ACOUSTIC AND PERCEPTUAL CORRELATES OF PHARYNGEAL AND UVULAR CONSONANTS

by

Abeer Abdul-Hussain Alwan

B.S.E.E., Northeastern University
(1983)

Submitted in partial fulfillment
of the requirements
for the degree of

Master of Science

at the

Massachusetts Institute of Technology

May, 1986

Signature of Author . . . .
Department of Electrical Engineering and Computer Science
May 12, 1986

Certified by . . . .
Kenneth N. Stevens
Thesis Supervisor

Accepted by . . . .
Arthur C. Smith
Chairman, Departmental Committee on Graduate Students

ACOUSTIC AND PERCEPTUAL CORRELATES OF PHARYNGEAL AND UVULAR CONSONANTS

by

Abeer Abdul-Hussain Alwan

Submitted to the Department of Electrical Engineering and Computer Science on May 12, 1986 in partial fulfillment of the requirements for the degree of Master of Science

ABSTRACT

This thesis investigates aspects of the production, the acoustics, and the perception of pharyngeal and uvular consonants. First, we introduce theoretical models of the vocal-tract area function during the production of these consonants. From these models we calculate the formant frequencies and the contributions of different vocal-tract losses (localized losses: due to the impedances of the glottis, the constriction, and the radiation; and distributed losses: due to heat conduction and viscosity, and to the impedance of the walls) to the bandwidths of the formants for both an open- and a closed-glottis case (voiced and voiceless, respectively). The presence of a noise source, modeled as a series pressure source, in the vicinity of a supraglottal constriction introduces zeros to U_o/p_s (transfer function from the volume velocity at the lips to the pressure source near the constriction). The zeros are in the vicinity of the back-cavity resonances (including the Helmholtz resonance). The location of the zero which is in the vicinity of the Helmholtz resonance is highly sensitive to the pressure-source location. Consequently, this resonance may or may not be cancelled. Other back-cavity resonances, on the other hand, are cancelled by zeros regardless of the pressure-source location.

Predictions based on the theoretical study were: (1) F1 for pharyngeals should be higher than that for uvulars, F3 should be lower, F2 should be approximately the same for both when the glottis is closed and should be higher for pharyngeals when the glottis is open; (2) for the pharyngeals F2 should be a Helmholtz resonance, and F1 and F3, front-cavity resonances, and for the uvulars, F1 should be a Helmholtz resonance, F2 and F4, front-cavity resonances, and F3, a back-cavity resonance; (3) for the closed-glottis case the Helmholtz resonances for both classes of consonants should be widened, compared to a no-constriction-loss case, due to constriction losses; and (4) in the case of noise generation in the vicinity of a supraglottal constriction the front-cavity resonances should be strongly excited and the Helmholtz resonances may or may not be excited depending on the noise-source location.

Secondly, an acoustic analysis of five Arabic consonants (two pharyngeal /ʕ,ħ/ and three uvular /ʁ,χ,q/) prevocalically with three vowels (/aa,ii,uu/) was carried out. Results of the analysis were in general agreement with the theoretical predictions. Spectral analysis reveals a compact spectral shape (narrow peaks in the mid-frequency region) for the voiceless uvular fricative /χ/ and stop /q/, whereas the spectrum for the voiceless pharyngeal /ħ/ is characterized by broad peaks at high frequencies. Both a continuant and a non-continuant allophone were found for the voiced consonants. Durational measurements of the consonantal intervals for the voiced and voiceless consonants were similar for both classes, and the voiceless consonants were longer than their voiced counterparts. Measurements of the fundamental frequency (f0) for the voiced consonants /ʕ,ʁ/ show a lower f0 than the adjacent vowels; this is attributed to the constricted pharynx in the former case, and to the acoustic and aerodynamic effects of introducing a narrow supraglottal constriction in the latter. Results of the analysis also show that, for each consonant, the F1 "target" is influenced by the height of the following vowel, whereas the F2 and F3 "targets" are influenced by the backness and rounding of the following vowel.

Thirdly, we investigate the perceptual cues for place of articulation for the voiced consonants /ʕ/ and /ʁ/ through perceptual experiments, using synthetic /Caa/ stimuli. Results show that the onset value of F1 ($F1_o$) is essential in discriminating between the two consonants, while F2 position and bandwidth are not. An $F1_o$ equal to or greater than the F1 in the steady state of the vowel ($F1_s$) results in the perception of the pharyngeal /ʕ/. When $F1_o$ is at least 130 Hz less than $F1_s$, the uvular /ʁ/ is perceived. Other values of $F1_o$ result in the perception of the glottal stop /ʔ/. Widening F1 bandwidth increases the uvular responses and improves the naturalness of the uvular stimuli, whereas it decreases substantially the pharyngeal responses. The increased bandwidth was predicted from the theoretical study.

Based on the results of the theoretical, acoustical, and perceptual studies we propose a set of binary features describing the two classes of consonants.

Thesis Supervisor: Kenneth N. Stevens
Title: Clarence J. LeBel Professor of Electrical Engineering

To my father and my mother

ACKNOWLEDGMENTS

My deepest gratitude goes to Professor Kenneth Stevens for his constant encouragement, critical comments and patience. During the two years that I have known Ken, he has been a true mentor and a valuable teacher to me.

Many thanks to all members of the Speech Communication Group for providing a stimulating and friendly atmosphere. I am deeply grateful to Corine Bickley for proofreading this thesis and for her support all along.

Special thanks to Maria-Gabriella Di Benedetto for her help in putting this thesis together and for her friendship, which has meant a great deal to me.

I thank all the subjects who participated as speakers and/or listeners (JM, HA, LT, MU, MA, BA, MR, MAL, and AM) for being so patient.

I would also like to thank my friends and relatives here and abroad who never failed to encourage me throughout my studies through their letters, "long-distance" phone calls, and care.

Finally, I want to thank my parents, sister, and brother; their faith and love has made this thesis possible.

This research was supported in part by an NSF grant.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem Statement

Consonants and vowels comprise the two major classes of speech sounds. While vowels are produced with a relatively unconstricted vocal tract, consonants are produced with a constriction at some point along the vocal tract. For most classes of consonants, the constriction is narrow enough to cause complete stoppage of the air or to produce a turbulent noise source (Jakobson, Fant and Halle, 1963). The acoustic properties of consonants depend mainly on the location and the cross sectional area of the constriction. The location of the constriction, which defines the place of articulation of consonants, could be anywhere between the glottis and the lips.

In English, the place of articulation of consonants ranges between the ' velar ' and the ' bilabial ' places corresponding to constriction locations in the oral cavity between the velum and the lips. Consonants produced with a constriction in the pharyngeal and the uvular regions between the velum and the glottis are found in some other languages.

There are known to be two pharyngeal consonants /ʕ/ and /ħ/ (voiced and voiceless, respectively) and three uvular consonants /ʁ/,/χ/ and /q/ (voiced, voiceless fricative, and voiceless stop, respectively). All five of these consonants are found in Arabic.[1] For

[1] From an inventory of 317 languages examined by Maddieson (1984), the voiced pharyngeal consonant is found in 9 languages,the voiceless in 13; the voiced uvular is found in 13 languages, the voiceless fricative in 29, and the voiceless uvular stop is found in 38 languages.

this reason previous researchers have focused attention on the Arabic language in order to study the properties of pharyngeal and uvular consonants; this thesis will further examine the pharyngeal and uvular consonants of Arabic.

Studying the relations between the production mechanisms and the acoustic properties of the speech waveforms of these consonants would help in developing a more detailed quantitative acoustic theory of speech production. W. Meyer-Eppler (1953), G. Fant (1960), J. M. Heinz and K.N. Stevens (1961) , and C. Shadle (1985) have theoretically formulated these relations for fricative consonants in English, Swedish, and German. Although the resulting theoretical models were based upon simplified assumptions regarding the production mechanisms involved, they were essential in clarifying which parameters (e.g., length of the front cavity, constriction location, etc. . .) are significant acoustically. However, none of the consonants analyzed was one with a constriction in the pharyngeal region. Similarly, at the perceptual level, no study has examined the perceptual correlates of uvular and pharyngeal consonants. Hence, it would be of interest to develop a theoretical model that would help us in understanding the mechanisms involved in producing these sounds and the parameters which are significant acoustically and perceptually.

The following is a summary of previous studies that dealt with the production, acoustics, perception and the phonology of pharyngeal and uvular consonants.

1.2 Literature Survey

1.2.1 Articulatory Mechanisms

Several researchers have examined X-ray films to determine the location and shape of the various articulators in the vocal tract during the production of pharyngeal and uvular consonants:

- Figure 1.1 (Al-Ani, 1970) shows tracings of X-ray films of a speaker (Iraqi) producing the pharyngeal consonants /ʕ/ and /ħ/ preceding the vowels /a/, /i/ and /u/. These tracings do not show the position of the tongue root, and hence it is difficult to conclude whether or not the two consonants are produced with a similar constriction location. Further investigations by Al-Ani (1985) indicate that the place of articulation for the voiceless pharyngeal /ħ/ is higher in the pharynx than that of /ʕ/. Similarly, he claims that the the three uvular consonants /ʁ/, /χ/ and /q/ differ in place of articulation with /q/ being the furthest back and /χ/, the furthest front.

- Figure 1.2 (Delattre, 1971) shows tracings of X-ray films illustrating a profile view of a speaker (Lebanese) producing pharyngeal and uvular consonants in initial position before /a/,/i/ and /u/. The frame on the left of each row shows the moment of maximal constriction; on the right of each row are sketches of the articulatory configurations of the following vowels. From these X-rays, Delattre concluded that the constriction location during the production of the voiceless pharyngeal /ħ/ is further back than that of the voiced pharyngeal /ʕ/. The uvular consonants /ʁ,χ,q/ seem to share the same place of articulation, differing only in the degree of constriction. The consonant /χ/ is produced with a more constricted tract than /ʁ/, and /q/ is produced with the most constricted. These findings, with regard to relative place of articulation within each group of sounds, are different from those of Al-Ani.

- X-ray tracings by Ghazeli (1977), in Figure 1.3a (North African speaker), show that the place of articulation of the two pharyngeal consonants /ʕ,ħ/ is similar. The center of the main constriction is approximately 3.5 *cm* from the glottis, at the level of the epiglottis. However, the constriction between the epiglottis and the back wall of the pharynx was observed to be narrower for /ħ/ than for /ʕ/ (3 *mm* vs. 4 *mm*), and was sustained 20-30 msec longer than for /ʕ/. The larynx was raised approximately 9 *mm* from rest position (7 *mm* from its position during speech) in both cases. Another narrowing of the vocal tract occurred 6 *cm* from the lips. This narrowing was formed by the portion of the tongue between the blade and the dorsum, and was of width 8 *mm* (from the hard palate). Uvular consonants (Figure 1.3b) were produced with different constriction locations in the uvular region: for /ʁ/ the location was midway between /q/ (the furthest back) and /χ/. A wider (4-5 *mm*) secondary constriction occurred at the level of the epiglottis. No larynx movements were observed for /ʁ/, whereas the larynx was raised 2 *mm* and 2.5 *mm* from rest position during the production of /χ/ and /q/, respectively.

Ghazeli also measured the air-flow rate for the voiceless fricatives /ħ/ and /χ/. He found that the average flow rate was 44 liters/min (733 cc/sec) and 29 liters/min (483 cc/sec) for /ħ/ and /χ/, respectively. He attributed the increase in the flow rate for the voiceless pharyngeal to a wider constriction and/or an increase in respiratory effort during the production of /ħ/.

In summary, these researchers seem to agree that the place of articulation is in the laryngopharynx region at the level of the epiglottis for the pharyngeal consonants, and is at the level of the uvula for uvular consonants. The constriction in the former case is formed by backing the tongue root toward the back wall of the pharynx, and in the latter by backing the tongue dorsum toward the uvula. However, there seems to be a disagreement as to whether or not the two pharyngeal consonants /ʕ,ħ/ or the three uvular consonants /ʁ,χ,q/ are produced with a similar constriction location. Dialectal differences might account for this disagreement, because the speakers in each case were from a different country.[2]

[2]Ghazeli (1977) claims that pharyngeal consonants are produced in an identical manner in all dialects; Al-Ani (1985), on the other hand, believes that /ʕ/, in particular, is produced differently in various regions.

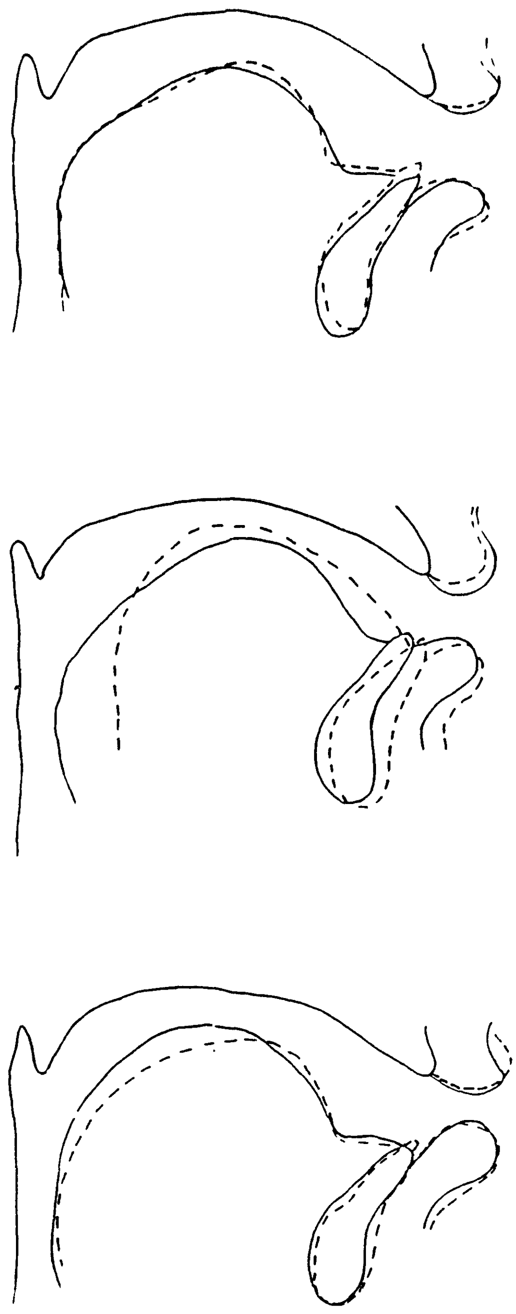


Figure 1.1: Tracings of X-ray films of the pharyngeal consonants /ʕ/ (solid line) and /ħ/ preceding the vowels /a/, /i/, and /u/, respectively (Al-Ani, 1970).

