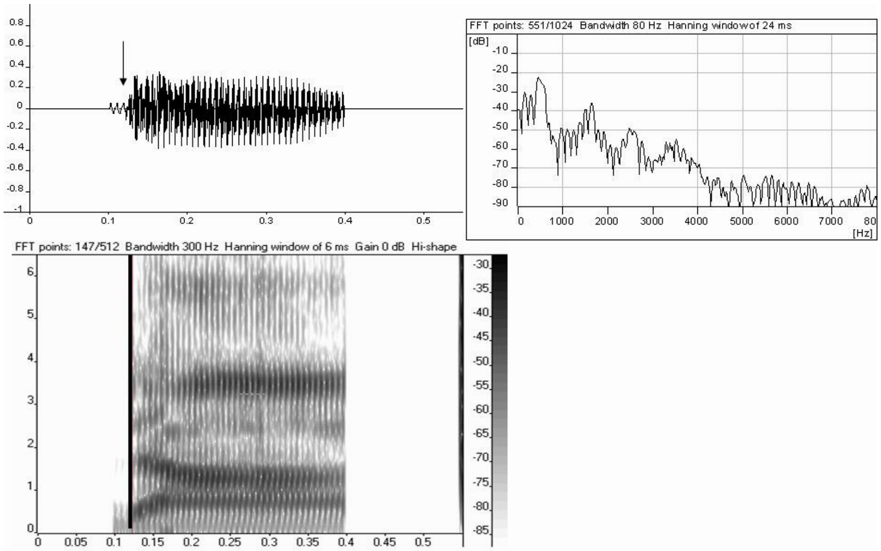
(i)



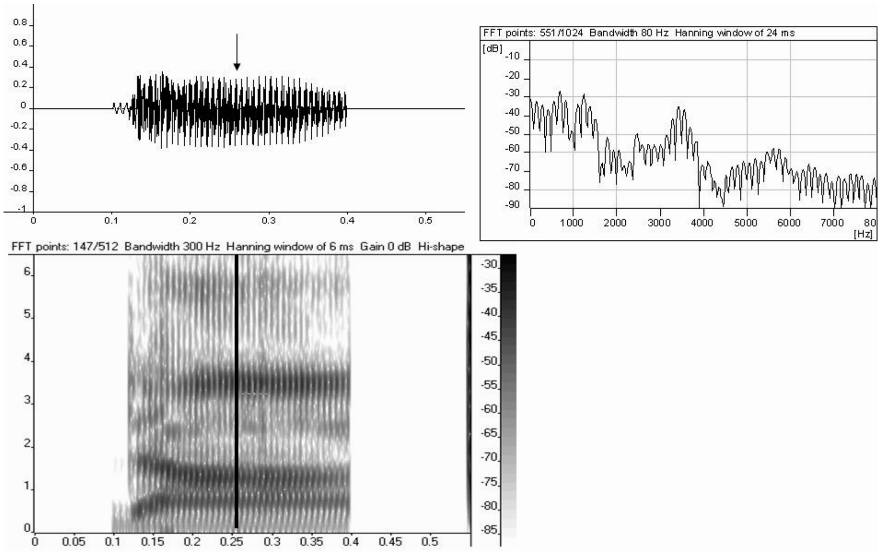
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(ii)

［da］

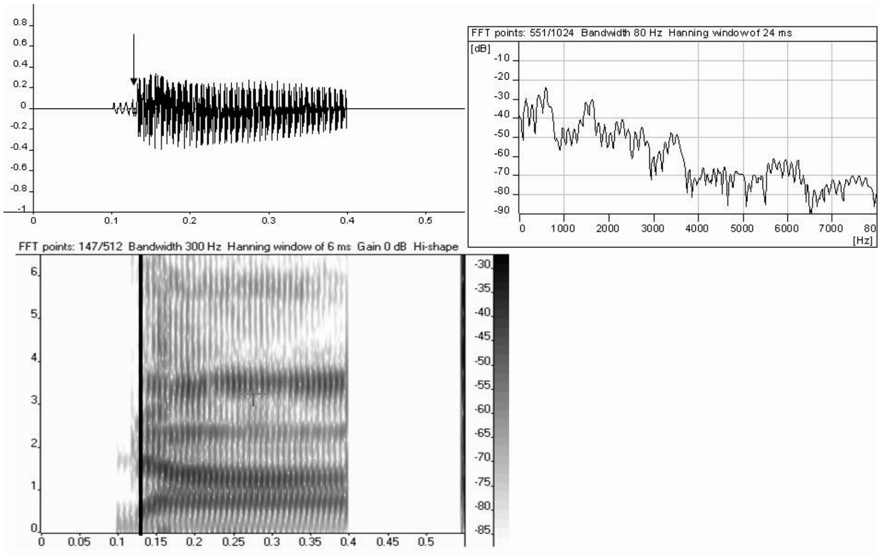

FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(i)

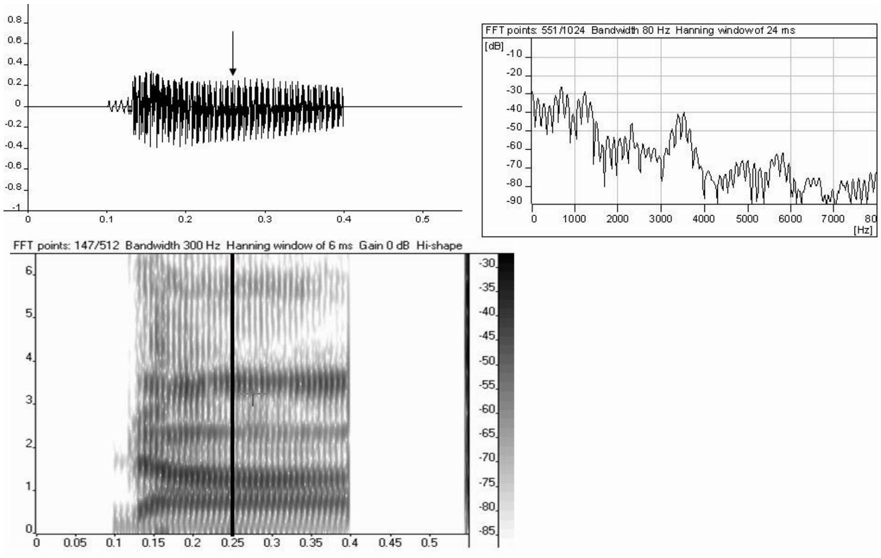

FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(ii)

[ɖa]

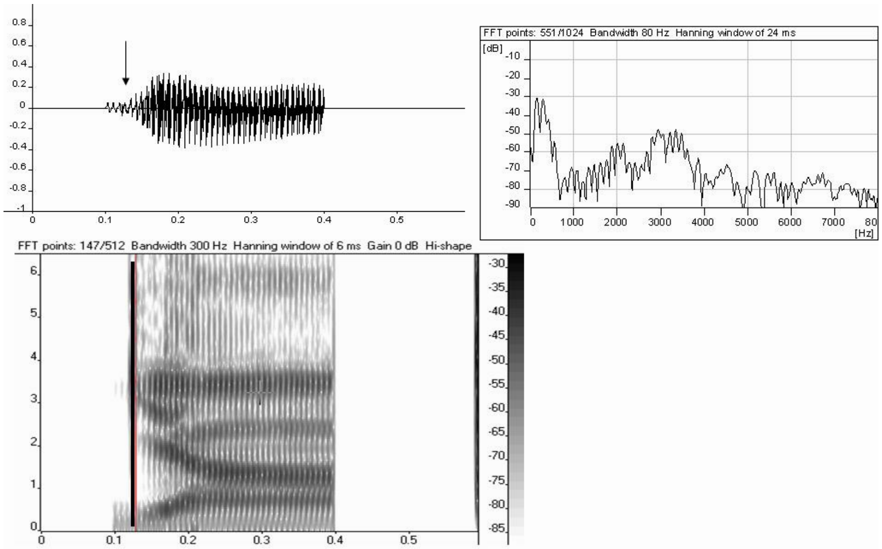

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(i)

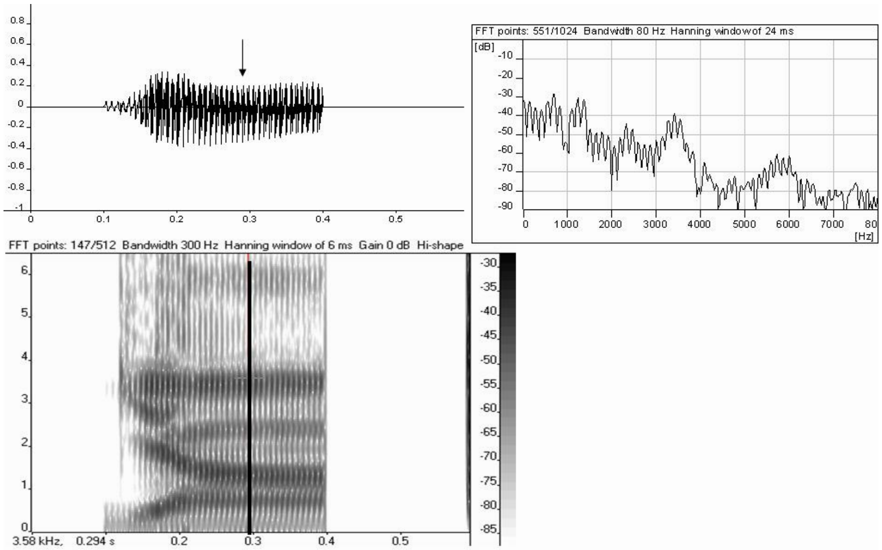

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(ii)

[ɟa]

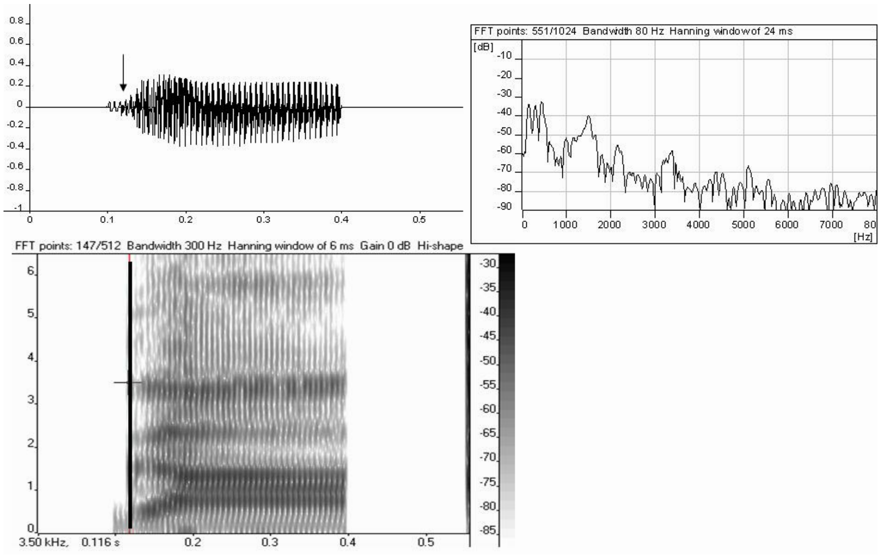

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
0
0.1
0.2
0.3
0.4
0.5
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(i)


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
3.58 kHz, 0.294 s
0.2
0.3
0.4
0.5
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(ii)

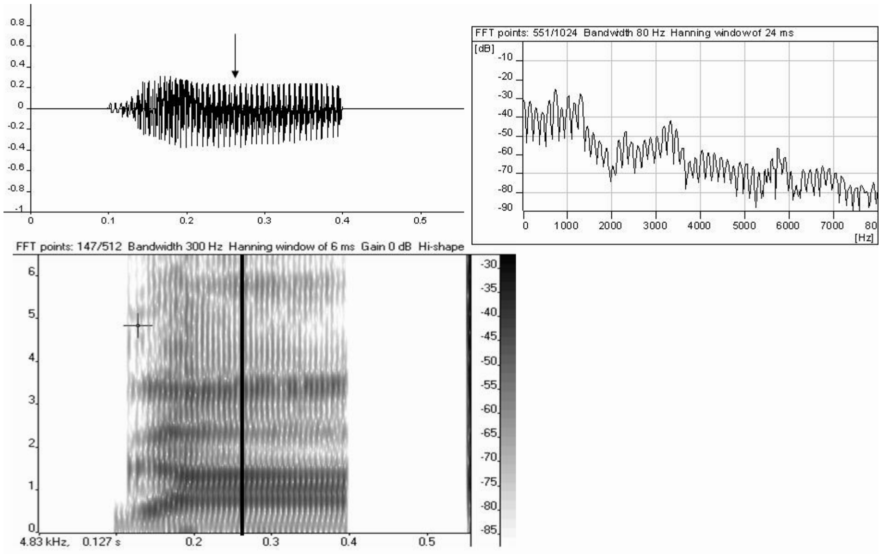
[ga]