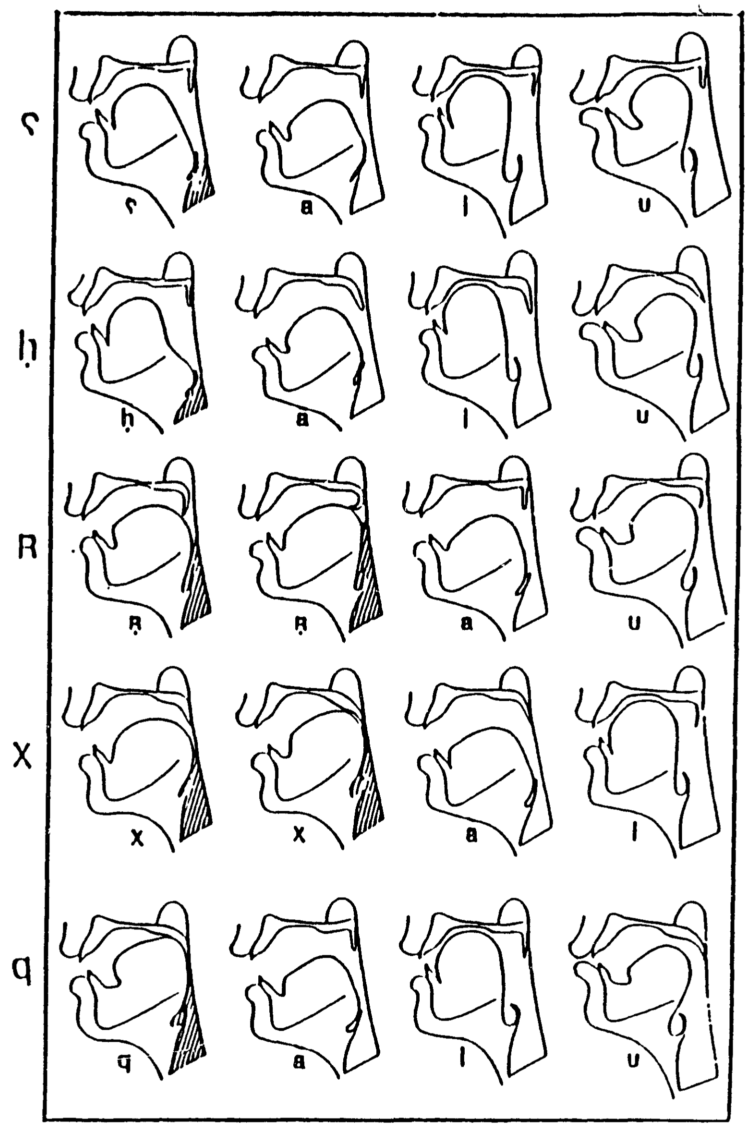


Figure 1.2: Tracings of frames taken from X-ray films showing a profile view of a speaker during the production of pharyngeal and uvular consonants in initial position before /a/,/i/, and /u/. The frame on the left of each row shows the moment of maximal constriction; on the right of each row are sketches of the articulatory configurations of the following vowels (Delattre, 1971).

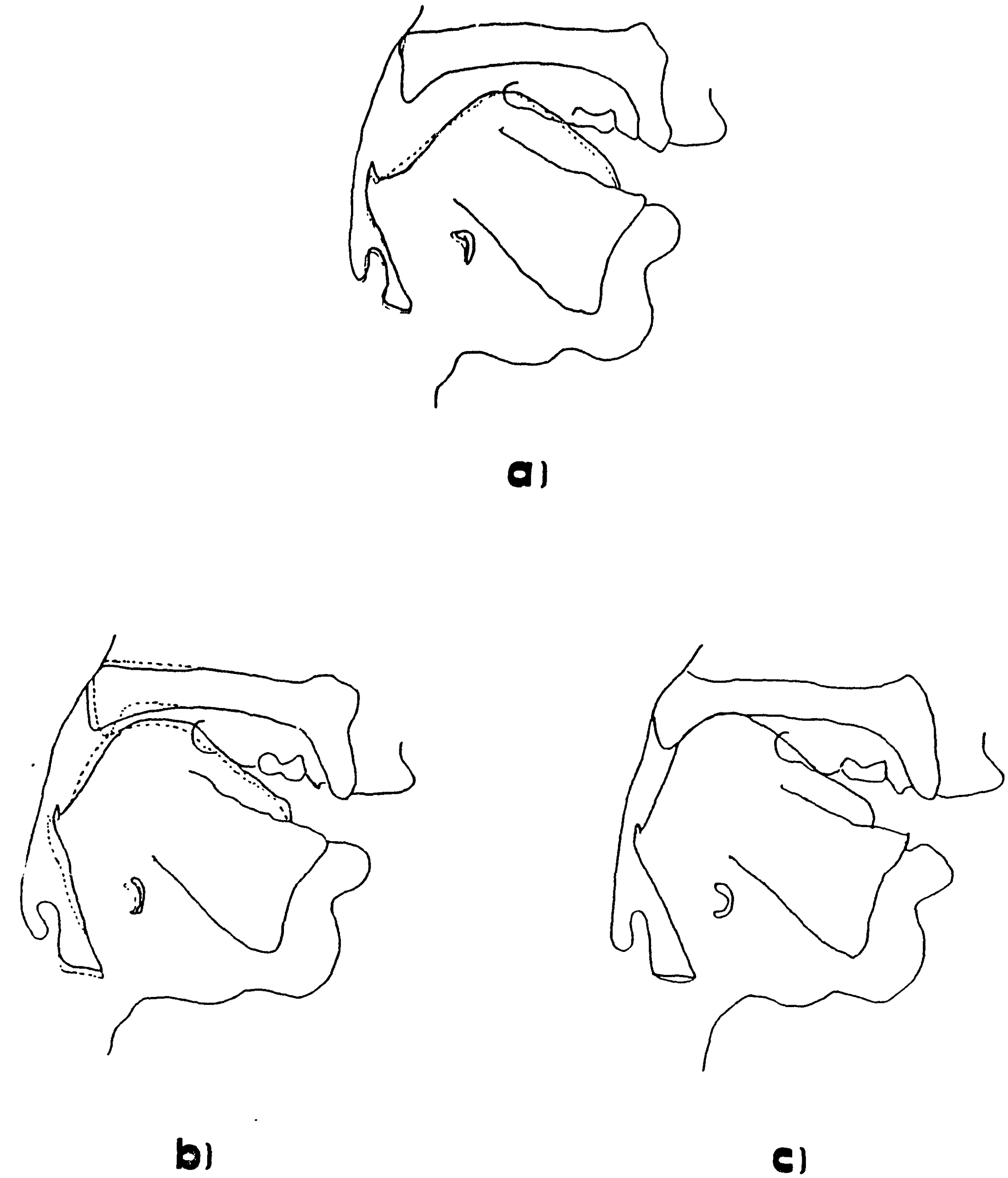


Figure 1.3: Vocal-tract shape during the articulation of a) /ʕ/ (dotted line) and /ħ/, b) /ʁ/ (dotted line) and /χ/, and c) /q/. The consonants are imbedded in the word /Caeli/, where C represents one of the five consonants /ʕ,ħ,ʁ,χ,q/ (Ghazeli, 1977).

1.2.2 Acoustics: Theoretical Modelling and Analysis

Klatt and Stevens (1969), introduced a theoretical model of the vocal-tract area function during the production of pharyngeal and uvular consonants. An idealized model (Figure 1.4a) was used to examine the acoustic effect of creating a narrow constriction at a point along the pharyngeal tract. The length of the back cavity (d) during the production of these consonants is within the range 3-7 *cm*. A length of d=3-4 *cm* is appropriate for the production of the consonants with the more posterior constriction position (pharyngeal /ʕ,ħ/), while a length of d=7 *cm* is appropriate for the consonants with the more anterior constriction position (uvular /ʁ,χ, q/).

Klatt and Stevens calculated the four lowest natural frequencies of the model as a function of d (length of the back cavity) and the cross-sectional area of the constriction (Figure 1.4b). Spectrograms of natural utterances spoken by a native speaker of Lebanese verified the predictions derived from the model (Figure 1.4a) regarding formant-cavity affiliation: F3 is a front-cavity resonance in the case of pharyngeal consonants, and F2 and F4 are front-cavity resonances in the case of uvular consonants. This was clear in the spectrograms of the voiceless fricatives /ħ,χ/, where noise strongly excited the formants that are front-cavity resonances. An explanation of excitation of formants by noise generated at a constriction, will be presented in Chapter 2. The discussion will be based on an acoustic model of the vocal tract during the production of consonants that are produced with a noise source at the constriction.

The important acoustic correlate attributed to pharyngeal consonants is a high F1; the F1 position during the articulation of pharyngeal consonants is the highest among all sounds in Arabic including the vowel /a/ (Klatt and Stevens, 1969; Ghazeli, 1977). Acoustic analyses by several researchers of the voiced pharyngeal consonant /ʕ/ has led to the conclusion that the most likely realizations of this consonant are as stops (Al-Ani, 1970) and approximants (Catford, 1968; Adamson, 1981). The acoustic analysis by Klatt and Stevens showed no evidence of noise in the spectrograms of /ʕ/, suggesting that this consonant could be categorized as a sonorant rather than as a fricative. Evidence of lower fundamental frequency than that of the vowels, laryngealization, and creakiness during the production of /ʕ/ was noted by Ghazeli (1977). This has led to his agreement with Ladefoged's (1975) suggestion that some intralaryngeal adjustments might occur during the production of pharyngeal consonants. In summary, /ʕ/ has several allophonic realizations due to its complex production mechanisms. This fact has led to an unresolved problem of categorization of /ʕ/.

Spectrographic analysis of the voiced uvular consonant /ʁ/ (Klatt and Stevens, 1969; Ghazeli, 1977) showed that this consonant is characterized by a clear formant structure. Evidence of weak noise excitation was observed by Al-Ani (1970); the noise excited F3 and higher formant frequencies. These observations suggest that two allophonic realizations are commonly found for /ʁ/: a sonorant allophone and a weak-fricative allophone. In addition, Ghazeli noted that in Cairo Arabic, /ʁ/ and /χ/ are produced with a velar rather than a uvular place of articulation, and Al-Ani noted that in Iraqi Arabic these two consonants (/χ/ with the vowel /i/ only) are produced with a velar place of articulation. The analyses of the voiceless consonants /ħ,χ/ have shown that these consonants have the acoustic attributes of fricatives, and no different allophonic realizations are found.

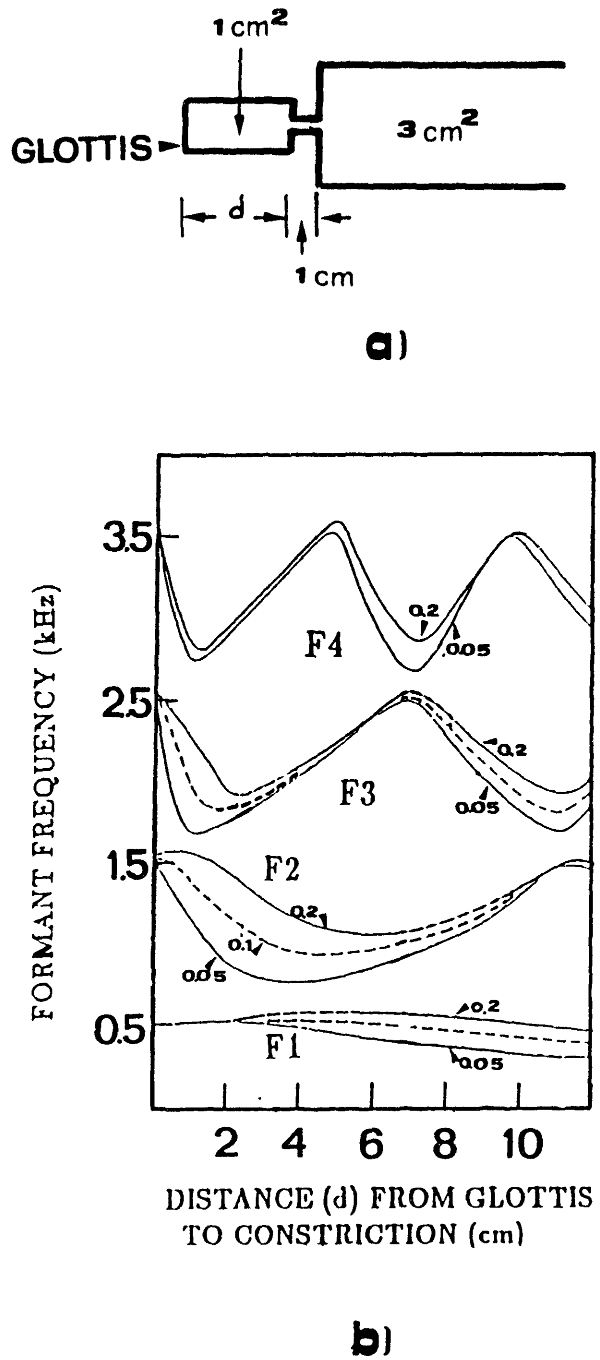


Figure 1.4: a) Idealized model of the vocal-tract area function with a narrow constriction at a point along the pharyngeal tract (the glottis is at the left end) . b) Plot of the four lowest natural frequencies as a function of d (length of the back cavity) and the cross-sectional area of the constriction (the dashed line corresponds to a constriction area of 0.1 cm^2) (Klatt and Stevens, 1969).

1.2.3 Perception

To the knowledge of the author, no study has been done to examine the perceptual correlates of pharyngeal and uvular consonants. Delattre (1971) claimed that formant transitions (into and out of the consonant) provide the important perceptual cue for the uvular fricatives /ʁ/ and /χ/. His claim was based on informal perceptual experiments done with synthetic stimuli. However, neither information about the synthetic stimuli and the procedures with which the experiments were carried out, nor quantitative results were reported.

The only systematic perceptual study of some Arabic consonants was done by Obrecht (1968), using synthetic stimuli. The results of his experiments show that the second formant transition is essential in discriminating velarized from non-velarized consonants in Arabic: the "locus" of F2 for the velarized consonants is lower than that of their non-velarized counterparts.[3] Because pharyngeal and uvular consonants are not considered to be the velarized counterparts of other consonants, they were not included in his study.

1.2.4 Phonological Considerations

The earliest descriptions of the Arabic phonological system were given by Sibawayhi in his book *al-Kitab* (750/1975), and later by Avicenna in *Makharij al-Huruf* (1333/1916). Both described the pharyngeals /ʕ,ħ/ as sharing a similar place of articulation in the lower pharynx. The voiceless pharyngeal was described as a fricative, whereas the voiced pharyngeal was described by both as being a frequentative.[4]

The two consonants /ʁ,χ/ were described as being a voiced fricative and a voiceless fricative, respectively, with a similar place of articulation in the upper pharynx. The consonant /q/ was described as a voiced stop with a place of articulation further front than /ʁ,χ/. Sibawayhi then distinguished two classes of consonants according to their

[3] Obrecht defines velarized consonants as those with a secondary place of articulation in the pharynx (pharyngealization). The minimal pairs chosen for his experiments were: (b,ḅ), (d.ḍ), (t,ṭ), (z,ẓ), (s,ṣ), (m,ṃ), (n,ṇ), (l,ḷ) and (r,ṛ). The synthetic stimuli used were CV syllables.

[4] Sibawayhi defines /ʕ/ as being "intermediate between a fricative and a stop, being a frequentative because of its similarity to /ħ/" (Vol. 4, p. 435).

place of articulation. The consonants /ʔ,h,ʕ,ħ,ʁ,χ/ were called "consonants of the pharynx" and all other consonants were called "consonants of the mouth".

Jakobson (1957) agreed with the general phonological description given by the early Arab grammarians. He described the Arabic consonants, as pronounced in Palestinian Arabic, by binary features (Table 1.1). Both pharyngeals /ʕ,ħ/ were called glides (non-consonantal, non-vocalic) with the *fortis/lenis* feature being distinctive between the two. The pharyngeals were deprived of the features *grave/acute* and *compact/diffuse* as these features are "generated in the mouth cavity" (p. 114). The consonants /ʁ,χ/ were described as uvular continuants and the voiceless stop /q/ was described as a voiceless pharyngealized velar. The feature attributed to all five consonants, and pharyngealized consonants, by Jakobson was the *flat* feature. In this case, *flat* indicates that a primary or secondary articulation takes place in the pharynx region during the production of the consonants.

In a more general linguistic framework, Chomsky and Halle (1968) described the two classes of sounds, pharyngeal and uvular, in terms of distinctive features as (−*anterior*, −*coronal*, −*high*, +*back*) with the *low* feature distinguishing the two classes; pharyngeals have the feature (+*low*) and uvulars, (−*low*). In their framework, the features *low* and *back* reflect the *compactness* and *gravity* of these sounds, respectively. Hence, this classification is in disagreement with Jakobson's (1957), in which the pharyngeals and uvulars were deprived of the feature *grave*, the pharyngeals were also deprived of the feature *compact*, and the uvulars had the feature *compact*.

Other phoneticians and linguists have agreed on characterizing /ħ,ʁ,χ/ as fricatives, and /q/ as a voiceless stop, in most dialects. The description of the manner of articulation for the voiced pharyngeal /ʕ/ is still not clear, mainly because of its different allophonic realizations (see Section 1.2.2).

It is hoped that this study will clarify the categorization of these sounds in terms of their distinctive features.

	ɔ	c	h	ḥ	d	t	ṭ	δ	δ̣	ə	z	s	ṣ	b	ḅ	f	(g)	k	q	ž	Γ	š	x	n	m	ṃ	r	l	ḷ	/	w
vocalic vs. non-voc.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
cons. vs. non-cons.	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
flat vs. plain	-	+	-	+	-	-	+	-	+	-	-	-	+	-	+	-	-	-	+	-	+	-	+	-	-	+	-	-	+	o	o
nasal vs. oral	o	o	o	o	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	o	o	o	o	o
compact vs. diffuse	o	o	o	o	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	o	o	o	o	o	o	o	o
grave vs. acute	o	o	o	o	-	-	-	-	-	-	-	-	-	+	+	+	o	o	o	o	o	o	o	-	+	+	o	o	o	-	+
fortis vs. lenis	-	-	+	+	-	+	+	-	-	+	-	+	+	-	-	+	-	+	+	-	-	+	+	o	o	o	o	o	o	o	o
continuant vs. abrupt	o	o	o	o	-	-	-	+	+	+	+	+	+	o	o	o	o	-	-	o	o	+	+	o	o	o	-	+	+	o	o
strident vs. mellow	o	o	o	o	-	o	o	-	-	-	+	+	+	o	o	o	-	o	o	+	+	o	o	o	o	o	o	o	o	o	o

Table 1.1: A description of the Arabic consonants by binary features (Jakobson, 1957).

1.3 Thesis Outline

The literature survey indicates a lack of good understanding of the production mechanisms and of the acoustic and perceptual properties specific for pharyngeal and uvular consonants. Consequently, the phonetic categorization of these consonants is yet to be clarified.

In the course of this thesis, we will investigate properties of pharyngeal and uvular consonants from three points of view: theoretical study (Chapter 2), acoustic analysis (Chapter 3), and perceptual experiments (Chapter 4).

In Chapter 2, we introduce theoretical models of the vocal-tract area function during the production of pharyngeal and uvular consonants. The dimensions of these models are based on anatomical and physiological data and constraints. From these models we calculate the formant frequencies and their bandwidths for both the voiced and the voiceless consonants. We also study the acoustic effects of introducing a noise source in the vicinity of a supraglottal constriction.

In Chapter 3, we discuss the validity of the theoretical predictions arrived at in Chapter 2 through acoustic analysis of prevocalic pharyngeal and uvular consonants in Arabic. We choose analysis techniques appropriate for each consonant and quantify our findings. The analysis is in the time domain (duration) and in the frequency domain (fundamental frequency, formant trajectories, formant bandwidths, and spectral shape).

In Chapter 4, we aim to find the relevant perceptual cues for identifying the place of articulation of pharyngeal and uvular consonants. Perceptual experiments are carried out using synthetic /Caa/ stimuli, where /C/ is one of the two voiced consonants: pharyngeal /ʕ/ or uvular /ʁ/.

Finally, in Chapter 5 we summarize the results of Chapters 2, 3, and 4 and attempt to describe these consonants in terms of their distinctive features. Suggestions for future work will conclude this thesis.

Chapter 2

Theoretical Considerations

In this chapter we will develop theoretical models of the vocal tract during the production of pharyngeal and uvular consonants. From these models we will calculate the formant frequencies for each configuration, determine the formant-cavity affiliation, and examine the acoustic effects when a noise source is present at the constriction. We will also calculate the contribution of the different losses in the vocal tract to the bandwidths of the calculated formant frequencies.

2.1 Idealized Models: Dimensions

An accurate model of the vocal tract during the production of any sound requires knowledge of the area function at each point along the tract. Due to the lack of reliable X-ray data describing the vocal tract during the production of pharyngeal and uvular consonants we will develop idealized models of the tract during the production of these consonants, taking into consideration certain anatomical and physiological data and constraints.

A simplified model the vocal-tract area function in the presence of a supraglottal constriction is shown in Figure 2.1a. The figure shows the vocal tract divided by the constriction into a back and a front cavity. In choosing the dimensions appropriate for the production of the pharyngeals and the uvulars, the following were considered:

- l (length of the vocal tract): X-ray data from Ghazeli (1977), see Section 1.2.1, show that the position of the larynx is 7 mm higher for pharyngeal consonants than it is for uvulars. Thus, if we assume a vocal-tract length of 17 cm, appropriate for a male's tract, for the uvulars, then according to Ghazeli's data, the length of the tract for the pharyngeals should be 16.3 cm.

- l_b, l_c, l_f (lengths of the back cavity, constriction, and the front cavity): Ghazeli's data also show that the location of the constriction is at the level of the epiglottis in the case of pharyngeal consonants (3 - 3.5 cm above the glottis), and at the level of the uvula (8 cm above the glottis) for uvular consonants. Hence, l_b was chosen to be 3 cm and 8 cm for the pharyngeals and the uvulars, respectively. We do not have accurate data on the length of the constriction (l_c); so we will leave this parameter as a variable with two values: 1 and 2 cm. After choosing l, l_b, and l_c then the length of the front cavity (l_f) would simply be $l_f = l - l_b - l_c$.

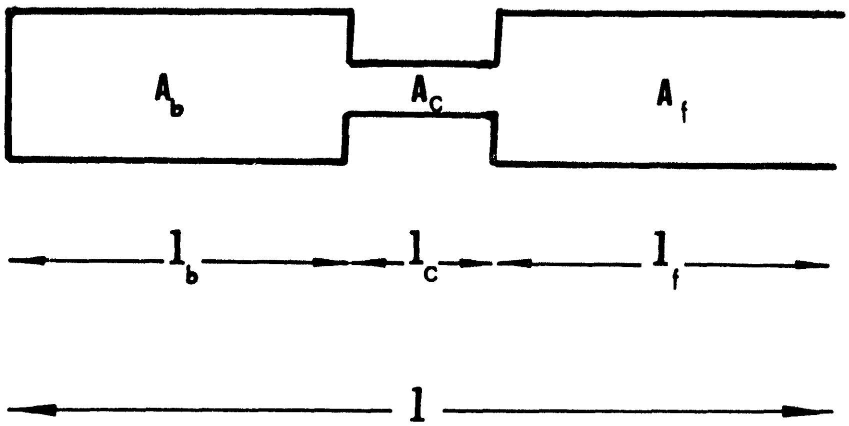
- A_b, A_f (Cross-sectional areas of the back and the front cavities): The choice of A_b and A_f was done in an ad-hoc manner; the area of the back cavity for the pharyngeal was chosen to be 1 cm^2 (similar to that used by Fant (1960) for the idealized model of the vowel /ɑ/). If we consider the total volume of the vocal tract to be approximately 65 cm^3,[1] then the area of the front cavity would be approximately 5 cm^2. For the uvulars, the area of the front cavity was chosen to be the same as that for the pharyngeals (5 cm^2), and the area of the back cavity to be 2 cm^2.

- A_c (Cross-sectional area of the constriction): The cross-sectional area of the constriction was assumed to be in the range of 0.15 to 0.25 cm^2. The upper limit on the area function (0.25 cm^2) is less than the minimum cross-sectional area of the constriction for vowels which was measured by Fant (1960) to be 0.3 cm^2.

The dimensions chosen for the models are summarized in Figure 2.1b, and the resulting models are shown in Figures 2.2a and 2.2b.

---

[1] The volume of a male's vocal tract was estimated by Stevens (in press) to be, on average, between 60 - 80 cm^3.



a)

	Pharyngeal	Uvular	
l	16.3	17	cm
l_b	3	8	cm
l_c	1, 2	1, 2	cm
l_f	13.3-l_c	9-l_c	cm
A_b	1	2	cm^2
A_c	.15, .2, .25	.15, .2, .25	cm^2
A_f	5	5	cm^2

b)

Figure 2.1: a) Idealized model of the vocal-tract area function in the presence of a supra-glottal constriction. The glottis is at the left end of the model. b) Parameters chosen for the pharyngeal and the uvular configurations. The subscripts b,c,f refer to the back cavity, constriction, and front cavity, respectively (see text for details).

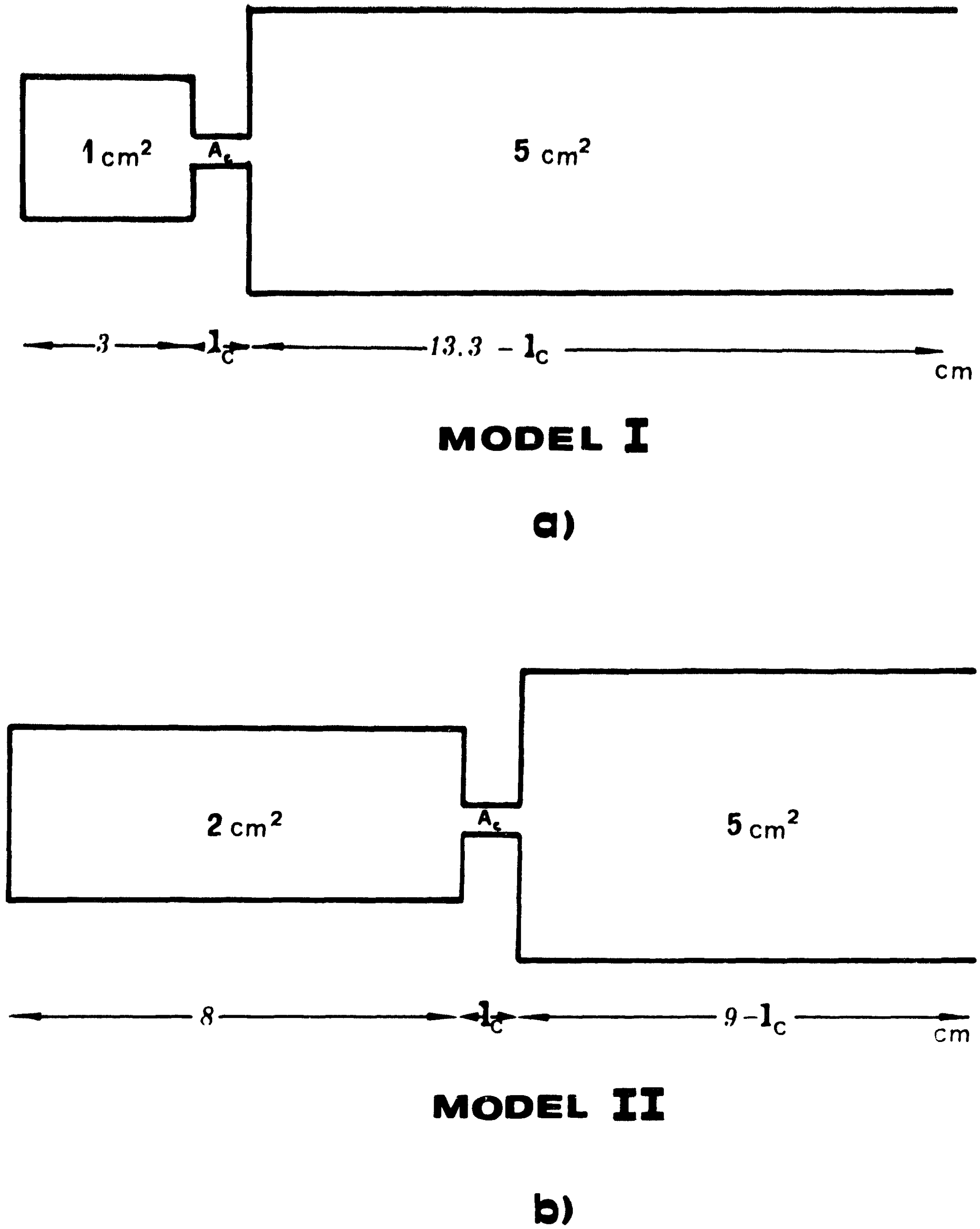


Figure 2.2: Idealized models for a) the pharyngeals (model I), and b) the uvulars (model II).