FFT points: 551/1024 Bandwidth 80 Hz Hanning windowof 24 ms

FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

3.50 kHz, 0.116 s

(i)

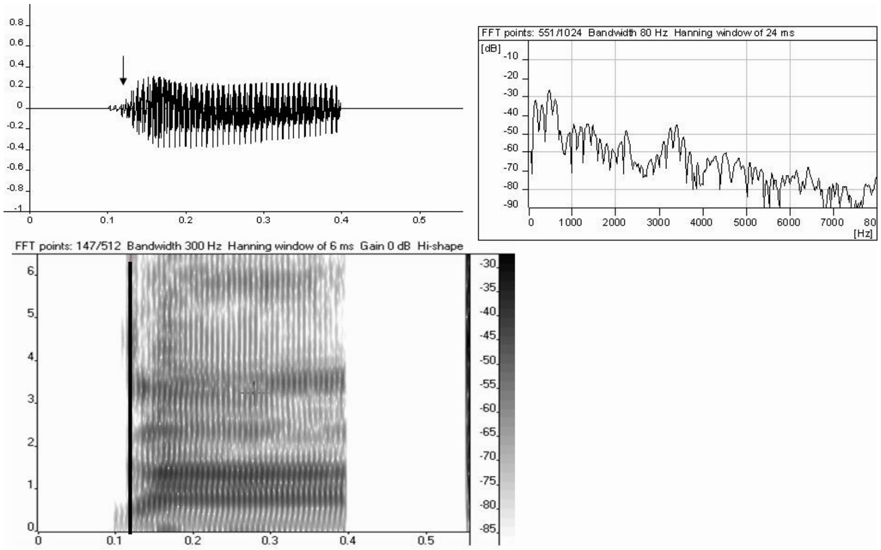

FFT points: 551/1024 Bandwidth 80 Hz Hanning windowof 24 ms

FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

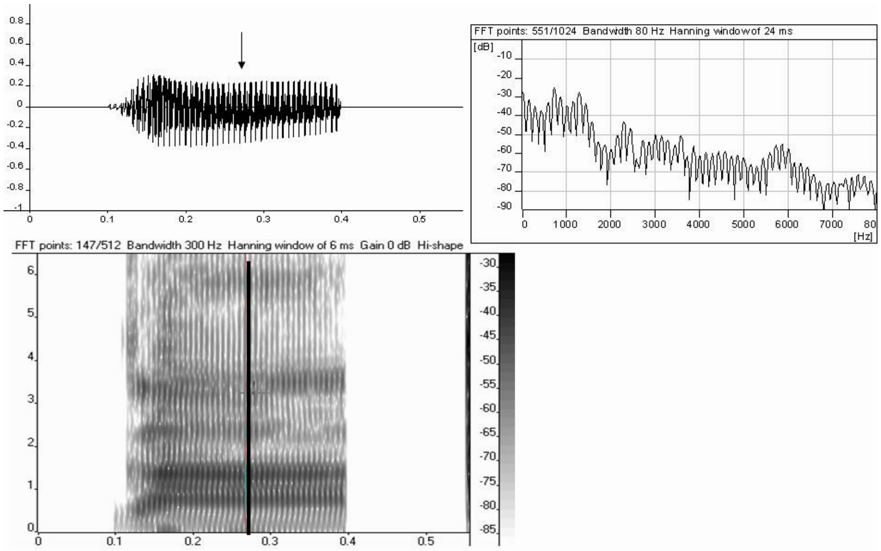
4.83 kHz, 0.127 s

(ii)

[ɢa]

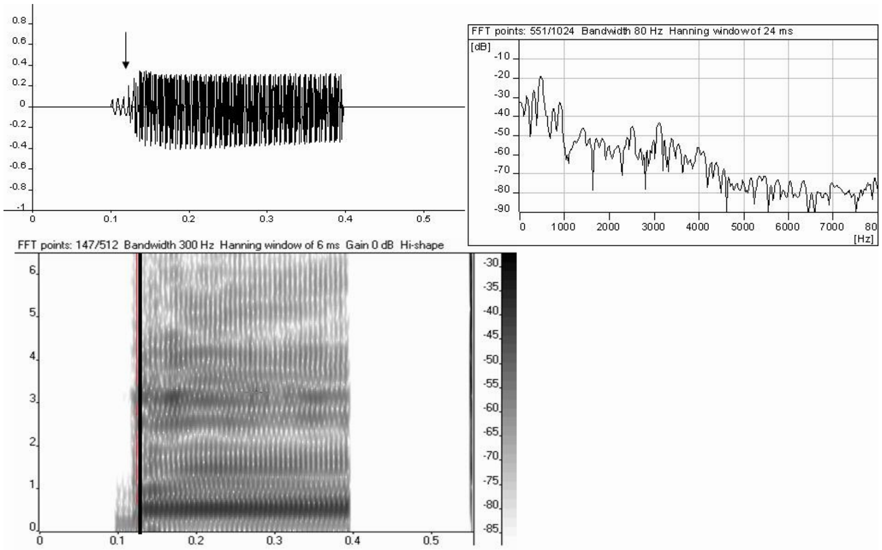

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(i)


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

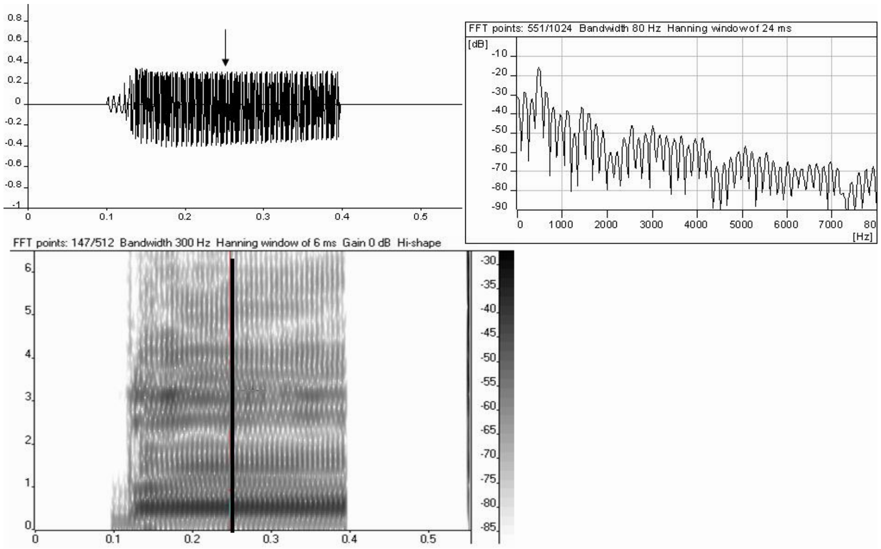
(ii)

[bɔ]


FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms

FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

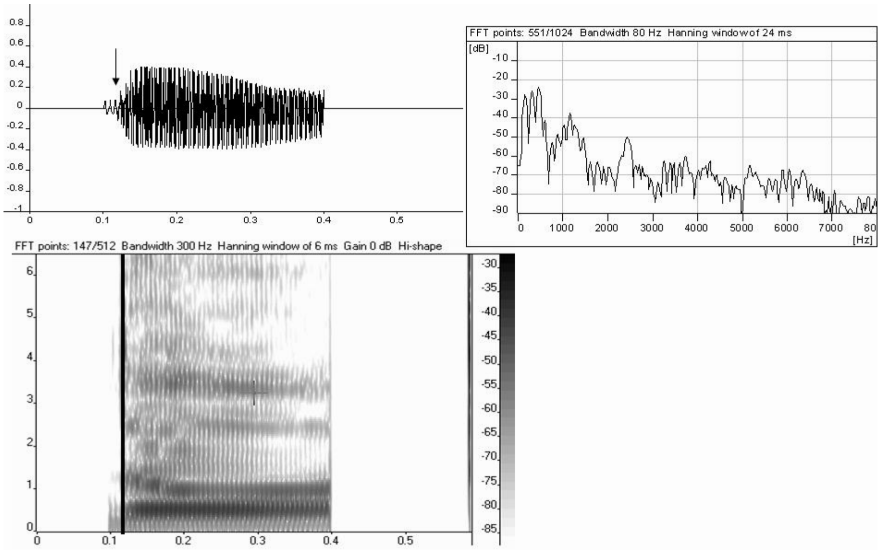
(i)


FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms

FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

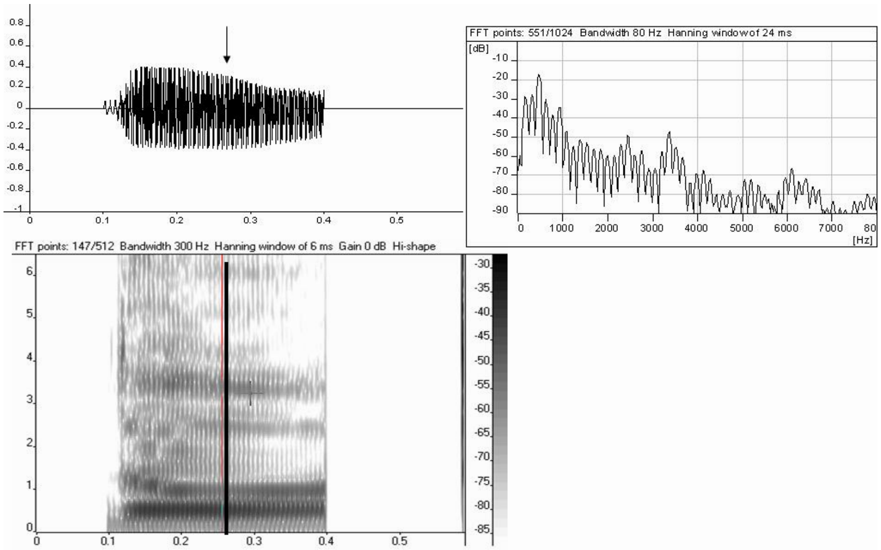
(ii)

[d̪ɔ]



FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

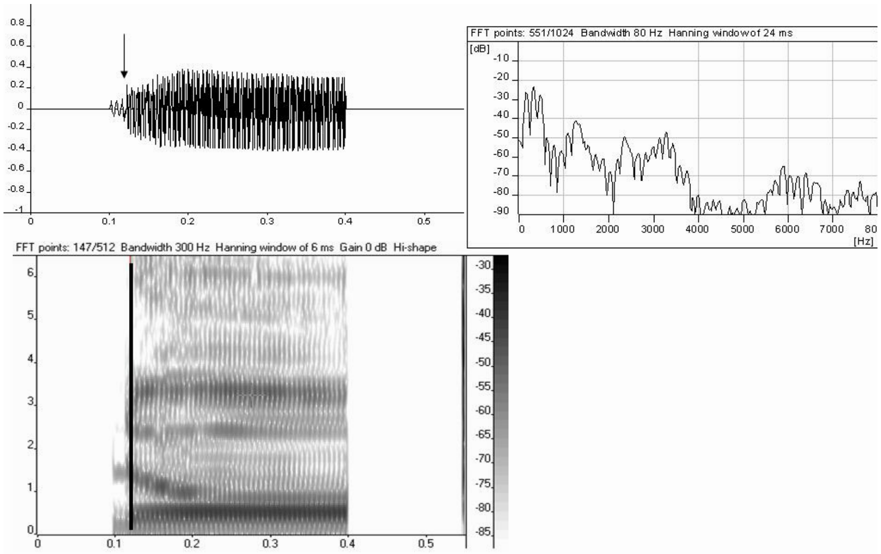
(i)



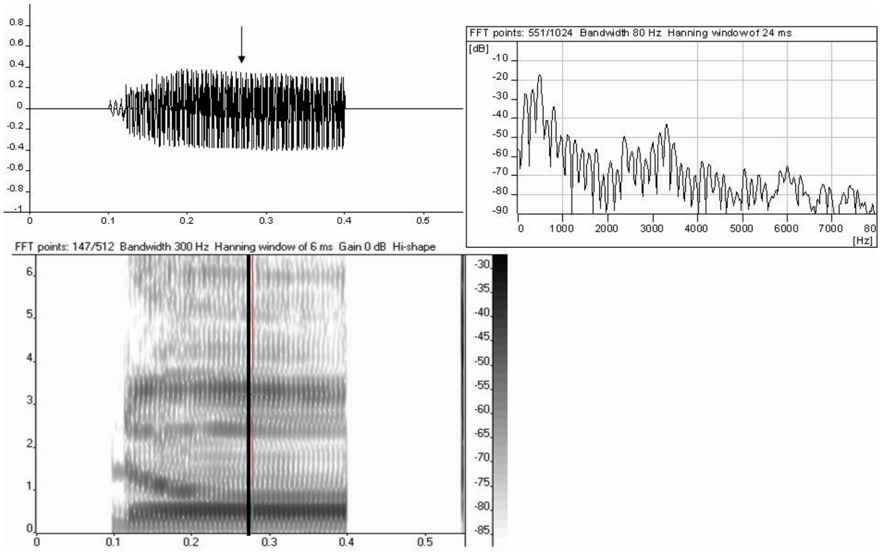
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(ii)

[dɔ]

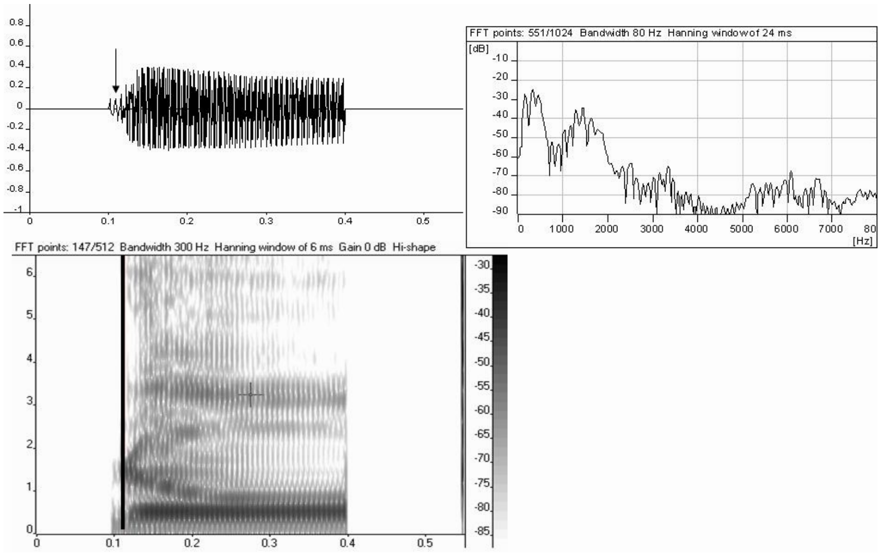

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(i)


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

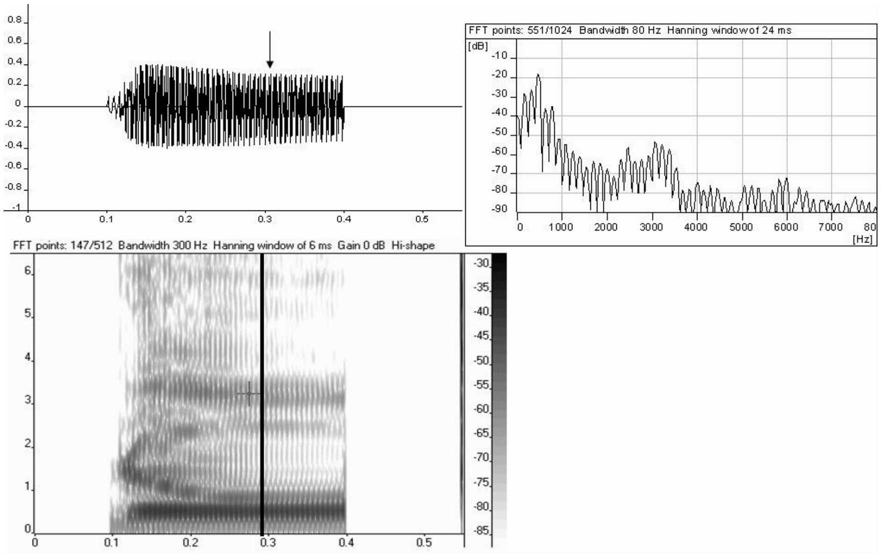
(ii)

[ɖɔ]



0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

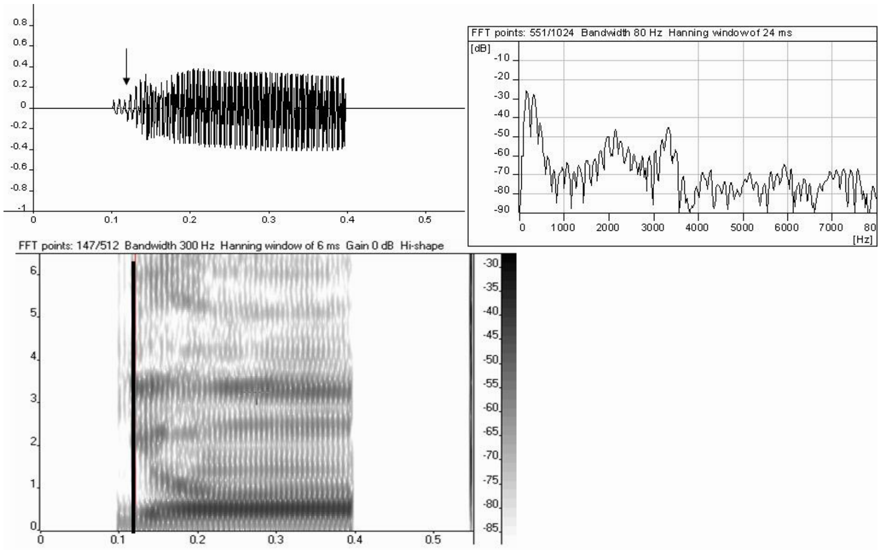
(i)



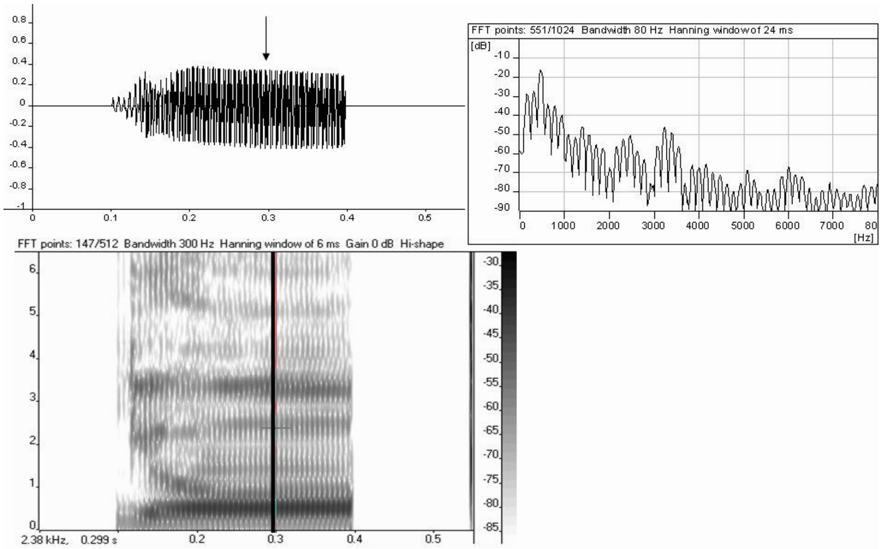
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(ii)

[ɟɔ]

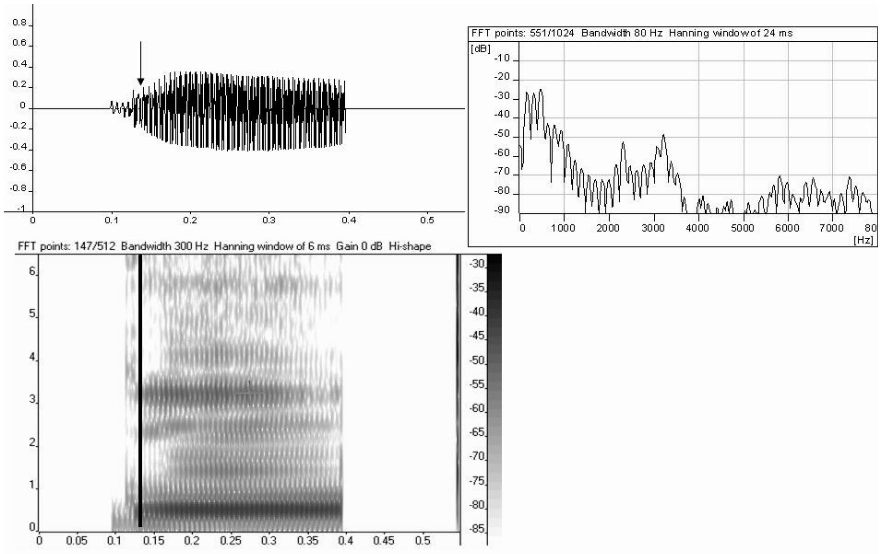

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(i)

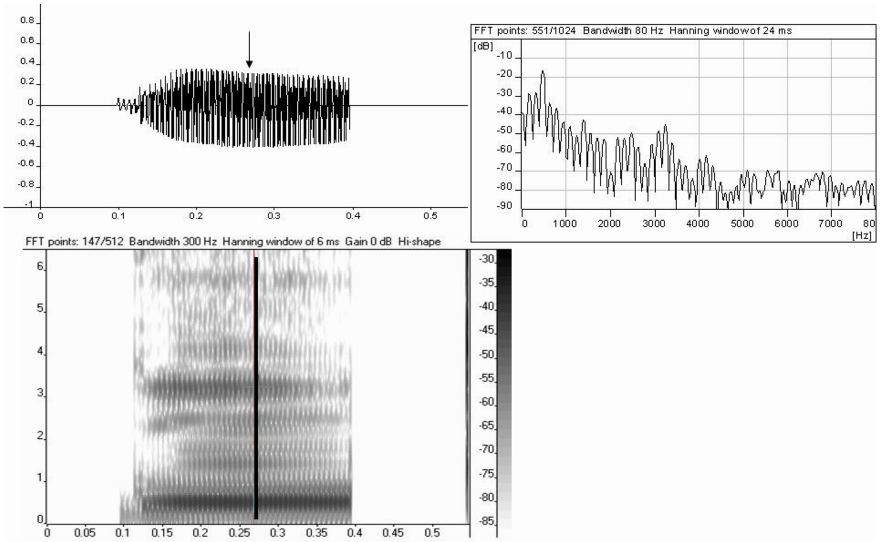

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
2.38 kHz, 0.299 s
0.2
0.3
0.4
0.5
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(ii)

[gɔ]

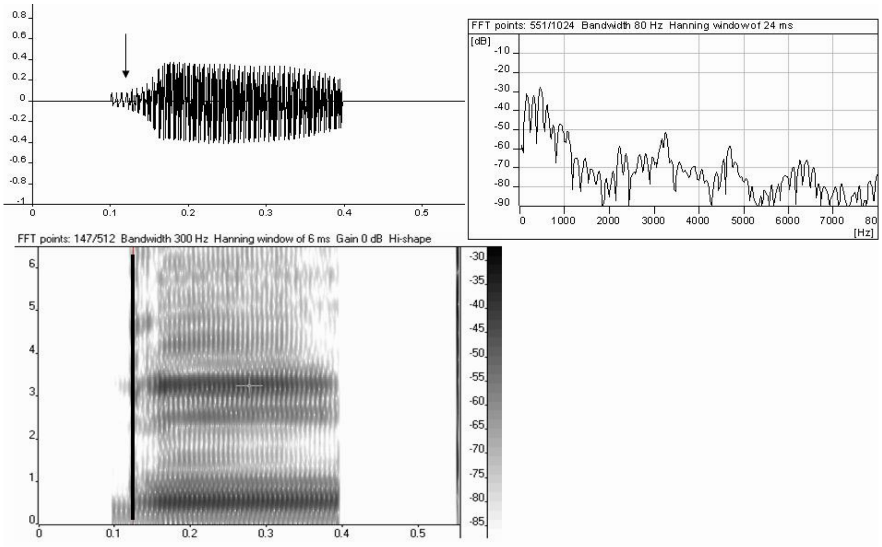

FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0 1000 2000 3000 4000 5000 6000 7000 80
[Hz]
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0 0.1 0.2 0.3 0.4 0.5
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6 5 4 3 2 1 0
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5
-30 -35 -40 -45 -50 -55 -60 -65 -70 -75 -80 -85

(i)