2.2 Idealized Models: Formant Frequencies

Formant frequencies were calculated from the area functions specified in models I, II (Figures 2.2a, 2.2b) using a program (TBFDA1) developed originally by Henke (1966) and modified by Hosein (1983). The program calculates the magnitude and the phase of the transfer function for an arbitrary shape and length of the vocal tract by considering the vocal tract to be a concatenation of arbitrary-length tubes, each having a fixed cross-sectional area. A formant frequency is determined when the phase of the transfer function changes from being greater than π (n+1)/2 to being less than that value. Formant frequencies were calculated taking into consideration the effects of the wall impedance (R_{wmech}=1060 $gcm^{-2}sec^{-1}$, M_{wmech}=1.5 gcm^{-2}), and assuming the lip impedance to be approximately that of a piston in a sphere.

The first four formant frequencies of the models were calculated for two cases: (1) an infinite glottal impedance (voiced), and (2) an open glottis with a cross-sectional area (A_g) of 0.1 cm^2 (voiceless).

2.2.1 Results and Discussion

Table 2.1 shows the calculated formant frequencies for the idealized models as functions of the length and the cross-sectional area of the constriction (l_c, A_c) for both the closed- and the open-glottis cases. For the pharyngeal model, F1, F3, and F4 are the first three front-cavity resonances, and F2 is a Helmholtz resonance. For the uvular model, F1 is a Helmholtz resonance, F2 and F4 are the first two front-cavity resonances, and F3 is the first resonance of the back cavity.

From the calculated values the following were observed:

- F1 and F4 for the pharyngeal model are higher than those for the uvular, and F3 is lower. This result is similar to that shown by Klatt and Stevens (1969). F2 for the pharyngeal is higher than that for the uvular for all the cases considered except when l_c is 2 cm and the glottis is closed (A_g is 0); in that case F2 values for the two models are approximately the same.

- The front-cavity resonances (F1, F3, and F4 for the pharyngeal model, and F2 and F4 for the uvular) are increased when either l_c or A_c increases. To understand why these effects occur, let us consider the front cavity to be a uniform tube open at one end (at the lips). The resonances of such a tube are:

$$f_{n+1} = \frac{c\,(2n+1)}{4\,l_{feff}} \tag{2.1}$$

Where

n= 0,1,2,...

c= velocity of sound in air

l_{feff}= effective length of the front cavity (taking into account the end correction introduced by the radiation impedance)

We notice from Eq. 2.1 that the front-cavity resonances vary inversely with l_{feff}; from the results we see that when l_c increases (and consequently l_{feff} decreases) these resonances are shifted upward.

For our models, the cross-sectional area of the constriction is not zero and therefore the front-cavity resonances will not be exactly the quarter-wavelength resonances shown in Eq. 2.1. The effect of the constriction is to introduce coupling between the front cavity and the rest of the vocal tract, and therefore shift the resonances upward, due to the acoustic mass of the constriction. The degree of coupling is related to the ratio of the cross-sectional area of the constriction to that of the front cavity (A_c/A_f), and this ratio determines the amount by which each frequency is shifted. This clarifies the reason for the increase in the front-cavity resonances when A_c is increased (keeping A_f fixed).

- An increase in A_c or a decrease in l_c causes the Helmholtz resonance (F2 for the pharyngeal model and F1 for the uvular) to increase. This result is rather intuitive since the Helmholtz resonance could be approximated by :

$$f_h \cong \frac{c}{2\pi}\sqrt{\frac{A_c}{V_b l_c}} \tag{2.2}$$

Where

A_c= cross-sectional area of the constriction

l_c= length of the constriction

V_b= volume of the back cavity

From this equation we notice that an increase in A_c or a decrease in l_c would shift the Helmholtz resonance upward.[2]

- For both models, the calculated formant frequencies for the open-glottis case are higher than those for the closed-glottis, due to the reactive part of the glottal impedance. The shift is greatest for the Helmholtz resonance (30% in F2 for the pharyngeal, and 11% in F1 for the uvular).

Formant frequencies measured from natural utterances by Ghazeli (1977) and Al-Ani (1970) show that the voiced pharyngeal F1 is in the range 700-900 Hz, F2, 1250-1400 Hz, F3, 2200-2300 Hz, and for the voiceless pharyngeal F2 is 1700 Hz and F3, 2300 Hz. For the voiced uvular, F1 was measured to be 500-600 Hz, F2, 1200-1300 Hz, F3, 2300-2600 Hz, and no formant frequencies were reported for the voiceless uvular. In order to obtain formant frequencies that match these data, we tapered the junctions between each cavity and the constriction (Figure 2.3). The tapering allows us to develop more realistic models of the vocal tract during the production of these consonants, since we are avoiding sharp discontinuities in the area function.

The formant frequencies were recalculated and the results are shown in Table 2.2. The same conclusions drawn from the uniform-tube models apply here, in terms of the relative values of the formant frequencies for both the pharyngeal and the uvular consonants, and the formant-cavity affiliation. From these results, we notice that the calculated frequencies are closely matched to the measured values for the voiced consonants, reported by Ghazeli and Al-Ani, when l_c is 1 cm and A_c 0.25 cm^2 (A_g=0). For the voiceless pharyngeal, the model gives a better match to the measured data when A_c=0.15 cm^2, and l_c=1 cm (A_g=0.1 cm^2).

[2]Equation 2.2 is an approximation to the Helmholtz resonance since it does not take into account the wall effects, nor coupling to the front cavity.

To be able to obtain formant frequencies from the models that match measurements from natural data is quite encouraging, because it suggests that these simplified models are reasonable models for the pharyngeal and uvular consonants.

Ghazeli's X-ray data, upon which the choice of the models' dimensions were partly based, were of the consonants preceding a low vowel. Hence, the models proposed in this section are appropriate for the production of the pharyngeal and the uvular consonants when adjacent to a low vowel (like the vowel /a/). When modeling these consonants adjacent to other vowels, such as /i/ and /u/, one should consider the perturbation effects that occur in anticipating the following vowel; for example rounding, as in the case of /u/, would lower the formant frequencies associated with the front cavity.

FORMANT FREQUENCIES (Hz)						
The Pharyngeal Model						
	A_c=0.15 cm^2		A_c=0.2 cm^2		A_c=0.25 cm^2	
	A_g =0	A_g=0.1 cm^2	A_g=0	A_g=0.1 cm^2	A_g=0	A_g=0.1 cm^2
l_c=1 cm						
F1	604	628	605	629	605	630
F2	1216	1595	1326	1657	1409	1694
F3	1916	2005	1939	2066	1965	2093
F4	3187	3195	3193	3203	3199	3211
l_c=2 cm						
F1	632	664	635	665	637	666
F2	966	1405	1058	1480	1136	1541
F3	2030	2063	2037	2077	2045	2131
F4	3432	3434	3433	3436	3434	3427
The Uvular Model						
	A_c=0.15 cm^2		A_c=0.2 cm^2		A_c=0.25 cm^2	
	A_g =0	A_g=0.1 cm^2	A_g=0	A_g=0.1 cm^2	A_g=0	A_g=0.1 cm^2
l_c=1 cm						
F1	519	574	536	590	550	603
F2	999	1018	1028	1051	1053	1082
F3	2309	2416	2336	2438	2359	2456
F4	2776	2794	2792	2813	2809	2832
l_c=2 cm						
F1	465	523	481	538	497	554
F2	1039	1047	1054	1065	1069	1082
F3	2253	2376	2267	2389	2281	2401
F4	3084	3090	3088	3094	3092	3098

Table 2.1: First four formant frequencies (Hz) of the idealized pharyngeal and uvular models (Figures 2.2a, 2.2b). The formant frequencies were calculated using the program TBFDA1 (see text for details).

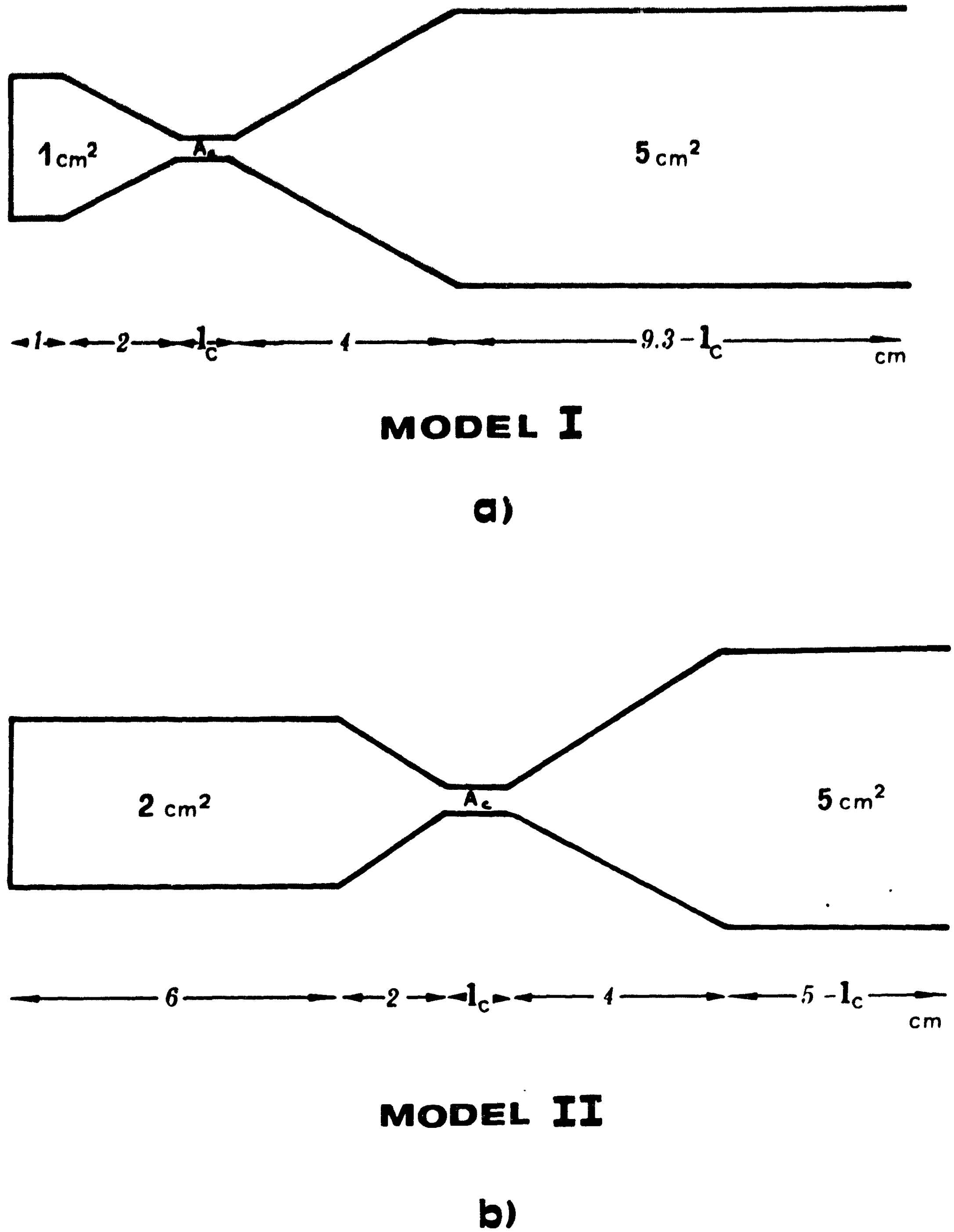


Figure 2.3: Idealized models I and II (pharyngeal and the uvular) with tapering.

FORMANT FREQUENCIES (Hz)						
The Pharyngeal Model						
	A_c=0.15 cm^2		A_c=0.2 cm^2		A_c=0.25 cm^2	
	A_g =0	A_g=0.1 cm^2	A_g=0	A_g=0.1 cm^2	A_g=0	A_g=0.1 cm^2
l_c=1 cm						
F1	715	747	713	745	710	743
F2	1152	1785	1226	1802	1284	1813
F3	2263	2344	2264	2354	2267	2365
F4	3606	3623	3572	3608	3587	3597
l_c=2 cm						
F1	749	801	748	798	745	796
F2	1015	1674	1078	1687	1131	1697
F3	2422	2473	2425	2475	2437	2481
F4	3784	3799	3744	3765	3711	3737
The Uvular Model						
	A_c=0.15 cm^2		A_c=0.2 cm^2		A_c=0.25 cm^2	
	A_g =0	A_g=0.1 cm^2	A_g=0	A_g=0.1 cm^2	A_g=0	A_g=0.1 cm^2
l_c=1 cm						
F1	483	553	499	568	513	582
F2	1232	1241	1238	1249	1243	1255
F3	2610	2737	2602	2725	2594	2713
F4	3415	3425	3395	3405	3379	3309
l_c=2 cm						
F1	447	517	460	529	474	542
F2	1379	1385	1380	1385	1380	1386
F3	2573	2706	2559	2689	2546	2673
F4	3677	3683	3627	3635	3587	3596

Table 2.2: First four formant frequencies (Hz) of the tapered models (Figures 2.3a, 2.3b). The formant frequencies were calculated using the program TBFDA1 (see text for details).

2.3 Noise Source at a Supraglottal Constriction

In the case of noise generation in the vicinity of a supraglottal constriction, we could model the vocal tract by the network model shown in Figure 2.4. As shown in this figure, the noise source is modeled as a pressure source (p_s) in series with the front and the back cavities, and the constriction itself is modeled as an acoustic mass and a kinetic resistance (M_c, R_c).[3] The transfer function of the volume velocity at the lips (U_o) to the pressure source (p_s) at the constriction has the following form:

$$\frac{U_o}{p_s} = K(s)\frac{\prod_{i=1}^{n}(s-s_i)(s-s_i^*)}{\prod_{j=1}^{m}(s-s_j)(s-s_j^*)} \tag{2.3}$$

Where $K(s)$ is the transfer function value at frequencies near the origin, and s is the complex frequency, with a real and an imaginary part $(\sigma+jw)$. As shown in Eq. 2.3, the zeros (s_i) and the poles (s_j) of the transfer function occur in complex conjugate pairs. The zeros are those frequencies for which the impedance looking back from the pressure source (Z_b) is infinite. The poles, on the other hand, are those frequencies for which the sum $(Z_f+Z_b+jwM_c)$ is zero. Hence, we can rewrite Eq. 2.3 as:

$$\frac{U_o}{p_s} = K(s)\frac{poles\ of\ (Z_b)}{zeros\ of\ (Z_b+Z_f+jwM_c)} \tag{2.4}$$

If wM_c is much larger than the characteristic impedances of the front and back cavities, then the poles of Z_b will cancel (or nearly cancel depending on the source location and the value of wM_c) from the numerator and the denominator in Eq. 2.4. The pole-zero cancellation of the back-cavity poles reduces the transfer function in Eq. 2.4 to:

$$\frac{U_o}{p_s} \cong K(s)\frac{1}{poles\ of\ (Z_f)} \tag{2.5}$$

The implication of the close proximity of the poles and zeros (which correspond to the back-cavity resonances) is that these poles (formant frequencies) will not be excited

[3] The acoustic mass of the constriction (M_c) is equal to $\rho l_c/A_c$, and the kinetic resistance (R_c) is equal to $\rho U_c/A_c^2$; l_c and A_c are the constriction length and cross-sectional area, respectively, ρ is the air density, and U_c is the volume velocity at the constriction.

strongly by the pressure source.

As stated earlier, the pole-zero cancellation will occur if wM_c is very large in comparison with the characteristic impedance of the cavities. Hence, for a given pressure source location we would expect that the longer and and the narrower the constriction is, the more likely we will achieve the pole-zero cancellation.

In this section, we will calculate the zeros of U_o/p_s for the untapered models I and II (corresponding to the pharyngeal and the uvular models, respectively). The calculations will be performed assuming the cross-sectional area of the constriction (A_c) to be 0.15 cm^2, and an open glottis with a cross-sectional area (A_g) of 0.1 cm^2. We will examine the effects of the pressure-source location, and the length of the constriction on the zero-frequency locations. The zeros of U_o/p_s were calculated using the program TBFDA1.

2.3.1 Results and Discussion

Figures 2.5 and 2.6 show plots of the zeros of U_o/p_s, as functions of the distance (d) between the pressure source and the constriction, and the length of the constriction (l_c), for the two models (I, II), respectively. In the figures the zeros are superimposed with the formant frequencies of the models calculated in Section 2.2. Only zeros and poles below 3600 Hz are shown.

For the two models, the first zero is at low frequencies (below F1). As the distance (d) between the pressure source and the constriction increases, the second zero for these models approaches the Helmholtz resonance (F2 for the pharyngeal and F1 for the uvular), and the third zero for the uvular approaches F3 (first resonance of the back cavity). As seen in the plots, the distance between the second zero and the Helmholtz resonance is rather sensitive to the pressure-source location; for source locations just beyond the constriction the pole-zero distance could be as high as 300 Hz. On the other hand, the third zero for the uvular is in the near vicinity of F3, suggesting that a pole-zero cancellation is likely to occur regardless of the source location.

The plots also show that for a given pressure-source location, the distance between the poles and the zeros decreases as the constriction becomes longer. This result verifies what we predicted earlier, that is, as M_c becomes larger (in this case by lengthening the constriction (l_c)) less coupling occurs between the front and the back cavities, and hence the distance between the poles and the zeros becomes smaller.

From the pole and zero locations (Figures 2.5, 2.6), we can predict that when a noise source is present at a supraglottal constriction, then: (1) for the pharyngeal model F1, F3, and F4 should be excited by the source. The degree of F1 excitation is expected to be less than that of F3 and F4 since F1 amplitude will be influenced by the presence of the low-frequency zero, (2) for the uvular model F2 and F4 are expected to be excited by the noise source, while F3 should not be excited since a pole-zero cancellation at that frequency is likely to occur, regardless of the source location and the length of the constriction, and (3) depending on the noise-source location, the Helmholtz resonance for both models (F2 for the pharyngeal and F1 for the uvular) may or may not be excited by the noise source, since the pole-zero distance in that frequency range is highly sensitive to the source location.

Two of the factors that influence formant amplitudes are the pole-zero locations and the bandwidths of the formant frequencies. We have explored in this section the first issue by calculating the pole and zero frequency locations for the pharyngeal and uvular models. In the following section, we will address the second issue by determining the bandwidths of the formant frequencies for both models.

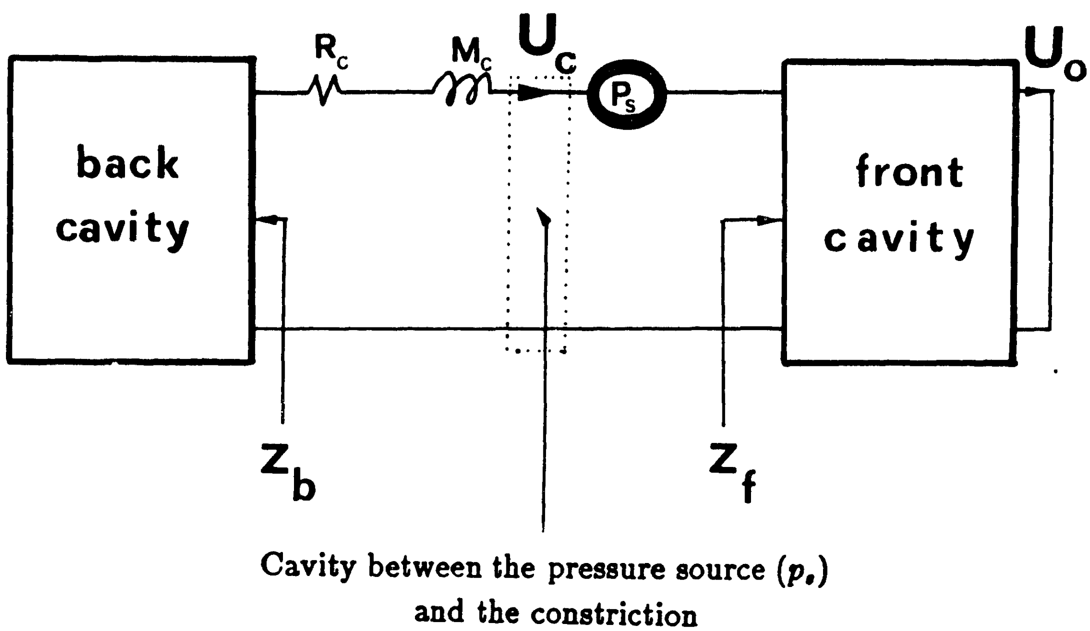


Figure 2.4: Network model of a constricted vocal tract in the presence of a noise source (p_s) at the vicinity of the constriction. M_c, R_c are the acoustic mass and resistance of the constriction, U_o, U_c are the volume velocity at the lips and the constriction, and Z_f, Z_b are the impedances seen by the pressure source of the front and the back cavities, respectively.

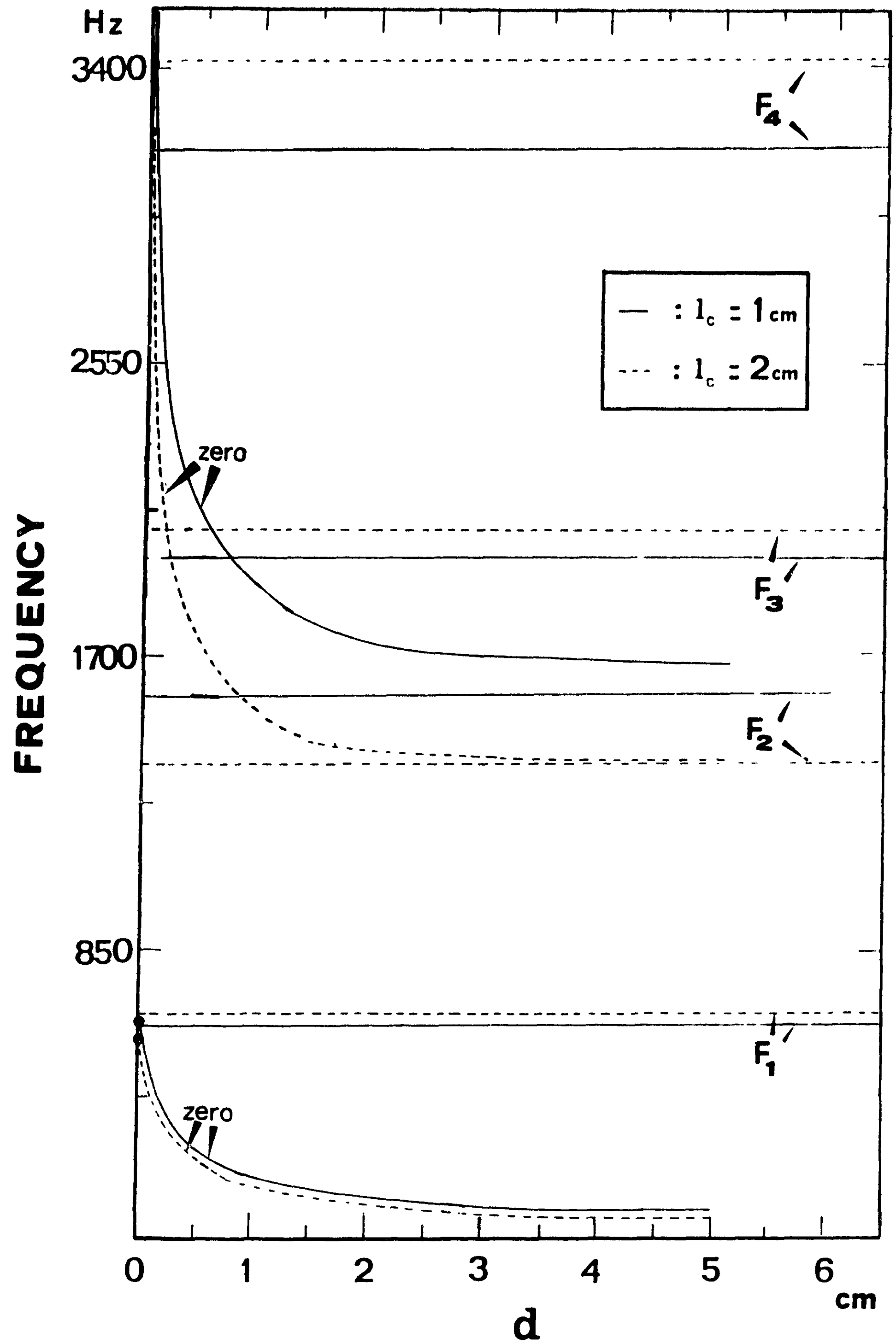


Figure 2.5: Plots of the first three zeros of U_o/p_s superimposed on the first four formant frequencies as a function of the distance between the pressure source and the constriction *d*. Model I (pharyngeal).

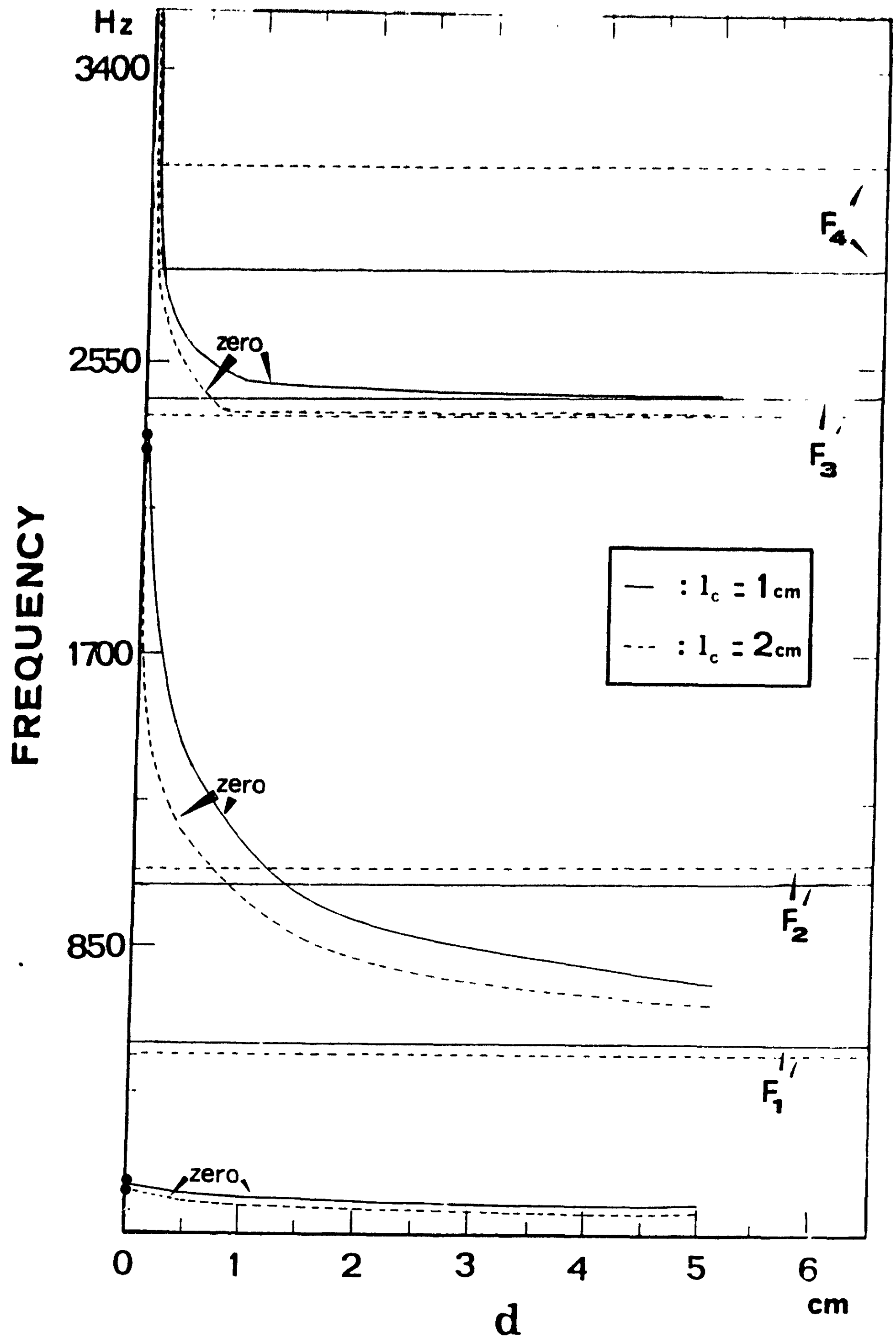


Figure 2.6: Plots of the first three zeros of U_o/p_s superimposed on the first four formant frequencies as a function of the distance between the pressure source and the constriction *d*. Model II (uvular).