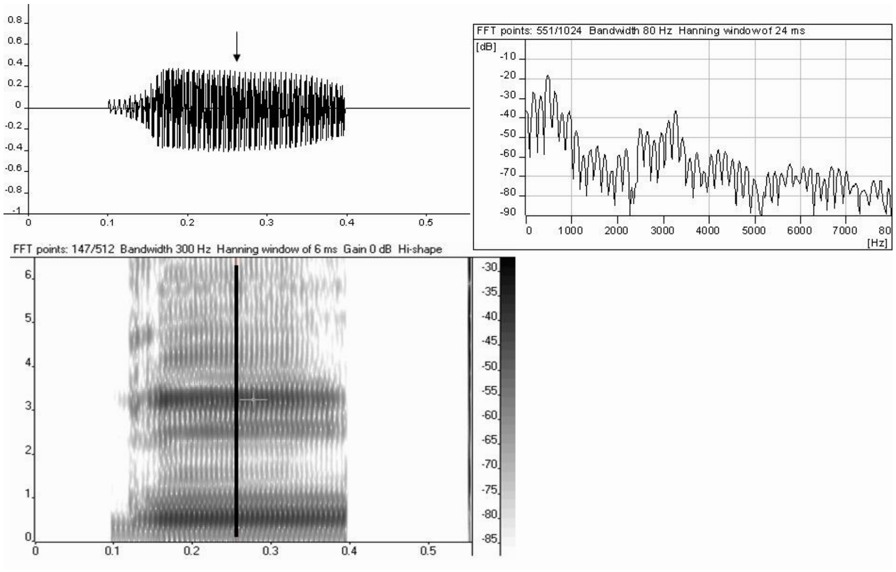

FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0 1000 2000 3000 4000 5000 6000 7000 80
[Hz]
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0 0.1 0.2 0.3 0.4 0.5
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6 5 4 3 2 1 0
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5
-30 -35 -40 -45 -50 -55 -60 -65 -70 -75 -80 -85

(ii)

[ɢɔ]

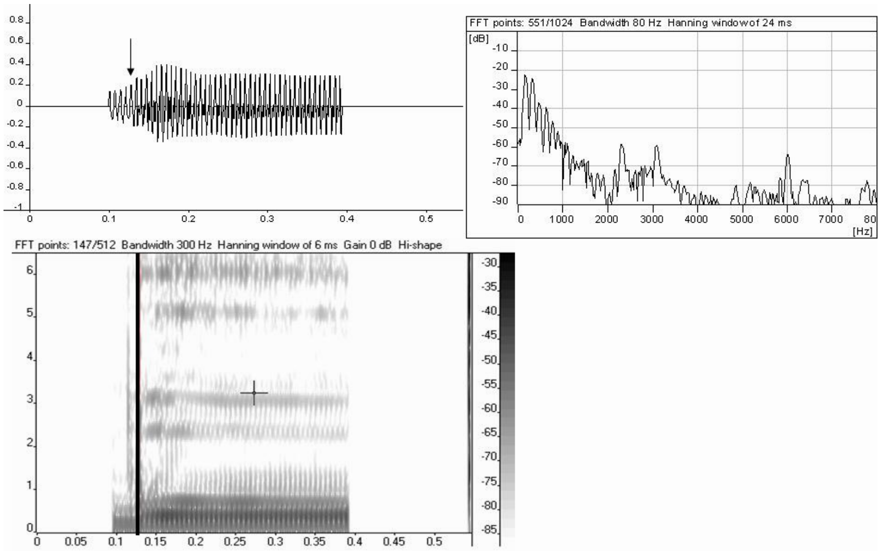

FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(i)


FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

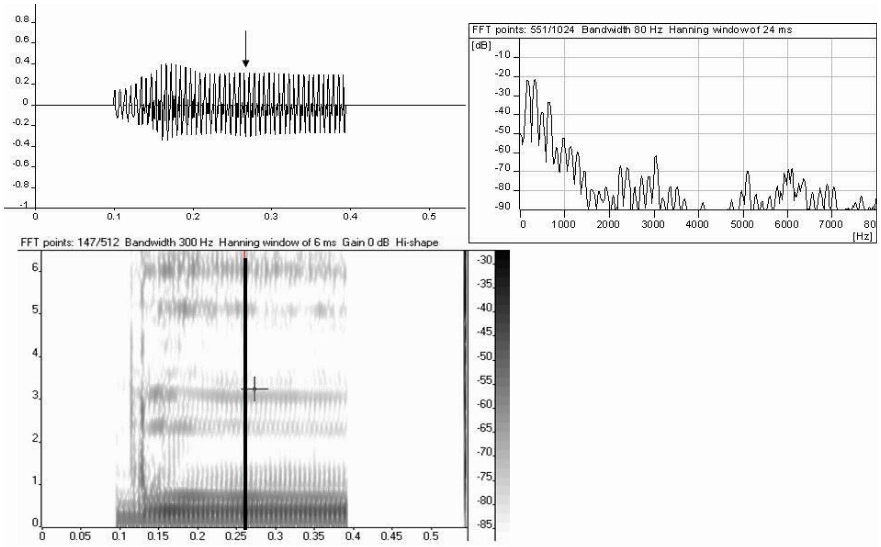
(ii)

[bu]



0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

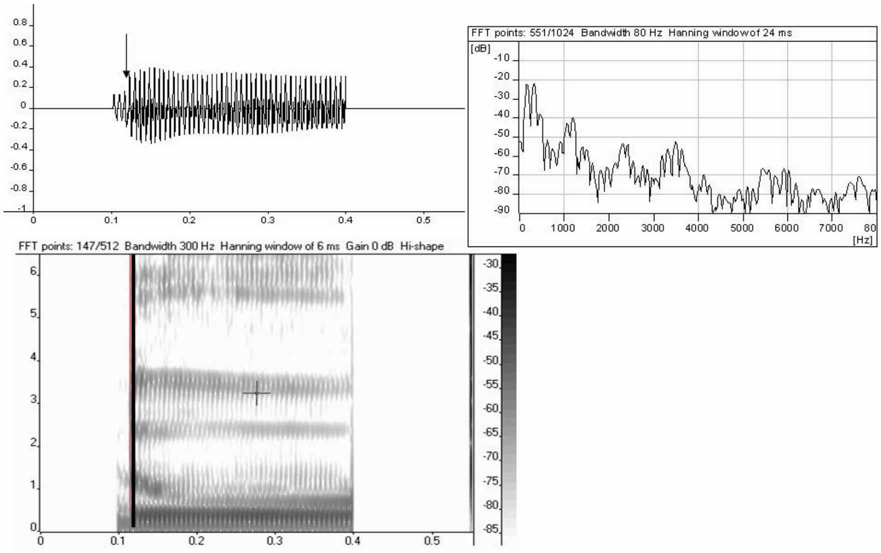
(i)



0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

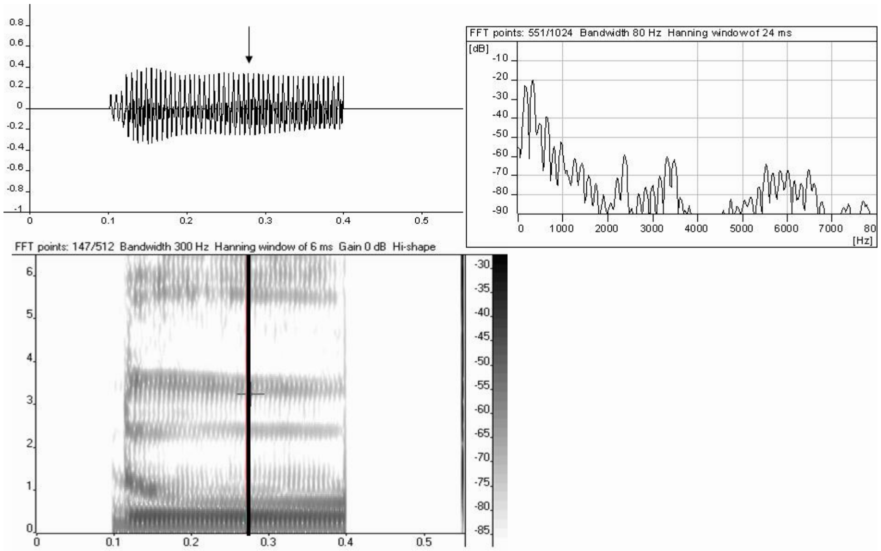
(ii)

[d̪u]


FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms

FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

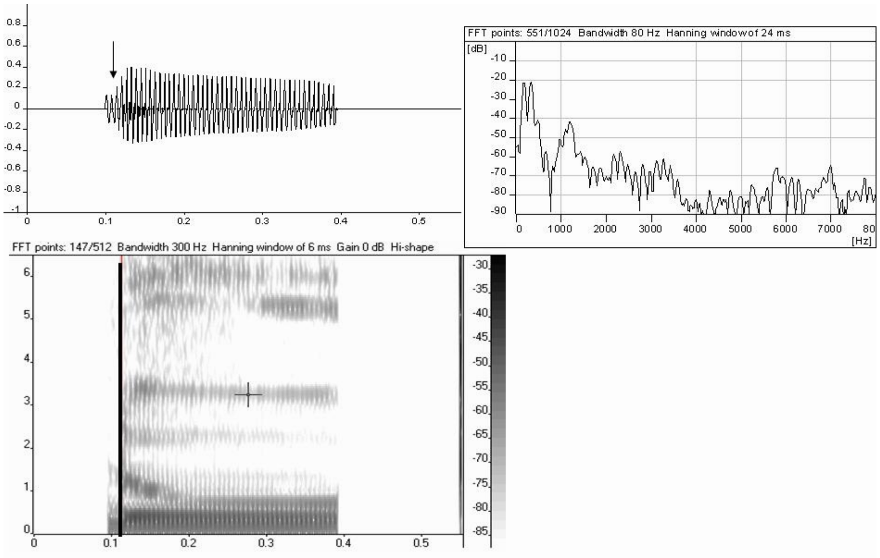
(i)


FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms

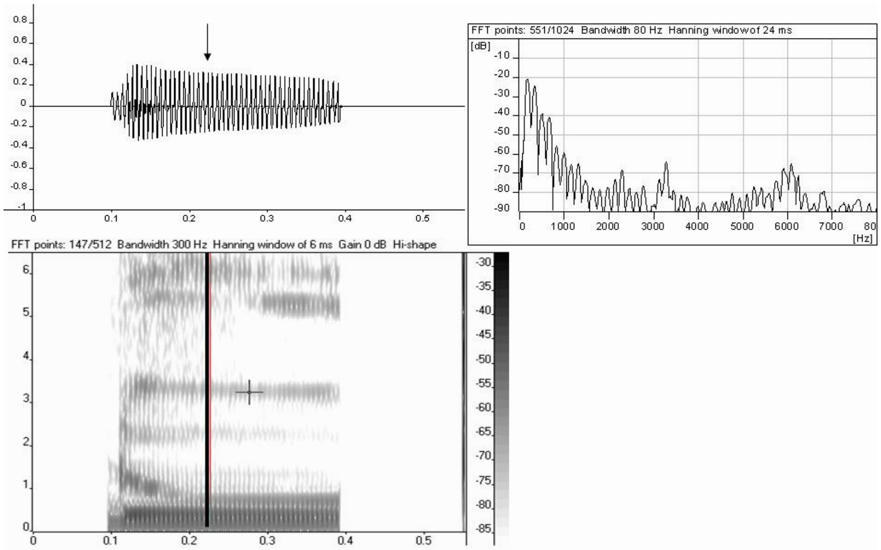
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(ii)

[du]

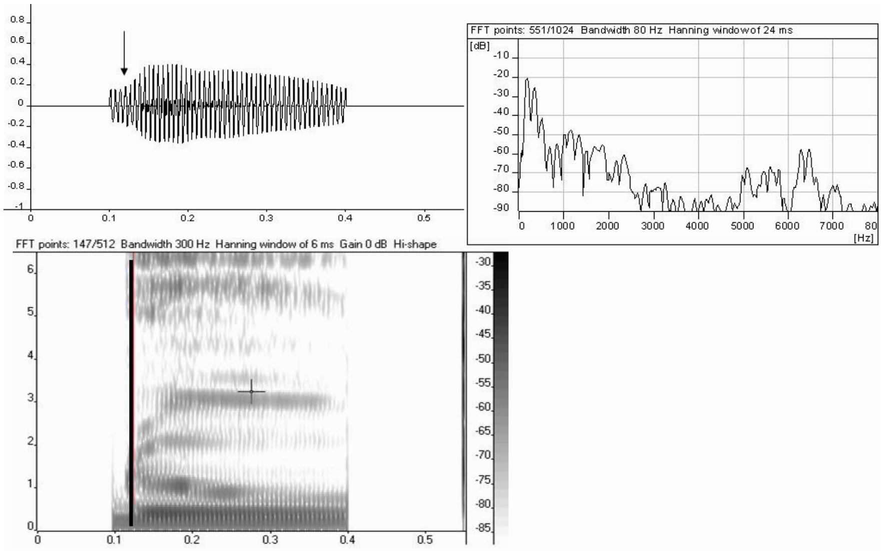

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

(i)


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
1000
2000
3000
4000
5000
6000
7000
80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6
5
4
3
2
1
0
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85