2.4 Losses in the Vocal Tract

Losses in the vocal tract arise from the resistive components of the impedances of the walls, radiation, glottis, and constriction, and from viscosity and heat conduction. In this section, we will calculate the contributions of the different losses to the bandwidths of the first four formant frequencies for the pharyngeal and uvular models. The program TBFDA1 calculates the bandwidth contributions of all losses except those due to the constriction. Hence, we will use this program to calculate the bandwidth contributions of all losses excluding the constriction, and we will derive expressions for calculating the constriction-losses contribution and perform these calculations separately. The purpose of these calculations is to gain an insight into the degree of damping of the pharyngeal and the uvular formant frequencies, and to determine which of the losses contribute the most to the formant bandwidths for both the closed- and open-glottis cases.

2.4.1 Method

At low frequencies we can model the vocal tract by the circuit shown in Figure 2.7a. In the circuit the volume velocity at the glottis (U_g) is modeled as a current source, the air behind the constriction by its compliance (C_a), the wall impedance as an acoustic mass and resistance (M_w, R_w), and the constriction by its acoustic mass and kinetic resistance (M_c, R_c).

The circuit elements are represented by the following expressions:

$$C_a = \frac{V_b}{\rho c^2}$$

$$R_w = \frac{R_{wmech}}{A_s}, \; M_w = \frac{M_{wmech}}{A_s}$$

$$R_c = \frac{\rho U_c}{A_c^2}, \; M_c = \frac{\rho l_c}{A_c}$$

Where

ρ= air density

c= velocity of sound in air

V_b= volume of air in the back cavity

R_{wmech}= mechanical resistance of the walls per unit area

M_{wmech}= mechanical compliance of the walls per unit area

A_s= surface area of the vocal tract behind the constriction

U_c= volume velocity of air at the constriction

A_c, l_c= cross-sectional area and length of the constriction

The resonance of this circuit (f_h), which is the Helmholtz resonance of the idealized models I and II, could be approximated by:

$$f_h \cong \frac{1}{2\pi\sqrt{C_a M_{tot}}} \tag{2.6}$$

Where M_{tot} is the parallel combination of M_c and M_w ($M_c || M_w$).

Assuming that the quality factor (Q=f/bandwidth) of the circuit is high, then we can calculate the contribution of the constriction resistance (R_c) to the bandwidth of the resonant frequency (f_h) as follows:

$$B_c = \frac{M_w R_c}{2\pi M_c (M_w + M_c)} \tag{2.7}$$

At high frequencies, we can think of each cavity of the vocal tract as a distributed transmission line of length l and cross-sectional area A (Figure 2.7b). Shadle (1985) showed that the bandwidth contribution of the constriction to the resonances of such a transmission line is:

$$B_c = \frac{\rho c^2 R_c}{\pi A l (R_c^2 + (2\pi f M_c)^2)} \tag{2.8}$$

The bandwidth contributions of the constriction losses to the first four formant frequencies for models I and II were calculated using Equation 2.7 and 2.8, with A_c=0.15 cm^2, and two values of l_c: 1 and 2 cm. The values for R_{wmech} and M_{wmech} were chosen to be similar to those used in Section 2.2 (1060 $gcm^{-2}sec^{-1}$ and 1.5 gcm^{-2}). The calculations were performed assuming the volume velocity U_g to be 150 cm^3/sec for the closed-glottis case, and 300 cm^3/sec for the open-glottis case. The bandwidth contributions of the walls, viscosity, heat conduction, radiation, and the glottis (for the open-glottis case) were calculated using the program TBFDA1. The results are shown in Table 2.3.

2.4.2 Results and Discussion

If we compare the bandwidth values in Table 2.3 for the closed-glottis case (A_g=0) to the formant frequency locations calculated in Section 2.2 (Table 2.1) we notice that the formant frequencies of both models are underdamped ($Q > 0.5$). However, the bandwidths of the Helmholtz resonances (F2 for the pharyngeal and F1 for the uvular) and F4 for both models have wider bandwidths than for the other two formant frequencies. The main contributor to the wide bandwidth of the Helmholtz resonance is the constriction loss (B_c), whereas radiation losses presumably account for the widened F4 bandwidth.[4] The constriction contribution (B_c) for formant frequencies other than the Helmholtz is less significant, and decreases as the formant frequency locations become higher.

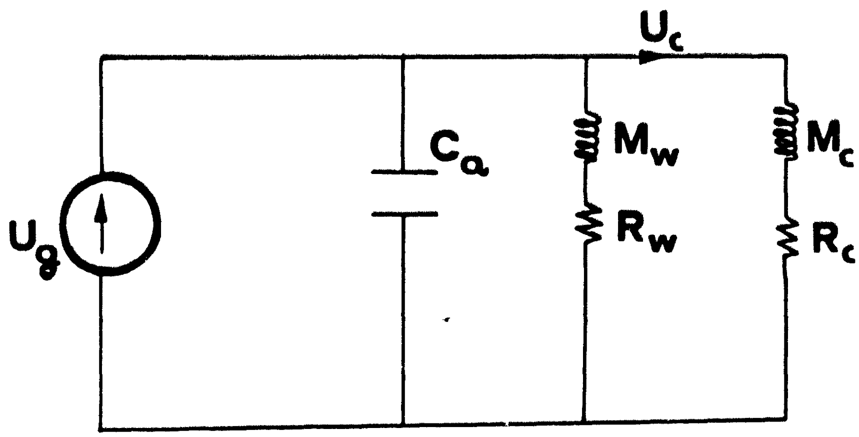
The results also show that the formant bandwidths for the open-glottis are higher than those for the closed-glottis. The increase in (B_c) occurs mainly because of the assumed higher flow rate for the open-glottis case, while the increase in the bandwidth contribution of the other losses (B_{wvhrg}) is due to the glottal losses which affect the Helmholtz and back-cavity resonances to a great extent.

We saw in Section 2.3 that the pole-zero distance for the Helmholtz resonance could be as high as 300 Hz depending on the source location. The results in this section indicate that the Helmholtz resonance is heavily damped by the glottal losses. So, even if it is not cancelled completely by the zero of the transfer function U_o/p_s, it is not expected to be strongly excited, since its bandwidth, for both models, is high (greater than 500 Hz).

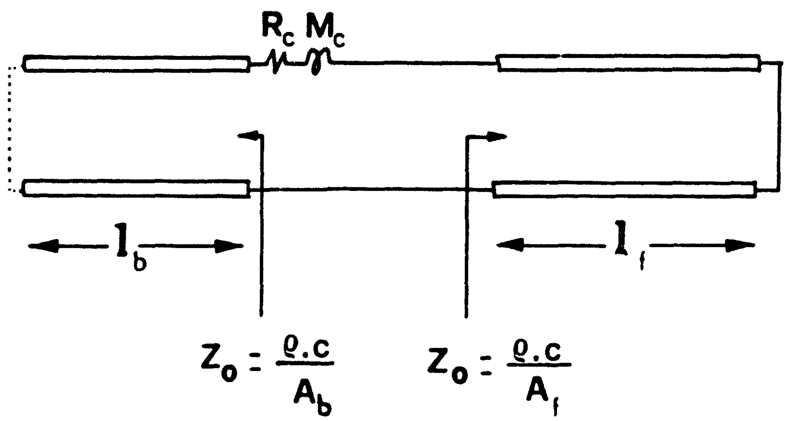
For both cases (A_g=0 and A_g= 0.1 cm^2), the constriction contribution B_c decreases

[4]Radiation losses vary with the square of the frequency for front-cavity resonances (Fant, 1960).

as l_c, or equivalently M_c (the acoustic mass of the constriction), increases. This can be seen in Equations 2.7, 2.8 where B_c varies inversely with M_c.



a)



b)

Figure 2.7: Two models of the vocal tract used for calculating the bandwidth contribution of the constriction resistance (R_c) to the formant-frequency bandwidths : a) Low-frequency model, and b) High-frequency model.

FORMANT-FREQUENCY BANDWIDTHS(Hz)				
The Pharyngeal Model				
	A_g=0		A_g=0.1 cm^2	
	B_c	B_{wvhr}	B_c	B_{wvhrg}
l_c=1 cm				
F1	60	30	95	156
F2	152	74	305	642
F3	7	104	11	282
F4	2	165	5	176
l_c=2 cm				
F1	16	36	27	138
F2	73	72	146	999
F3	2	117	3	172
F4	.83	192	1	191
The Uvular Model				
	A_g=0		A_g=0.1 cm^2	
	B_c	B_{wvhr}	B_c	B_{wvhrg}
l_c=1 cm				
F1	137	63	275	312
F2	35	54	63	99
F3	17	49	30	162
F4	5	213	9	215
l_c=2 cm				
F1	60	80	121	317
F2	9	63	18	69
F3	6	46	8	173
F4	1	253	2	262

Table 2.3: Bandwidth contributions of the constriction (B_c), and the contributions of the other losses (B_{wvhr}) to the bandwidths the first four formant frequencies (Hz) for the pharyngeal and uvular models. For the open-glottis case the glottal losses are also included with the "other" losses (B_{wvhrg}) (see text for details).

2.5 Summary and Discussion

In this chapter, we have proposed idealized models of the vocal tract during the production of pharyngeal and uvular consonants (Section 2.1). From these models we have made certain predictions regarding the formant frequencies and their cavity affiliation (Section 2.2), the bandwidths of these formant frequencies (Section 2.4), and the effects of creating a noise source in the vicinity of a supraglottal constriction on the transfer function of the volume velocity at the lips to the pressure at the noise source (Section 2.3).

We have shown that for both the closed- and the open-glottis cases F1 and F4 for the pharyngeal model are higher than those for the uvular and F3 is lower. F2 was undistinguishable between the two consonants when the glottis is closed. However, if the glottis is open, F2 for the pharyngeal is higher than that for the uvular.

For the closed-glottis case, we showed that the calculated bandwidths of the Helmholtz resonance (F2 for the pharyngeal and F1 for the uvular) and of F4 for both models is higher than those of the other two formant frequencies. We attributed the widened bandwidth in the Helmholtz resonance case to the bandwidth contribution of the constriction resistance, whereas radiation losses accounted for the widened F4 bandwidth.

When the glottis is open, we showed that the presence of a noise source in the vicinity of the constriction resulted in introducing zeros in the transfer function U_o/p_s, which are at low frequencies, in the vicinity of the Helmholtz resonances, and at the frequency of F3 for the uvular model, resulting in F3 cancellation. The glottal losses were shown to have a large effect on damping the Helmholtz resonances and the back-cavity resonances. Based on these results, we could predict for the voiceless pharyngeal that only F3 and F4 should be strongly excited, since F1 amplitude will be influenced by the low-frequency zero and because its bandwidth is widened by the different losses, and F2 amplitude will be reduced by the glottal losses. For the voiceless uvular, we can predict that only F2 and F4 will be strongly excited by the pressure source, since F1 is heavily damped and F3 is cancelled by a zero.

In the following chapter, we will examine the acoustic properties of the pharyngeal and uvular consonants in naturally-spoken utterances and compare our findings with the theoretical predictions arrived at in this chapter.

Chapter 3

Acoustic Analysis: Methods and Results

In this chapter, the methods and results of the acoustic analysis of pharyngeal and uvular consonants are presented. The first two sections include brief descriptions of the speakers' backgrounds, the corpus, and the recording method used in this study. In the following sections, data analysis procedures are described and the results are interpreted in terms of the articulatory mechanisms involved in the production of these consonants. We will compare the results of the acoustic analysis to the theoretical predictions introduced in Chapter 2.

3.1 Speakers

Four adult males (HA, JM, LT, MU), whose native language was Arabic, participated as speakers. Three speakers (HA, LT, MU) were from the city of Baghdad, Iraq. The fourth speaker (JM) was from Southern Lebanon.[1] None of the speakers had known speech or hearing impairments.

[1] At the early stages of this study, four speakers from four different countries were chosen. In that case, it was difficult to attribute the variability observed in the acoustic properties of the speech sounds to individual or/and to dialectal differences. Consequently, we decided to choose three of the four speakers to be from the same country (Iraq). The speech data of the Lebanese speaker (JM) was considered, as it showed similar acoustic properties to those obtained from the speech data of the Iraqi speakers.

3.2 Corpus and Recording Method

The corpus chosen for this study included consonant-vowel (CV) and glottal stop-vowel-consonant-vowel(ʔVCV) isolated, nonsense syllables. The consonant was one of the two pharyngeal consonants, /ʕ,ħ/, or one of the three uvular consonants, /ʁ,χ,q/. The vowel, one of the three long vowels of Arabic, /aa,ii,uu/. This selection of utterances resulted in a total number of sixty utterances (considering all possible vowel combinations). In Arabic, no word starts with a vowel. This is why a glottal stop was inserted at the beginning of the VCV utterances.

The subjects were recorded in a sound-treated room (signal to noise ratio was approximately 30 dB) using an Altec microphone, a Shure microphone mixer and a Nakamichi Lx-5 tape recorder. The microphone was placed 20 cm away from the speaker's mouth.