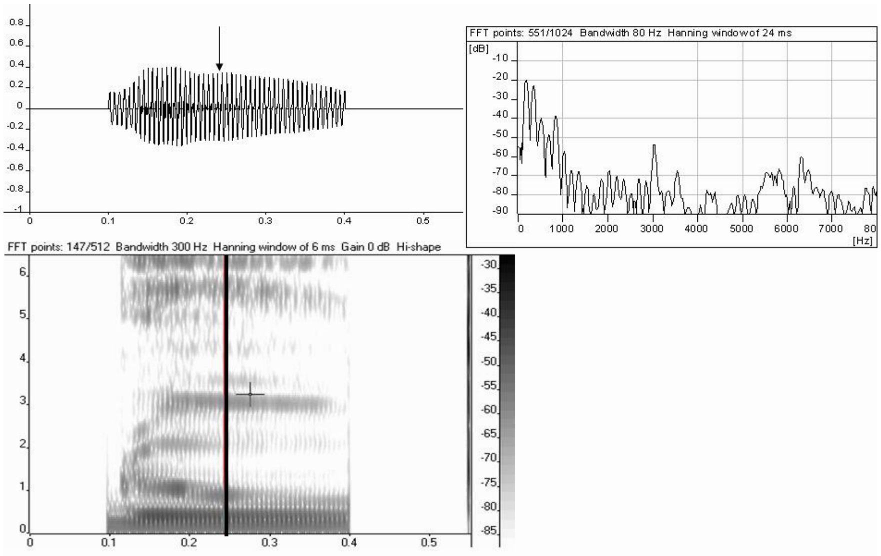
(ii)

[ɖu]


FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms

FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

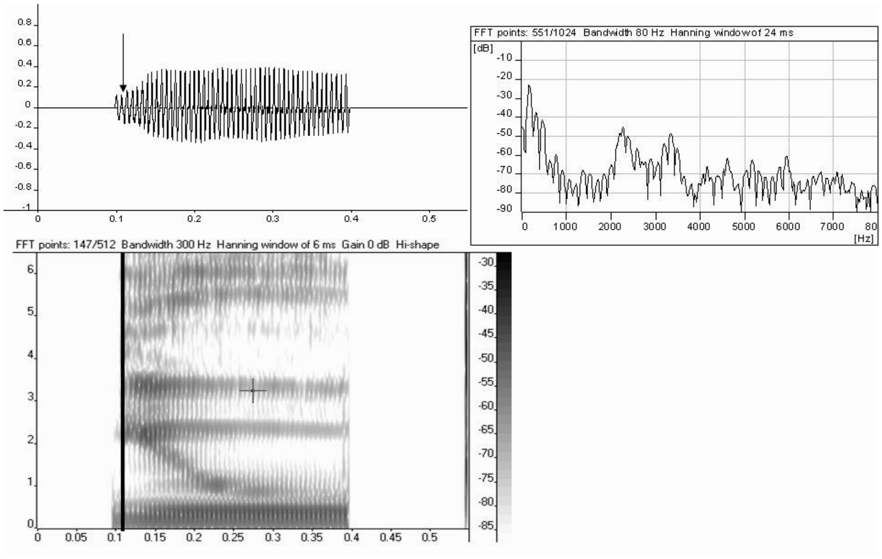
(i)


FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms

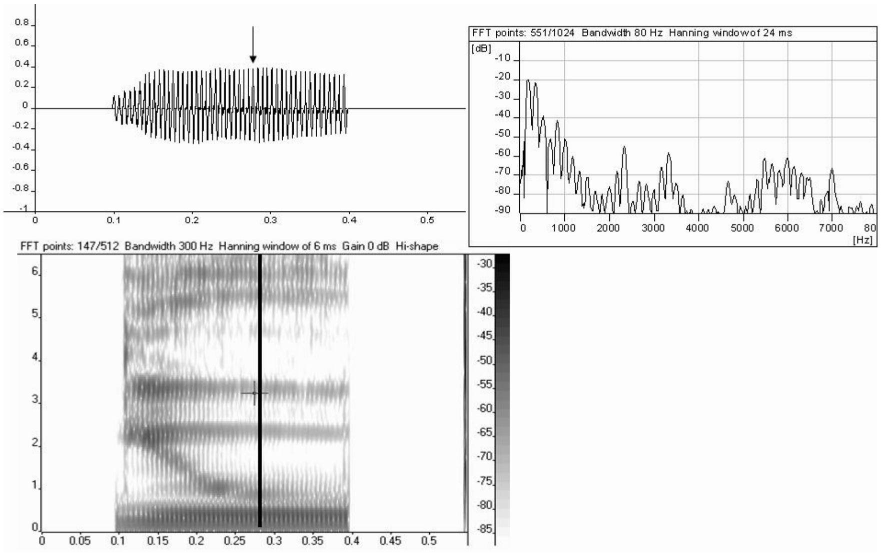
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(ii)

[ɟu]

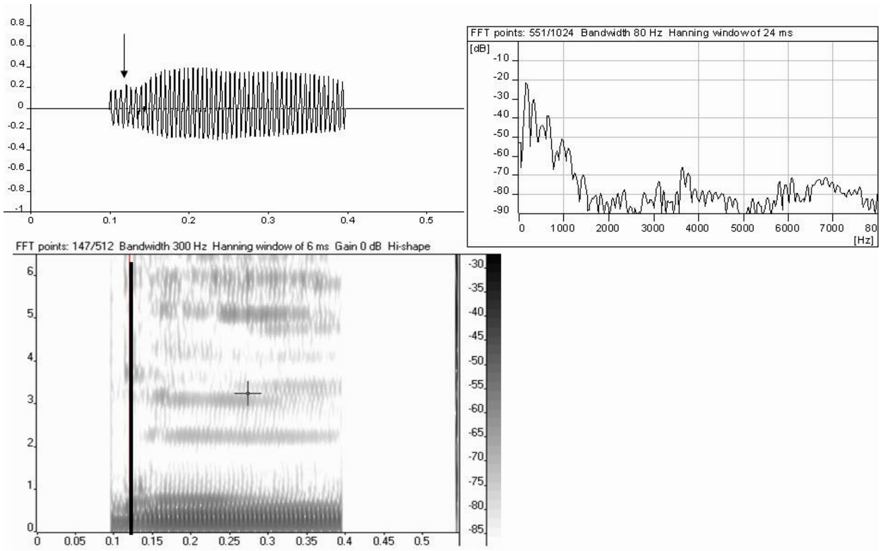

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0 0.1 0.2 0.3 0.4 0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10 -20 -30 -40 -50 -60 -70 -80 -90
0 1000 2000 3000 4000 5000 6000 7000 80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6 5 4 3 2 1 0
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5
-30 -35 -40 -45 -50 -55 -60 -65 -70 -75 -80 -85

(i)


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0 0.1 0.2 0.3 0.4 0.5
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10 -20 -30 -40 -50 -60 -70 -80 -90
0 1000 2000 3000 4000 5000 6000 7000 80
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6 5 4 3 2 1 0
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5
-30 -35 -40 -45 -50 -55 -60 -65 -70 -75 -80 -85

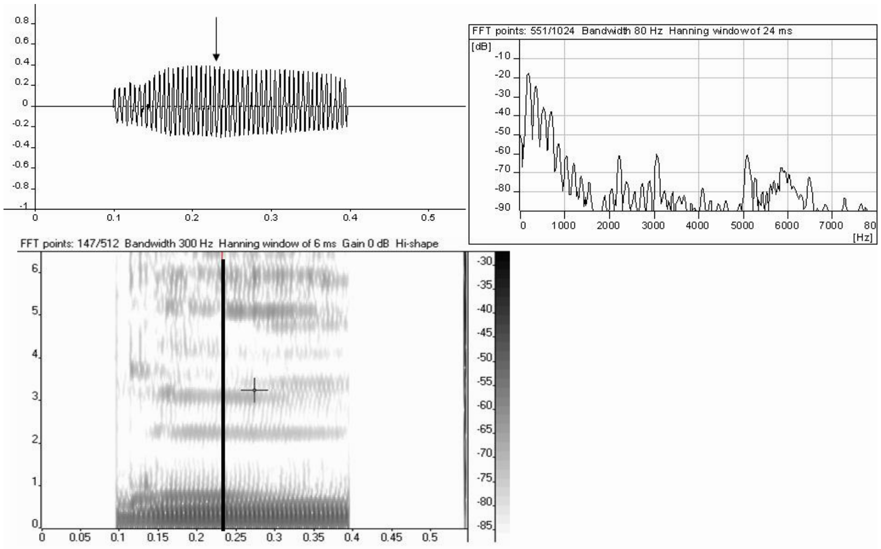
(ii)

[gu]



FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10 -20 -30 -40 -50 -60 -70 -80 -90
0 1000 2000 3000 4000 5000 6000 7000 80
[Hz]
0.8 0.6 0.4 0.2 0 -0.2 -0.4 -0.6 -0.8 -1
0 0.1 0.2 0.3 0.4 0.5
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6 5 4 3 2 1 0
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5
-30 -35 -40 -45 -50 -55 -60 -65 -70 -75 -80 -85

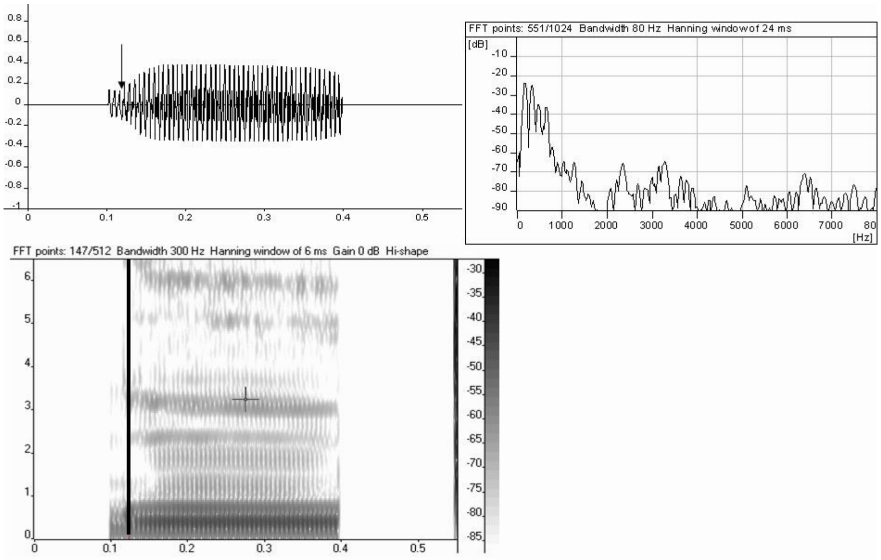
(i)



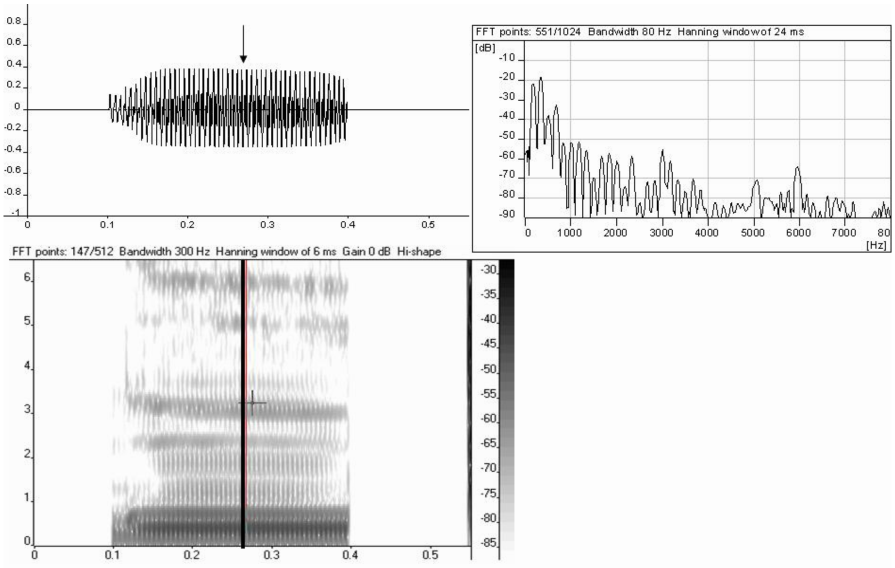
FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
-10 -20 -30 -40 -50 -60 -70 -80 -90
0 1000 2000 3000 4000 5000 6000 7000 80
[Hz]
0.8 0.6 0.4 0.2 0 -0.2 -0.4 -0.6 -0.8 -1
0 0.1 0.2 0.3 0.4 0.5
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape
6 5 4 3 2 1 0
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5
-30 -35 -40 -45 -50 -55 -60 -65 -70 -75 -80 -85