The subjects were instructed to read the list of utterances twice, at a moderate rate, keeping the pitch constant, and inserting pauses between utterances. The list consisted of CV and ʔVCV "sequences", where a CV sequence is defined to be a set of three CV utterances of one consonant with the three vowels. Similarly, a ʔVCV sequence consists of three ʔVCV utterances each with a different second vowel. At the beginning and at the end of each sequence, an extra CV utterance was inserted. These extra utterances were not considered as part of the data analysis corpus but were included for possible intonation rise/fall at the beginning/end of a sequence. The utterances were then low-pass filtered at 4.8 kHz, digitized at 10 kHz and recorded into the SPEECHVAX and the LISP machines. Both are facilities of the MIT Speech Group.

3.3 Time-Domain Analysis: Duration

Durational measurements of the consonantal interval were made from the natural ʔVCV utterances in the corpus. These utterances establish clear left and right boundaries for each consonant from which its duration could be measured accurately, as will be described below. Using SPIRE,[2] displays of spectrograms, temporal waveforms,

[2]SPIRE is a software package on the Speech Group Lisp Machines used for the acoustic analysis of

plots of the calculated total energy and of the energy in the low frequency region (125 to 750 Hz) of the speech signals were obtained.[3] From these displays, the consonantal boundaries were determined and labeled manually from several cues. In the case of the voiceless consonants /ħ,χ,q/, rapid spectral changes were indicative of the consonants' left boundaries. These changes were illustrated in the onset of frication noise for the fricatives /ħ,χ/, and in the beginning of the silence gap for the stop /q/. The right boundaries were considered to be the beginning of voicing in the following vowel. Consequently, aspiration, if present, was considered to be part of the consonantal interval. Examples of these utterances are shown in Figure 3.1. Voice onset time (VOT) for the voiceless stop /q/ was defined as the interval between the release of the consonant and the onset of voicing in the following vowel.

Spectrographic displays of /ʕ/ showed two allophones of this consonant. The first is a stop-like allophone (Figure 3.2a), and the second is a continuant (Figure 3.2b). The boundaries of the stop-like allophone were determined the same way as in the case of the stop consonant /q/. The spectrograms of the continuant allophone showed a lower fundamental frequency (f0) for the consonant (vertical striations farther apart) than that for the following vowel. For some speakers, the plots of the low frequency energy showed lower values for the consonant relative to the vowel, as well as a lower f0. These two cues were used (simultaneously or alternatively, depending on the speaker), for boundary detection for the continuant allophone of /ʕ/.

The voiced consonant /ʁ/ appeared spectrographically to have clear first formant structure and its waveform envelope was lower in amplitude than that of the surrounding vowels. However, formants above F1 were very weak. This discontinuity in higher formant structure along with the change in the waveform envelope were cues for the consonant boundaries (Figure 3.2c).

---

speech waveforms.

[3]Energy was computed using the following equation:

$$E(t) = 10 \times log \int W(f) * |S(f)|^2 df \qquad (3.1)$$

Where $E(t)$ is the energy, $W(f)$ and $S(f)$ are the spectrum at time t of the weighting window and the signal, respectively.

Absolute values of consonant duration have no significant meaning, particularly when the contextual environment is restricted as it is in the selected utterances. However, a durational contrast between the voiced and the voiceless consonants would be of interest as would durational differences between the uvular and the pharyngeal consonants.

3.3.1 Results and Discussion

Table 3.1 summarizes the results of the durational measurements for all the consonants considered with the three vowels. In this table, the average duration values and the standard deviation (calculated across all speakers) are shown. Two measurements were made for the voiceless uvular stop /q/: the VOT and the total duration of the consonant. From these results, the following were concluded:

- The voiceless consonants /ħ,χ,q/ are longer in duration than the voiced consonants /ʕ,ʁ/. However, these differences in duration could not be attributed to voicing alone, since the pairs (ħ,ʕ) and ((χ,q),(ʁ)) are not minimal pairs, as the manner of articulation was observed to be different as well.

- The two classes of sounds (uvular and pharyngeal) cannot be distinguished from one another on a durational basis. In fact, the voiced and voiceless consonants in both classes have similar average durations (117 msec. and 113 msec. for /ʕ/ and /ʁ/, respectively, and 159 msec., 169 msec. and 158 msec. for /ħ/, /χ/ and /q/, respectively.)

- The identity of the following vowel did not significantly affect the duration of the consonant.

- The VOT of the voiceless uvular stop /q/ was measured between the release of the consonant (typically a weak burst) and the onset of the following vowel. In some cases (in 1% of the total number of utterances), the stop was unreleased (no burst following the stop closure, Figure 3.3a). The VOT of the consonant consisted of either a silence gap (Figure 3.3b) or of aspiration (Figure 3.3c), in 74% and 25%

of the utterances for the two cases, respectively. This indicates that /q/ has an aspirated realization, which is in disagreement with Al-Ani's classification of /q/ as an unaspirated voiceless stop (Al-Ani, 1970).

Comparing these results with the results of other studies on the duration of consonants in English and in other languages (Lisker and Abramson, 1964; Umeda, 1977), one finds that they agree in terms of the general tendency for the voiced consonants to be shorter than their voiceless counterparts within a class of sounds.[4]

It is worthwhile mentioning that in the process of labeling the different phonemes manually, three observations regarding the production of these consonants were made. First, there was no evidence of voicing in the time waveform or in the spectrograms for the intervocalic voiceless consonants. This suggests that the voicing distinction remains unchanged in intervocalic position. Second, in a number of utterances (5% of the total number) the vowel preceding the voiceless uvular stop /q/ was followed by a brief period of weak noise (Figure 3.4a) preceding the stop closure. This noise could be attributed to partial devoicing of the vowel or to preaspiration of /q/. Third, irregular bursts were observed during the production of the voiced pharyngeal consonant /ʕ/ in 90% of the utterances, regardless of its realization (Figure 3.4b).

[4]It should be noted that the context in these studies was considerably different than that considered here. However, the general conclusion regarding the "relative duration" of the voiced and voiceless consonants could still be drawn from these studies.

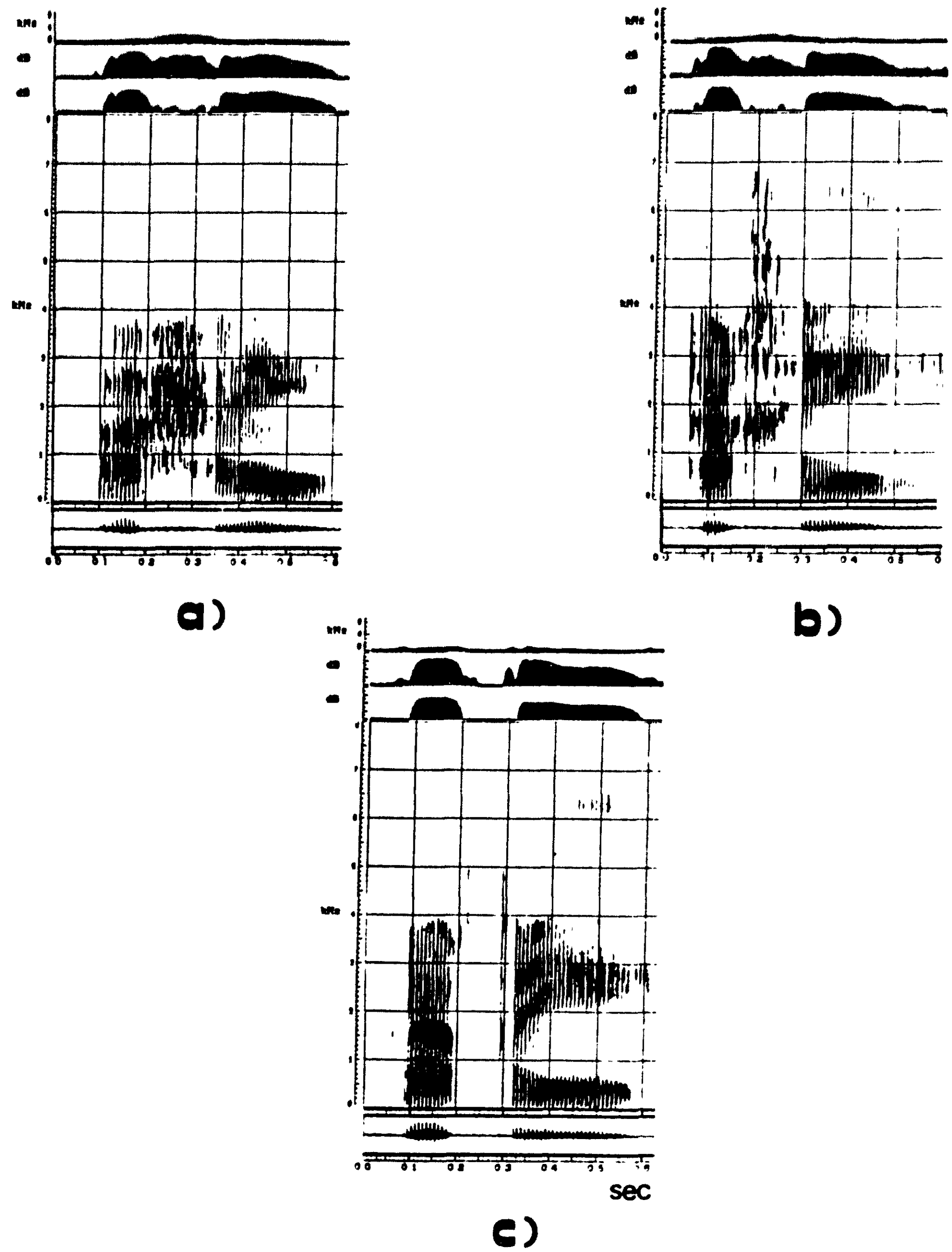


Figure 3.1: Examples of displays used in determining the consonantal boundaries in ʔVCV utterances. The top half of each display shows plots of the zero-crossing rate, total energy and the low-frequency energy (125 to 750 Hz) in the signals. The spectrograms and time waveforms of these utterances are shown at the center and bottom of each display. a) /ʔaaħii/ b) /ʔaaχii/ c) /ʔaaqii/. All utterances were spoken by speaker LT.

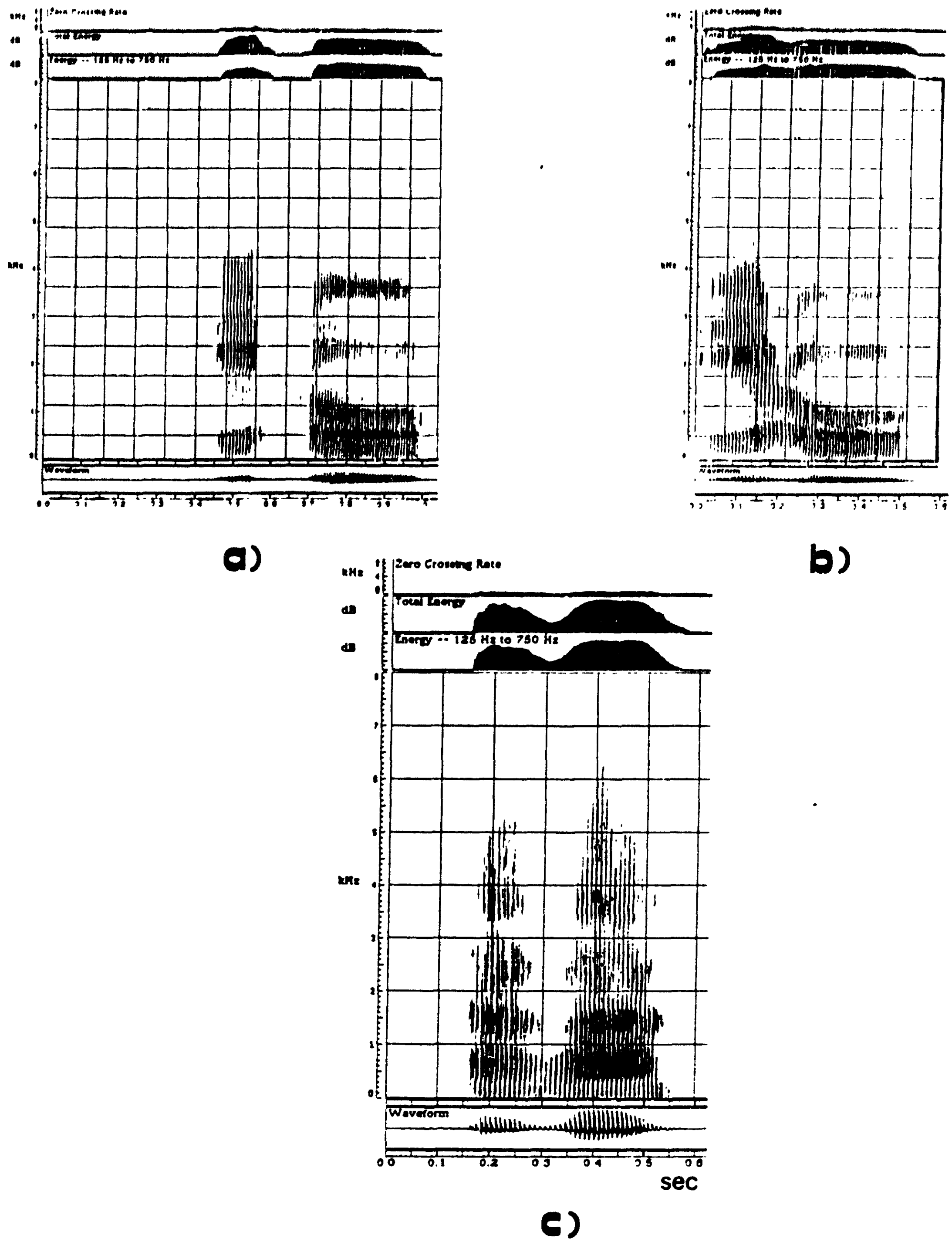


Figure 3.2: Examples of displays used in determining the consonantal boundaries in ʔVCV utterances. For a description of these displays see Figure 3.1. a) /ʔiiʕuu/ -stop allophone-, b) /ʔiiʕuu/ -continuant allophone-, c) /ʔaaʁaa/. All utterances were spoken by speaker MU.