(ii)

[ɢu]


FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(i)


FFT points: 551/1024 Bandwidth 80 Hz Hanning window of 24 ms
[dB]
[Hz]
FFT points: 147/512 Bandwidth 300 Hz Hanning window of 6 ms Gain 0 dB Hi-shape

(ii)

Appendix B.1 Results of single factor ANOVA for the time constants (when the independent variable is consonant place)

The time constants for F1

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.03	6	0.01	1.43	0.24	2.45
Within Groups	0.11	28	0.00			
Total	0.15	34				

The time constants for F2

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.00	6	0.00	0.41	0.87	2.45
Within Groups	0.01	28	0.00			
Total	0.01	34				

The time constants for F3

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.00	6	0.00	0.83	0.56	2.45
Within Groups	0.01	28	0.00			
Total	0.01	34				

Appendix B.2 Results of single factor ANOVA for the time constants (when the independent variable is vowel context)

The time constants for F1

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.03	4	0.01	2.22	0.09	2.69
Within Groups	0.11	30	0.00			
Total	0.15	34				

The time constants for F2

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.00	4	0.00	0.84	0.51	2.69
Within Groups	0.01	30	0.00			
Total	0.01	34				

The time constants for F3

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.00	4	0.00	0.92	0.47	2.69
Within Groups	0.01	30	0.00			
Total	0.01	34				

Appendix C.1 Confusion matrices before symmetrization: Korean Subjects (The number in each cell represents the number of responses)

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	25	0	0	0	0	0	0
d̪i	2	10	8	5	0	0	0
di	0	2	19	4	0	0	0
ɖi	8	2	12	3	0	0	0
ɟi	0	0	0	0	17	1	7
gi	0	0	0	0	6	5	14
ɢi	0	0	0	0	1	0	24

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	22	3	0	0	0	0	0
d̪ɛ	0	5	16	4	0	0	0
dɛ	0	1	19	5	0	0	0
ɖɛ	0	2	16	7	0	0	0
ɟɛ	0	0	0	0	10	9	6
gɛ	4	0	0	0	1	3	17
ɢɛ	5	0	2	0	0	0	18

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	25	0	0	0	0	0	0
d̪a	0	13	10	2	0	0	0
da	0	6	15	4	0	0	0
ɖa	0	0	9	16	0	0	0
ɟa	0	0	0	0	25	0	0
ga	0	0	0	0	5	11	9
ɢa	0	0	0	0	2	3	20

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	25	0	0	0	0	0	0
d̪ɔ	1	14	9	1	0	0	0
dɔ	0	3	16	6	0	0	0
ɖɔ	0	1	6	18	0	0	0
ɟɔ	0	2	0	0	22	0	1
gɔ	0	0	0	0	0	12	13
ɢɔ	0	0	0	0	0	6	19

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	25	0	0	0	0	0	0
d̪u	1	11	8	5	0	0	0
du	0	3	14	8	0	0	0
ɖu	0	2	2	21	0	0	0
ɟu	0	0	0	0	19	5	1
gu	0	0	0	0	0	14	11
ɢu	0	0	0	0	0	9	16

Appendix C.2 Confusion matrices before symmetrization: English Subjects (The number in each cell represents the number of responses)

	bi	d̪i	di	ɖi	ɟi	ɡi	ɢi
bi	25	0	0	0	0	0	0
d̪i	4	13	6	0	1	0	1
di	4	2	14	3	0	0	2
ɖi	10	1	8	4	1	0	1
ɟi	0	0	0	0	9	15	1
ɡi	0	0	0	0	5	15	5
ɢi	0	0	0	0	1	4	20

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	ɡɛ	ɢɛ
bɛ	20	4	0	0	0	0	1
d̪ɛ	1	16	8	0	0	0	0
dɛ	0	6	17	2	0	0	0
ɖɛ	2	6	14	3	0	0	0
ɟɛ	0	1	0	0	12	11	1
ɡɛ	0	0	0	0	0	13	12
ɢɛ	0	2	0	0	0	4	19

	ba	d̪a	da	ɖa	ɟa	ɡa	ɢa
ba	25	0	0	0	0	0	0
d̪a	0	8	16	1	0	0	0
da	0	10	14	1	0	0	0
ɖa	0	3	8	14	0	0	0
ɟa	0	0	0	0	25	0	0
ɡa	0	0	0	0	1	22	2
ɢa	0	0	0	0	1	8	16

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	25	0	0	0	0	0	0
d̪ɔ	0	16	8	1	0	0	0
dɔ	0	0	8	16	1	0	0
ɖɔ	0	0	2	23	0	0	0
ɟɔ	0	0	0	0	25	0	0
gɔ	0	0	1	0	0	16	8
ɢɔ	0	0	0	0	0	3	22

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	25	0	0	0	0	0	0
d̪u	0	13	8	4	0	0	0
du	0	5	9	11	0	0	0
ɖu	2	1	2	19	0	1	0
ɟu	0	0	0	0	25	0	0
gu	0	0	0	1	1	21	2
ɢu	0	0	1	1	0	13	10

Appendix C.3 Confusion matrices before symmetrization: Hindi Subjects (The number in each cell represents the number of responses)

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	17	2	1	0	1	1	3
d̪i	0	14	2	2	1	3	3
di	0	5	14	2	0	0	4
ɖi	2	2	14	4	1	0	2
ɟi	0	0	0	0	8	15	2
gi	1	0	1	3	6	11	3
ɢi	0	0	0	2	1	12	10

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	17	3	1	0	0	1	3
d̪ɛ	1	16	5	0	0	1	2
dɛ	0	6	13	3	1	1	1
ɖɛ	1	7	12	3	0	0	2
ɟɛ	0	0	0	1	11	12	1
gɛ	0	0	0	1	3	7	14
ɢɛ	0	2	4	2	3	5	9

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	24	0	0	0	0	0	1
d̪a	0	23	0	0	0	0	2
da	0	1	18	5	0	1	0
ɖa	0	3	11	11	0	0	0
ɟa	3	2	0	0	16	2	2
ga	1	0	1	2	2	17	2
ɢa	4	1	0	3	3	11	3

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	ɡɔ	ɢɔ
bɔ	21	0	0	0	0	1	3
d̪ɔ	0	22	2	0	0	0	1
dɔ	0	2	16	7	0	0	0
ɖɔ	0	4	10	11	0	0	0
ɟɔ	0	0	0	0	20	1	4
ɡɔ	0	0	0	2	4	17	2
ɢɔ	0	0	1	2	4	13	5

	bu	d̪u	du	ɖu	ɟu	ɡu	ɢu
bu	24	0	0	0	0	0	1
d̪u	0	5	16	3	0	0	1
du	0	3	19	3	0	0	0
ɖu	0	3	7	13	0	1	1
ɟu	0	0	0	5	19	0	1
ɡu	0	0	0	3	3	16	3
ɢu	1	0	0	0	0	12	12

Appendix C.3 Confusion matrices before symmetrization: Spanish Subjects (The number in each cell represents the number of responses)

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	25	0	0	0	0	0	0
d̪i	0	16	9	0	0	0	0
di	5	8	6	2	2	2	0
ɖi	8	11	1	1	1	1	2
ɟi	0	5	0	2	6	9	3
gi	0	0	1	0	4	7	13
ɢi	0	1	0	0	5	6	13

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	19	3	1	1	1	0	0
d̪ɛ	1	11	11	2	0	0	0
dɛ	1	7	12	4	1	0	0
ɖɛ	4	6	10	4	0	1	0
ɟɛ	0	4	0	2	7	11	1
gɛ	0	2	1	2	9	9	2
ɢɛ	6	1	1	3	3	7	4

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	20	0	0	2	2	0	1
d̪a	0	13	12	0	0	0	0
da	0	8	15	2	0	0	0
ɖa	0	4	12	5	1	3	0
ɟa	0	2	1	1	14	4	3
ga	0	5	0	0	6	13	1
ɢa	1	3	1	0	8	11	1

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	21	2	1	0	1	0	0
d̪ɔ	0	16	8	1	0	0	0
dɔ	0	6	10	6	2	1	0
ɖɔ	0	0	9	10	2	4	0
ɟɔ	0	0	1	2	14	5	3
gɔ	0	5	0	0	7	13	0
ɢɔ	1	1	1	1	8	10	3

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	25	0	0	0	0	0	0
d̪u	0	12	9	1	2	1	0
du	0	3	11	7	3	1	0
ɖu	0	1	9	10	3	2	0
ɟu	0	1	3	1	11	4	5
gu	0	4	0	0	6	14	1
ɢu	3	4	1	0	4	8	5

Appendix C.5 Confusion matrices before symmetrization: All Subjects (The number in each cell represents the number of responses)

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	92	2	1	0	1	1	3
d̪i	6	53	25	7	2	3	4
di	9	17	53	11	2	2	6
ɖi	28	16	35	12	3	1	5
ɟi	0	5	0	2	40	40	13
gi	1	0	2	3	21	38	35
ɢi	0	1	0	2	8	22	67

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	78	13	2	1	1	1	4
d̪ɛ	3	48	40	6	0	1	2
dɛ	1	20	61	14	2	1	1
ɖɛ	7	21	52	17	0	1	2
ɟɛ	0	5	0	3	40	43	9
gɛ	4	2	1	3	13	32	45
ɢɛ	11	5	7	5	6	16	50

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	94	0	0	2	2	0	2
d̪a	0	57	38	3	0	0	2
da	0	25	62	12	0	1	0
ɖa	0	10	40	46	1	3	0
ɟa	3	4	1	1	80	6	5
ga	1	5	1	2	14	63	14
ɢa	5	4	1	3	14	33	40

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	92	2	1	0	1	1	3
d̪ɔ	1	68	27	3	0	0	1
dɔ	0	11	50	35	3	1	0
ɖɔ	0	5	27	62	2	4	0
ɟɔ	0	2	1	2	81	6	8
gɔ	0	5	1	2	11	58	23
ɢɔ	1	1	2	3	12	32	49

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	99	0	0	0	0	0	1
d̪u	1	41	41	13	2	1	1
du	0	14	53	29	3	1	0
ɖu	2	7	20	63	3	4	1
ɟu	0	1	3	6	74	9	7
gu	0	4	0	4	10	65	17
ɢu	4	4	2	1	4	42	43

Appendix D.1 Dissimilarity matrices before symmetrization: Korean Subjects

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	0.24	2.68	3.52	2.24	4.24	5.24	5.84
d̪i	2	0.48	1.24	1.56	3.64	5.32	5.4
di	3.04	1.12	0.24	0.44	3.88	4.68	5.12
ɖi	2.4	1.36	0.4	0.2	3.72	4.96	5.04
ɟi	4.32	4.32	3.96	3.68	0.04	4.68	5.32
gi	5.4	5.28	4.52	4.72	4.48	0.08	2.96
ɢi	5.72	5.6	5	5.24	5.28	2.96	0.04

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	0.16	2.2	2	2.44	4.8	4.12	3.8
d̪ɛ	2.4	0.28	0.44	0.4	4	3.24	3.68
dɛ	2.36	0.36	0.4	1.16	4.12	3.68	3.4
ɖɛ	2.68	1.12	0.8	0.32	3.72	3.44	3.48
ɟɛ	4.92	4.72	4.28	4.4	0.12	4.28	4.72
gɛ	4.64	3.48	3.4	3.32	4.2	0.12	1.64
ɢɛ	4.12	3.96	3.48	3.92	4.28	1.52	0.2

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	0	3.52	3.68	3.76	4.64	5.16	4.96
d̪a	3.24	0.6	0.76	1.76	3.96	4.4	4.52
da	3.24	0.56	0.44	0.52	4.16	4.2	4.72
ɖa	3.76	1.12	0.48	0.32	3.8	4.56	3.96
ɟa	4.6	4.24	4.12	4.4	0.04	3.68	3.76
ga	5	4.56	4.36	4.24	3.32	0.08	0.24
ɢa	5.16	4.8	4.28	4.2	3.6	0.2	0.08

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	0.2	3.72	4.04	3.6	4.92	4.96	5.04
d̪ɔ	3.68	0.32	1.52	1.6	4.08	4.16	4.32
dɔ	3.76	1.8	0.48	1.2	3.8	3.96	4.12
ɖɔ	3.72	2.08	0.76	0.24	3.44	3.96	4.12
ɟɔ	4.2	4.32	3.8	4.16	0	4.24	3.84
gɔ	4.88	4.16	4.2	3.72	3.92	0.08	0.48
ɢɔ	4.2	4.08	4.2	4.16	4.52	0.44	0.04

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	0.04	3.4	3.76	3.96	4.68	4.56	4.36
d̪u	3.44	0.2	1.08	2.88	3.8	4.08	4.28
du	3.44	1.44	0.4	1.4	4.12	3.88	4.16
ɖu	3.72	3	1.32	0.24	3.64	3.76	3.96
ɟu	4.64	4.4	4.44	3.84	0.04	4.68	4.36
gu	4.4	4.56	3.84	3.64	4.48	0.08	2.08
ɢu	4.4	4.04	4.04	3.84	4.56	1.56	0.2

Appendix D.2 Dissimilarity matrices before symmetrization: English Subjects

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	0.16	2.32	2.32	1.8	4.08	4.96	5.52
d̪i	2.48	0.28	2.32	2.16	3.64	4.28	5
di	2.24	1.72	0.04	0.24	4.16	4.88	5.12
ɖi	1.92	2	0.32	0.04	4.2	5	5.44
ɟi	4.4	3.68	4.4	4.84	0.04	4.16	4.16
gi	5.28	5.12	5.32	5	3.92	0.04	2.8
ɢi	5.68	5.12	5.24	5	4.2	2.28	0.08

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	0.2	2	2.52	2.6	4.28	4.28	3.04
d̪ɛ	2	0.16	0.44	0.88	3.68	4	2.68
dɛ	3.04	0.32	0.04	0.96	4.12	4.16	2.8
ɖɛ	2.76	1.64	0.4	0.12	4.2	4.04	3.4
ɟɛ	5.08	4.72	4.64	4.32	0.2	4.08	4.6
gɛ	4.04	3.6	3.88	4.12	4.16	0.12	3.32
ɢɛ	3.4	2.96	2.72	3.52	4.16	2.96	0.04

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	0.04	3.6	3.92	4	5	4.56	4.52
d̪a	3.24	0.08	0.72	1.44	4.36	4.48	4.44
da	3.88	1.2	0.12	1.44	4.44	4.12	4.68
ɖa	3.72	1.16	1.24	0.04	4.72	4.28	4.4
ɟa	5.12	4.92	4.52	4.92	0	3.92	3.88
ga	4.68	4.88	4.36	4.68	3.6	0.08	1.04
ɢa	4.96	5	4.4	4.68	3.8	0.88	0.08

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	ɡɔ	ɢɔ
bɔ	0	4.4	4.04	4.44	5	4.96	4.76
d̪ɔ	3.84	0	2.2	2.28	4.12	4.12	3.96
dɔ	3.96	2.56	0.48	1.68	3.84	4.4	4.44
ɖɔ	4.32	2.96	0.6	0.12	3.96	4.36	4.48
ɟɔ	4.76	4.36	4.2	4.4	0.2	4.12	4
ɡɔ	4.56	4.12	4.56	4.28	3.8	0	1.36
ɢɔ	4.96	4.36	4.92	4.92	4.44	1.4	0.04

	bu	d̪u	du	ɖu	ɟu	ɡu	ɢu
bu	0.08	3.36	3.68	3.84	4.56	3.88	4
d̪u	3.88	0.04	0.64	1.88	4.24	4.16	4.16
du	3.72	1.04	0.2	1.56	4.24	4.44	4.32
ɖu	3.48	2.68	1.84	0.16	4.12	4.12	4.16
ɟu	4.88	4.32	4.84	4.36	0.04	4	4.08
ɡu	3.68	4.84	4.08	4.28	4.16	0.12	1.96
ɢu	3.68	4.08	4.16	4.12	4.2	1.52	0.08

Appendix D.3 Dissimilarity matrices before symmetrization: Hindi Subjects

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	0.28	1.76	3.08	2.24	3.16	3.76	4.56
d̪i	2.44	0.56	2.88	2.96	3.44	3.96	5
di	2.28	2.76	0.92	1.4	3.92	4.32	4.96
ɖi	2.16	2.68	0.64	0.52	4.48	4.12	4.88
ɟi	3.84	3.76	4.24	4.24	0.08	4.28	4.64
gi	4.72	4.44	4.56	4.32	4.48	0.36	3.44
ɢi	5.12	4.96	5.24	5.04	4.96	3.48	0.52

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	0.36	1.6	2.28	3.12	3.88	3.24	2.88
d̪ɛ	1.8	0.64	1.56	1.92	3.72	2.88	3.04
dɛ	2.04	1.24	0.68	1.08	3.52	3.72	3.16
ɖɛ	3.16	2.32	1.08	0.92	3.72	3.52	3.24
ɟɛ	4.96	4.96	4.72	4.76	0.88	4.28	4.96
gɛ	3.4	3.56	3.8	3.56	4.04	0.4	2.88
ɢɛ	3.6	3.36	3.4	4.08	4.64	2.76	0.76

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	0.2	3.96	4.28	4.16	4.04	4.2	4.24
d̪a	3.88	0.36	2.32	3.04	4.52	4.44	3.96
da	4	2.44	0.28	1.2	4.72	4.16	4.44
ɖa	4.52	2.72	0.8	0.16	4.24	4.44	3.92
ɟa	4.6	4.84	4.64	4.64	0.32	4.64	5.12
ga	3.92	4.6	4.88	4.44	4.48	0.2	1.4
ɢa	4.76	4.16	4.36	4.48	4.36	0.8	0.84

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	0.24	4.44	4.6	3.84	4.4	4.44	3.52
d̪ɔ	4	0.68	2.84	3.04	4.36	3.64	4.08
dɔ	4.64	3.24	0.56	1.76	4.12	4.24	4.08
ɖɔ	4.28	3.56	1.48	0.64	4.28	3.96	4.56
ɟɔ	4.52	4.08	3.92	4.68	0.36	4.88	4.4
gɔ	4.56	3.72	4	3.64	4.88	0.44	0.64
ɢɔ	4.04	4	4.4	4.6	4.84	0.96	0.24

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	0.2	3.36	3.88	3.92	4.2	4	3.56
d̪u	4.08	0.44	1.48	3.12	4.28	4.36	3.92
du	4.2	1.76	0.36	2.24	4.52	4.8	4.68
ɖu	4.2	3.24	1.4	0.64	3.88	3.92	4.44
ɟu	4.76	4.72	4.92	4.76	0.32	5.24	4.8
gu	3.64	4.44	4.08	3.8	4.96	0.32	2.2
ɢu	2.96	3.8	4.12	3.6	4.36	1.52	0.36

Appendix D.4 Dissimilarity matrices before symmetrization: Spanish Subjects

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	0.68	3.24	2.88	2.32	4.36	5.12	5.16
d̪i	3.04	0.48	1.68	2	4.44	4.12	5.16
di	2.76	1.08	0.24	0.92	3.64	4.16	5.32
ɖi	2.2	1.44	0.52	0.48	4.52	4.92	5.12
ɟi	4.2	4.64	4.28	4.32	0.16	4.04	5.08
gi	5	5.12	4	4.72	4.48	0.28	3.52
ɢi	5.36	5.08	5.48	4.96	5	1.84	0.12

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	0.24	2.4	2.52	2.32	4.6	3.4	2.12
d̪ɛ	2.64	0.68	0.56	1.36	3.88	2.84	3.28
dɛ	1.96	0.52	0.48	0.76	3.88	3.32	2.44
ɖɛ	2.96	1.36	1.44	0.48	3.72	3.88	3.36
ɟɛ	4.64	4.88	4.72	4.52	0.32	3.96	4.68
gɛ	3.8	3.28	3.4	3.12	3.8	0.56	2.6
ɢɛ	2.8	3.36	3.36	4.16	4.32	2.52	0.36

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	0.24	3.6	3.4	3.52	4.52	4.36	4.04
d̪a	3.32	0.28	0.72	1.44	4.92	4.04	4.08
da	3.2	0.6	0.48	0.96	4.28	3.64	4.6
ɖa	4.04	1.28	1	0.48	4.48	4.04	3.72
ɟa	5.2	4.24	4.28	4.08	0.2	4.12	4.4
ga	4	4	4.4	3.6	3.8	0.32	0.44
ɢa	4.32	4.76	4.12	4.8	4.24	0.48	0.68

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	0.28	3.76	3.96	3.4	4.6	4.4	4.12
d̪ɔ	3.28	0.4	1.4	2.6	3.96	4.24	3.8
dɔ	3.92	1.04	0.76	1.64	3.72	3.72	4.4
ɖɔ	3.88	3.04	0.96	0.36	4.16	3.8	4.68
ɟɔ	4.12	4.8	4.28	4.36	0.28	4.28	3.8
gɔ	4.52	3.68	4.36	3.72	4.8	0.2	0.88
ɢɔ	4.36	4.28	4.72	4.48	4.8	0.68	0.24

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	0.16	3.36	4	4.2	4.92	4.24	3.8
d̪u	3.76	0.28	1.08	3.72	3.44	3.88	4.04
du	3.8	1.36	0.48	1.44	4.2	4.6	4.52
ɖu	4	3.12	1.72	0.44	3.88	4.04	4.36
ɟu	4.96	4.52	4.72	4.92	0.32	4.64	4.2
gu	4.4	4.72	4.36	4.4	4.4	0.32	2.08
ɢu	3.6	4.56	4.4	4.04	4.48	1.72	0.48

Appendix D.5 Dissimilarity matrices before symmetrization: Pooled Data (Each cell is averaged over vowel context)

	bi	d̪i	di	ɖi	ɟi	gi	ɢi
bi	0.34	2.5	2.95	2.15	3.96	4.77	5.27
d̪i	2.49	0.45	2.03	2.17	3.79	4.42	5.14
di	2.58	1.67	0.36	0.75	3.9	4.51	5.13
ɖi	2.17	1.87	0.47	0.31	4.23	4.75	5.12
ɟi	4.19	4.1	4.22	4.27	0.08	4.29	4.8
gi	5.1	4.99	4.6	4.69	4.34	0.19	3.18
ɢi	5.47	5.19	5.24	5.06	4.86	2.64	0.19

	bɛ	d̪ɛ	dɛ	ɖɛ	ɟɛ	gɛ	ɢɛ
bɛ	0.24	2.05	2.33	2.62	4.39	3.76	2.96
d̪ɛ	2.21	0.44	0.75	1.14	3.82	3.24	3.17
dɛ	2.35	0.61	0.4	0.99	3.91	3.72	2.95
ɖɛ	2.89	1.61	0.93	0.46	3.84	3.72	3.37
ɟɛ	4.9	4.82	4.59	4.5	0.38	4.15	4.74
gɛ	3.97	3.48	3.62	3.53	4.05	0.3	2.61
ɢɛ	3.48	3.41	3.24	3.92	4.35	2.44	0.34

	ba	d̪a	da	ɖa	ɟa	ga	ɢa
ba	0.12	3.67	3.82	3.86	4.55	4.57	4.44
d̪a	3.42	0.33	1.13	1.92	4.44	4.34	4.25
da	3.58	1.2	0.33	1.03	4.4	4.03	4.61
ɖa	4.01	1.57	0.88	0.25	4.31	4.33	4
ɟa	4.88	4.56	4.39	4.51	0.14	4.09	4.29
ga	4.4	4.51	4.5	4.24	3.8	0.17	0.78
ɢa	4.8	4.68	4.29	4.54	4	0.59	0.42

	bɔ	d̪ɔ	dɔ	ɖɔ	ɟɔ	gɔ	ɢɔ
bɔ	0.18	4.08	4.16	3.82	4.73	4.69	4.36
d̪ɔ	3.70	0.35	1.99	2.38	4.13	4.04	4.04
dɔ	4.07	2.16	0.57	1.57	3.87	4.08	4.26
ɖɔ	4.05	2.91	0.95	0.34	3.96	4.02	4.46
ɟɔ	4.40	4.39	4.05	4.40	0.21	4.38	4.01
gɔ	4.63	3.92	4.28	3.84	4.35	0.18	0.84
ɢɔ	4.39	4.18	4.56	4.54	4.65	0.87	0.14

	bu	d̪u	du	ɖu	ɟu	gu	ɢu
bu	0.12	3.37	3.83	3.98	4.59	4.17	3.93
d̪u	3.79	0.24	1.07	2.9	3.94	4.12	4.10
du	3.79	1.40	0.36	1.66	4.27	4.43	4.42
ɖu	3.85	3.01	1.57	0.37	3.88	3.96	4.23
ɟu	4.81	4.49	4.73	4.47	0.18	4.64	4.36
gu	4.03	4.64	4.09	4.03	4.50	0.21	2.08
ɢu	3.66	4.12	4.18	3.9	4.40	1.58	0.28

Appendix E Results of multiple regression analyses with six independent variables (Formant-based distance, the time constant distances for F1, F2, and F3, burst spectra distance and articulatory distance)

/i/ vowel context

Regression Statistics	
Multiple R	0.98
R Square	**0.96**
Adjusted R Square	0.96
Standard Error	0.14
Observations	50

ANOVA

	df	SS	MS	F	Significance F
Regression	6	23.54	3.92	195.93	0.00
Residual	43	0.86	0.02		
Total	49	24.40			

/e/ vowel context

Regression Statistics	
Multiple R	0.95
R Square	**0.91**
Adjusted R Square	0.90
Standard Error	0.14
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	5	9.08	1.82	87.61	0.00
Residual	43	0.89	0.02		
Total	48	9.98			

/a/ vowel context

Regression Statistics	
Multiple R	0.92
R Square	**0.84**
Adjusted R Square	0.82
Standard Error	0.23
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	6	12.09	2.02	36.79	0.00
Residual	42	2.30	0.05		
Total	48	14.39			

/o/ vowel context

Regression Statistics	
Multiple R	0.88
R Square	**0.78**
Adjusted R Square	0.75
Standard Error	0.25
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	6	9.33	1.55	24.77	0.00
Residual	42	2.64	0.06		
Total	48	11.96			

/u/ vowel context

Regression Statistics	
Multiple R	0.95
R Square	**0.89**
Adjusted R Square	0.88
Standard Error	0.17
Observations	49

ANOVA

	df	SS	MS	F	Significance F
Regression	6	10.18	1.70	58.94	0.00
Residual	42	1.21	0.03		
Total	48	11.39			

Vowel pooled

Regression Statistics	
Multiple R	0.88
R Square	**0.77**
Adjusted R Square	0.77
Standard Error	0.26
Observations	245

ANOVA

	df	SS	MS	F	Significance F
Regression	6	56.99	9.50	136.60	0.00
Residual	238	16.55	0.07		
Total	244	73.54			

Bibliography

Al-Tamimi, J. (2007). *Indices dynamiques et perception des voyelles: Étude translinguistique en arabe dialectal et en français*, doctoral dissertation, Université Lumière Lyon 2.

Branderud, P., Lundberg, H.-J., Lander, J., Djamshidpey, H., Wäneland, I., Krull, D. & Lindblom, B. (1998). "X-ray analyses of speech: Methodological aspects," in *Proceedings of the XIIIth Swedish Phonetics Conference (FONETIK 1998)*, KTH, Stockholm.

de Groot, A. W. (1931). *Phonologie und Phonetik als Funktionswissenschaften* (TCLP 4.), Prague.

Diehl, R. L., Kluender, K.R. & Walsh, M. A. (1990). "Some auditory basis of speech perception and production," in W. A. Ainsworth (ed.), *Advances in Speech, Hearing and Language Processing*, 243-268, London, JAI Press.

Diehl, R. L. & Kingston, J. (1991). "Phonetic covariation as auditory enhancement: The case of the [+ voice]/[- voice] distinction," *PERILUS: Papers from the Conference on Current Phonetic Research Paradigms: Implications for Speech Motor Control*, Stockholm University, August 13-

16.

Diehl, R. L., Lindblom, B. and Creeger, C. P. (2003). "Increasing realism of auditory representations yields further insights into vowel phonetics," *Proceedings of the 15th International Congress of Phonetic Sciences, Vol2,* 1381-1384, Adelaide, Causal Publications.

Engstrand, O., Frid, J. and Lindblom, B. (2007). "A perceptual bridge between coronal and dorsal /r/," in P. Beddor, M. Ohala and M.-J. Solé (eds.), *Experimental Approaches to Phonology*, 175-191, Oxford University Press.

Ericsdotter, C. (2005). *Articulatory-Acoustic Relationships in Swedish Vowel Sounds*, doctoral dissertation, Stockholm University.

Fant, G. (1960). *Acoustic Theory of Speech Production*, Hague, the Netherlands, Mouton.

Fant, G. (1973). *Speech Sounds and Features*, Cambridge, MA, MIT Press.

Flemming, E. (2005). "Speech Perception and Phonological Contrast," in D. Pisoni and R. Remez (eds.), *The Handbook of Speech Perception*, 156-181, Blackwell.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F. and Rizzolatti, G.

(2005) “Parietal lobe: From action organization to intention understanding,” *Science* 29, Vol. 308, no. 5722, 662 – 667.

Gallese V., Fadiga, L., Fogassi, L. and Rizzolatti, G (1996). “Action recognition in the premotor cortex,” *Brain* 119, 593 – 609.

Hanson G (1967). *Dimensions in speech sound perception*, Ericsson Technics 23.

Hume, E. & Johnson, K. (2001). “A model in the interplay of speech perception in phonology,” in E. Hume and K. Johnson (ed.), *The Role of Speech Perception in Phonology*, 3-26, Academic Press, New York.

Iverson, P. and Evans, B. G. (2007). “Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration,” *Journal of the Acoustical Society of America* 122(5), 2842-2854.

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*, Almqvist & Wilsell, Uppsala, (Reprinted in R. Jakobson, 1962, *Selected Writings I*, 328-401, Mouton, the Hague.)

Klatt, D. H. (1979). “Speech perception: A model of acoustic-phonetic analysis and lexical access,” *Journal of Phonetics*, 7, 279-312.

Klatt, D. H. and Stevens, K. N. (1969). "Pharyngeal consonants" *MIT QPR 93*, 207-219.

Klein, W., Plomp, R. and Pols, L. C. W. (1970). "Vowel spectra, vowel space, and vowel identification," *Journal of the Acoustical Society of America,* 48(4), 999-1009.

Krull, D. (1988). "Acoustic properties as predictors of perceptual responses: A study of Swedish voiced stops," *PERILUS* 7.

Krull, D. (1990). "Relating acoustic properties to perceptual responses: A study of Swedish voiced stops," *Journal of the Acoustical Society of America*, 88(6), 2557-2570.

Liljencrants, J. & Lindblom, B. (1972). "Numerical simulation of vowel quality systems: The role of perceptual contrast," *Language, 48*, 839-862.

Lindblom, B. (1963a). *On Vowel Reduction,* Fil. Lic. Thesis University of Uppsala, Rep. No. 29, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden.

Lindblom, B. (1963b). "Spectrographic Study of Vowel Reduction," *Journal of the Acoustical Society of America,* 35(11), 1773-1781.

Lindblom, B. (1986). “Phonetic universals in vowel systems,” in J. J. Ohala and J. Jaeger (Eds.), *Experimental Phonology*, 13-44, Orlando, FL, Academic Press.

Lindblom, B. (2003). “A numerical model of coarticulation based on a principal components analysis of tongue shapes,” *Proceedings of the 15th International Congress of the Phonetics Science*, CDROM ISBN 1-876346-48-5 © 2003 UAB.

Lindblom, B. (2007). “The target hypothesis, dynamic specification and segmental independence,” in B. L. Davis and K. Zajdó, Mahwah (eds.), *Syllable Development: The Frame/Content Theory and Beyond*, New Jersey, Lawrence Earlbaum Associates.

Lindblom, B., Diehl, R. & Creeger, C. (2006). “Do ‘Dominant Frequencies’ explain the listener’s response to formant and spectrum shape variations?”, submitted to special issue of *Speech Communication*.

Lindblom, B. & Sundberg, J. (1971). "Acoustical consequences of lip, tongue, jaw, and larynx movement," *Journal of the Acoustical Society of America, 50(4B)*, 1166-1179.

Luce, R. D. (1963). “Detection and recognition,” in R. D. Luce, R. R. Bush & E. Galanter (eds.), *Handbook of Mathematical Psychology,* Vol. 1, 103–189, New York, Wiley.

Maddieson, I. (1984). *Patterns of Sounds*, Cambridge, UK, Cambridge University Press.

Maddieson, I. and Precoda K. (1989). “Updating UPSID,” *Journal of the Acoustical Society of America,* 86*,* Suppl. 1, S19.

Martinet, A. (1955). *Économie des changements phonetiques*, Francke, Berne.

Martinet, A. (1964). *Elements of General Linguistics* (Elizabeth Palmer, Trans.), Chicago, University of Chicago Press. (Original work published 1960)

Moulton, W. G. (1962). “The vowels of Dutch: Phonetic and distributional classes,” *Lingua*, *11*, 294-312.

Ohala, J. J. (1980). “Moderator's summary of symposium on 'Phonetic universals in phonological systems and their explanation,” *Proceedings of the Ninth International Congress of Phonetic Sciences, 3*, 181-194, Copenhagen, Institute of Phonetics.

Ohala, J. J. (1981). "The listener as a source of sound change," in C. S. Masek, R. A. Hendrick, & M. F. Miller (eds.), *Papers from the Parasession on Language and Behavior*, 178 – 203, Chicago, Chicago Linguistic Society.

Ohala, J. J. (1993). "The phonetics of sound change," in Charles Jones (ed.), *Historical Linguistics: Problems and Perspectives,* 237-278, London, Longman.

Plomp, R. (1970). "Timber as multidimensional attribute of complex tones," in R. Plomp and G. F. Smoorenburg (eds.), *Frequency Analysis and Periodicity Detection in Hearing*, Sijthoff, Leiden.

Rizzolatti, G. and Arbib, M. A. (1998). "Language within our grasp," *Trends in Neuroscience,* 21, 188-194.

Schwartz, J.-L, Boë, L.–J., Vallée, N. & Abry, C. (1997). "Major trends in vowel system inventories," *Journal of Phonetics, 25 (3)*, 255-286.

Shepard, R. N. (1972). "Psychological representation of speech sounds," in P.B. Denes & E. E. David Jr. (eds.) *Human Communication: A unified View*, 67-113, New York, McGraw-Hill.

Sidwell, A and Summerfield, Q. (1986). “The auditory representation of symmetrical CVC syllables,” *Speech Communication,* 5, 283-297.

Singer, T., Seymour, B., O’Doherty, J., Kaube, H., Dolan, R. J. and Frith, C. D. (2004). “Empathy for pain involves the affective but not sensory components of pain,” *Science,* 20, Vol. 303, no. 5661, 1157 – 1162.

Skoyles, J. R. (2000). “Gesture, language origins, and right handedness,” *Psycoloquy,* 11(024), Language Gesture (2).

Stark, J., Ericsdotter, C., Branderud, P., Sundberg, J., Lundberg, H.-J. & Lander, J. (1999). "The APEX model as a tool in the specification of speaker-specific articulatory behaviour." *Proceedings of the XIVth ICPhS*, 2279-2282, San Francisco, California, Aug. 1-7.

Stevens, K. N. (1998). *Acoustic Phonetics*, Cambridge, MA, MIT Press.

Stevens, K. N. and Blumstein, S. (1981). “The search for invariant acoustic correlates of phonetic features,” in Eimas and Miller (eds.), *Perspectives on the Study of Speech,* 1-38, Erlbaum Assoc., New Jersey.

Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (1991). “An investigation of locus equations as a source of relational invariance for stop place categorization," *Journal of the Acoustical Society of America*, 90(3), 1309-1325.

Sussman, H. M., Fruchter, D., Hilbert, J. and Sirosh, J. (1998). "Linear correlates in the speech signal: The orderly output constraint," *Behavioral and Brain Sciences*, 21, 241-299.

Ten Bosch, L. F. M. (1991). *On the Structure of Vowel Systems: Aspects of an extended vowel model using effort and contrast*, Ph. D. thesis, U of Amsterdam.

Wang, W. S. -Y. (1968). *The basis of speech*. Project on Linguistic Analysis Reports, University of California at Berkeley. (Reprinted in *The Learning of Language*, ed. by C. E. Reed)

Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V. and Rizzolatti, G. (2003). "Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust," *Neuron*, Vol. 40, 655–664.

VITA

Sang-Hoon Park was born in Seoul, Korea, on January 21, 1969, the eldest son of Myung Soon Lee and Joon Bong Park. After completing his work at Jaehyeon High School, he entered Yonsei University in Seoul, Korea. He majored in English language and literature and received his Bachelor of Arts in February 1991. He entered the Master's Program in English linguistics at Yonsei University in Seoul, Korea. He earned Master of Arts with a thesis titled "A Study on the Meaning of English Generic Expressions." He served on Korean Navy and completed his service as a lieutenant junior grade in 1996. While he was in the navy, he taught English at Korea Naval Academy as a lecturer from 1994 to 1995, and as a full-time instructor from 1995 to 1997. He taught English at Yonsei University at Weonju from 1998 to 2000 and at Yonsei University at Seoul from 1999 to 2000 as a part-time instructor. He entered the Ph.D. program in linguistics at the University of Texas at Austin in August 2000. He was a teaching assistant working as a phonetics lab manager from 2002 to 2003. He taught Korean in Asian Studies department at the University of Texas at Austin as an Assistant Instructor from 2003 to 2006. In fall 2007, he worked as a lab manager in Auditory Cognition and Speech Perception Lab.

Permanent Address: 3500 Greystone Dr. Apt. 212, Austin, Texas 78731

This dissertation was typed by the author.