Consonant Duration (msec.)				
C	ʔVCaa	ʔVCii	ʔVCuu	ʔVCV
ʕ	120 (18)	115 (20)	116 (25)	117 (21)
ħ	162 (15)	159 (24)	157 (18)	159 (17)
ʁ	110 (18)	115 (27)	116 (16)	113 (19)
χ	170 (15)	167 (18)	171 (22)	169 (16)
q	160 (27)	158 (24)	156 (30)	158 (25)

VOT Duration (msec.) for /q/				
VOT	43 (5)	37 (7)	40 (7)	40 (5)

Table 3.1: Results of durational measurements of the five consonants. Average values in msec. pooled across four speakers are shown with standard deviations (in parentheses). In the first three columns, the duration measurements were made with the consonant preceding the vowels /aa/, /ii/, and /uu/, respectively. In the last column the values shown reflect measurements from all ʔVCV utterances.

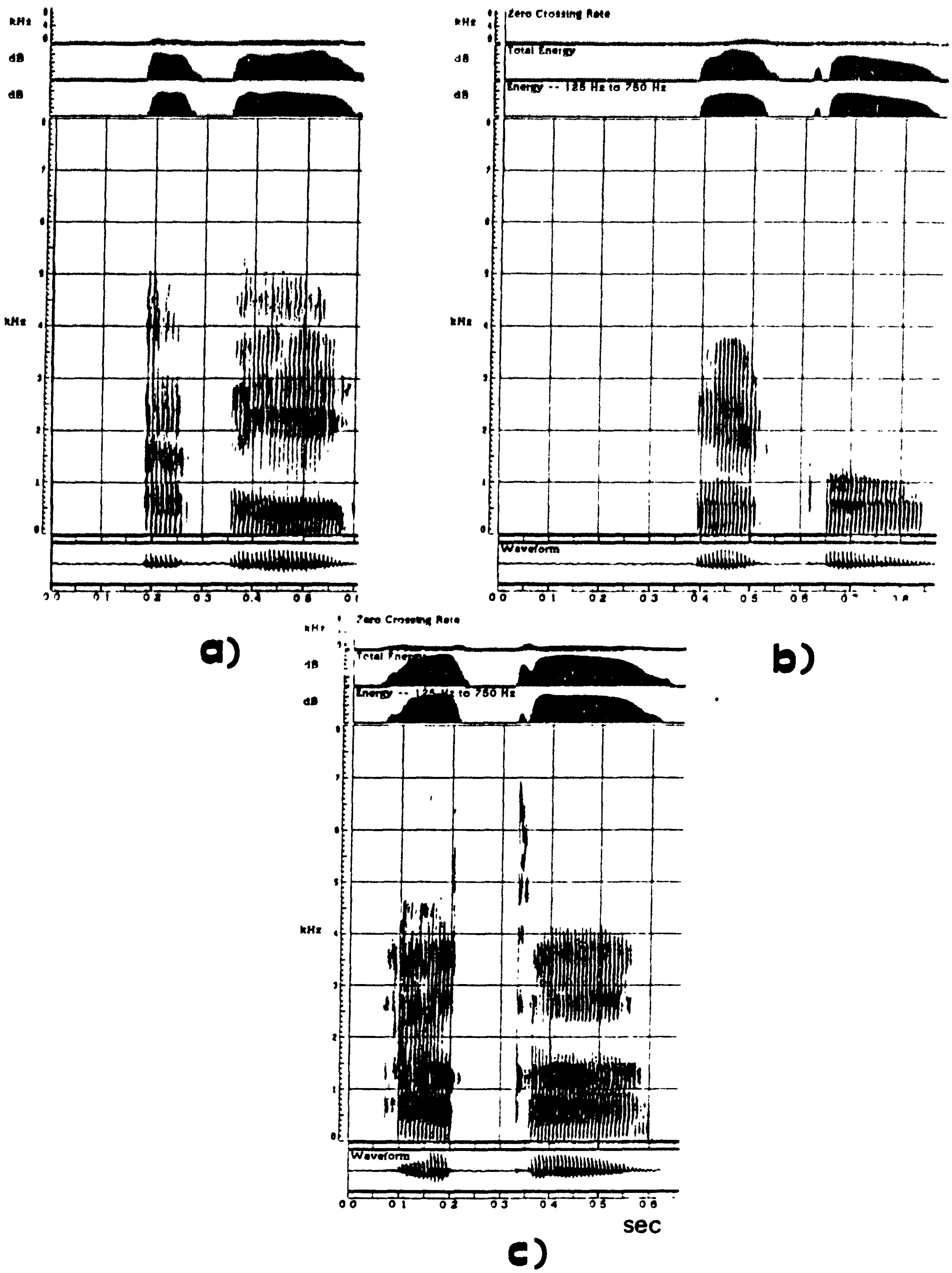


Figure 3.3: Three different realizations of the voiceless uvular stop /q/: a) unreleased, b) released unaspirated, c) released aspirated.

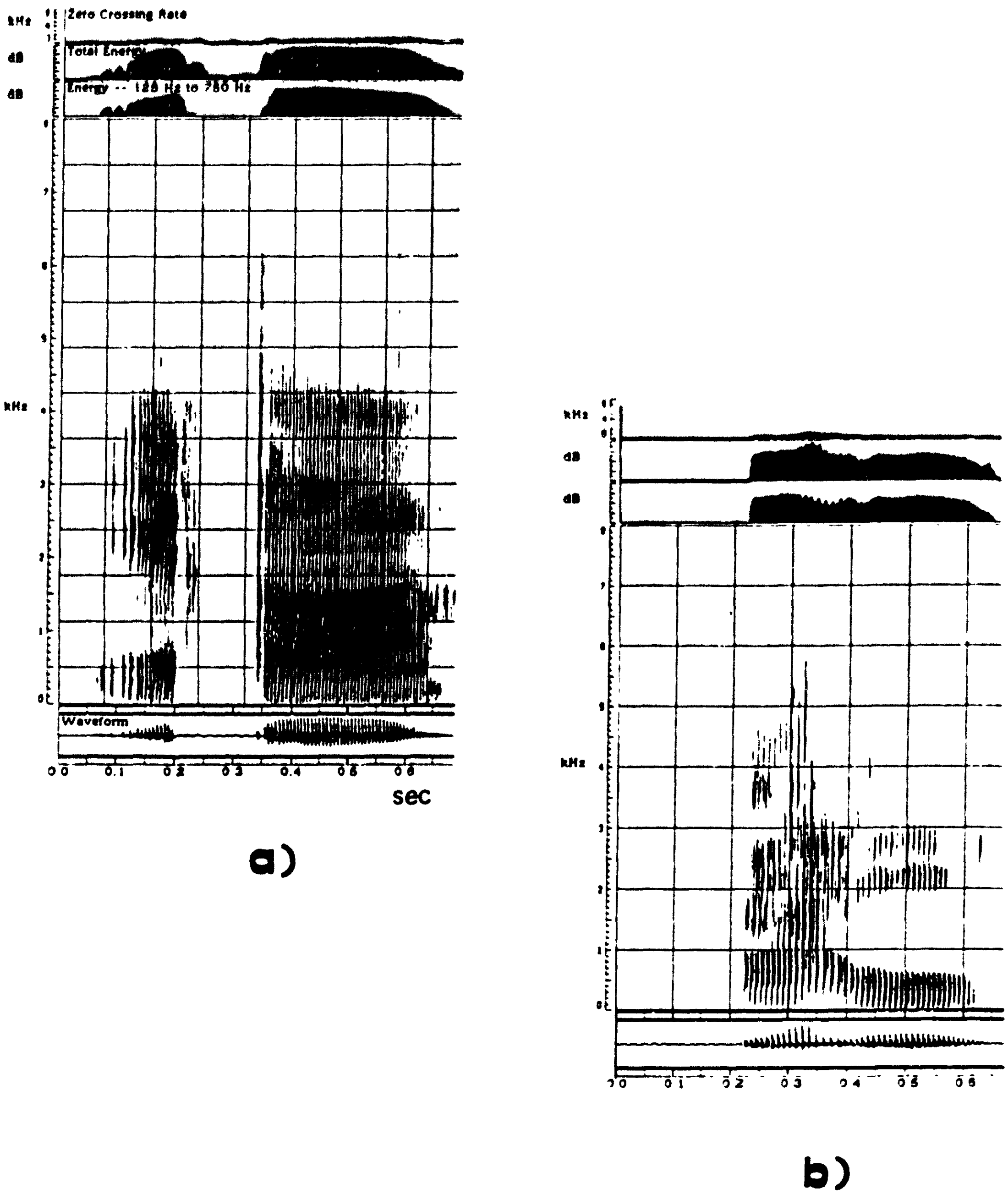


Figure 3.4: a) Weak noise preceding the stop closure for /q/, indicating preaspiration or devoicing the preceding vowel. The utterance shown is /ʔiiqaa/ by speaker HA. b) Irregular bursts during the consonantal interval of /ʕ/. The utterance shown is /ʔaaʕii/ by speaker MU.

3.4 Frequency-Domain Analysis

3.4.1 Method

Frequency-domain analysis was performed using KLSPEC, a software package developed by D.H.Klatt, availabl. . the MIT SPEECHVAX. Four kinds of spectral representations were used to describe the spectral properties of the consonants and the vowels. These four representations are: linear-prediction (LPC), critical-band (CB), discrete Fourier transform (DFT) magnitude, and spectrogram-like spectra (S).

The waveforms, digitized at 10 kHz, analyzed were initially first-differenced, then multiplied by a Hamming window of an appropriate duration. The duration of the Hamming window typically used was 256 samples for computing all spectra except the DFT magnitude spectra. The duration in the latter case was somewhat longer (299 samples) to be able to track fundamental frequencies.

A Kaiser weighting window of duration 256 samples was used for the computation of the 14th order LPC spectra. The critical-band spectra and the spectrogram-like spectra were computed by forming a weighted sum of adjacent DFT energies for each of the 36 CB filters, and 128 S filters used in computing these spectra, respectively. The critical-band spectra employ a Mel frequency scale and filter bandwidths that increase with increasing frequencies. The spectrogram-like filters have a frequency-domain shape that is approximately Gaussian. All the parameters mentioned earlier (e.g., window duration, choice of preemphasis, etc.) are the default values used by the analysis programs and could be adjusted by the user.[5]

3.4.2 Fundamental Frequency Measurements of /ʕ/ and /ʁ/

The voiced consonants /ʕ/ and /ʁ/ exhibit lower fundamental frequencies than the surrounding vowels. This observation is not new. Gardiner (1925) and Jackobson (1957) describe /ʕ/ as being of lower pitch, perceptually, in comparison with surround-

[5]For further detail on the analysis techniques used in KLSPEC, the reader is referred to Klatt, D.H. (1983).

ing vowels.[6] Spectrographic analysis by Ghaseli (1977), showed that "glottal pulses" during the consonantal interval of /ʕ/ are farther apart than they are in vocalic portions.

This has led to f0 measurements of /ʕ/ and /ʁ/ and the following vowels in ʔVCV context, in an attempt to describe quantitatively differences in f0 between these consonants and the three vowels.

Method

The ʔVCV utterances were chosen for this part of the analysis, partly because of greater accuracy of locating consonantal boundaries. The CV utterances were not considered as the lower fundamental frequency of the consonant could be attributed, partially, to its initial position.

Fundamental frequencies were measured using SPECTO, a program available in the KLSPEC software package (see Section 3.4.1). The fundamental is computed by collecting frequencies of local maxima in the DFT spectrum. Only peaks below 3000 Hz contribute to this pool, and the fundamental is the component that accounts for the most peaks as harmonics. The duration of the analysis window typically used (Hamming, in this case) is 299 samples. As the fundamental of the consonants /ʕ, ʁ/ could be as low as 100 Hz, the window duration was adjusted accordingly. It was increased to 350 samples, which corresponds to 35 msec. at a sampling rate of 10,000 samples/sec. The effective duration of such a window is about half this value.

Two measurements were made for each utterance. One, at the midpoint of the consonant, and the other 50 msec. after the consonant/vowel boundary (Figures 3.5 and 3.6). The difference between these two measurements was called Δf0.

[6] Gardiner (1925) described the change in "voice pitch" of /ʕ/ by the following: "In passing to /ʕ/ from a preceding vowel the voice has to descend rapidly, often through more than an octave, and is cut off at its lowest pitch. If a vowel follows, the pitch begins at its lowest and rises quickly, through a similar interval, to a normal vowel pitch" (p. 28).

Results and Discussion

Seventy-two tokens were available for each consonant. Forty-six, and sixty tokens were measurable for the pharyngeal /ʕ/ and the uvular /ʁ/, respectively. Partial devoicing of the consonants accounted for all the unmeasurable tokens of /ʁ/ and six of twenty six unmeasurable tokens of /ʕ/. Stop-like allophones and irregularity of the waveforms accounted for the rest of the unmeasurable tokens of /ʕ/.

The calculated Δf0, which reflects the difference in f0 between the consonant and the following vowel, was averaged across all speakers. Table 3.2 shows the average Δf0 along with the standard deviations.

What could be concluded from these measurements? The vibration of the vocal folds appears to be slower during the production of these consonants than for vowels. One explanation for the lower f0 is that some source-filter interaction occurs during the production of these consonants.

In the case of /ʕ/, the lower f0 is evidence that the constriction in the lower part of the pharynx affects the entire region, including the glottis. This occurs, presumably, because of adducting the vocal folds. Laryngealization, manifested by the irregularities of the waveforms, is yet additional evidence for this effect. For /ʁ/, where the lowering of f0 is not as large as it is for /ʕ/, the slower vibrational pattern could be attributed to the acoustic and/or aerodynamic effects of creating a narrow constriction in the uvular region. A similar phenomenon was observed by Bickley and Stevens (in press) for some English consonants. In their study, acoustic analysis of liquids, glides, /ʋ/ and /ð/ showed that these sounds exhibit a lower f0 than the following vowels in similar context, that is, in a VCV context. The vowel considered in their study was /ɪ/. Their results show that Δf0, averaged across 4 male and 2 female speakers, was 4 Hz for the liquids and glides, 7 Hz for /ʋ/, and 17 Hz for /ð/. There, the difference in f0 was attributed to acoustic or aerodynamic effects of creating a supraglottal constriction on the glottal waveform and the vocal-fold vibrational pattern.

In summary, the results of this section show that the voiced consonants /ʕ/ and /ʁ/ exhibit lower fundamental frequencies than do vowels. The change in the vibrational modes of the vocal folds was attributed to the constricted pharynx in the former case, and to the acoustic and/or aerodynamic effects of creating a narrow constriction in the uvular region of the vocal tract in the latter.

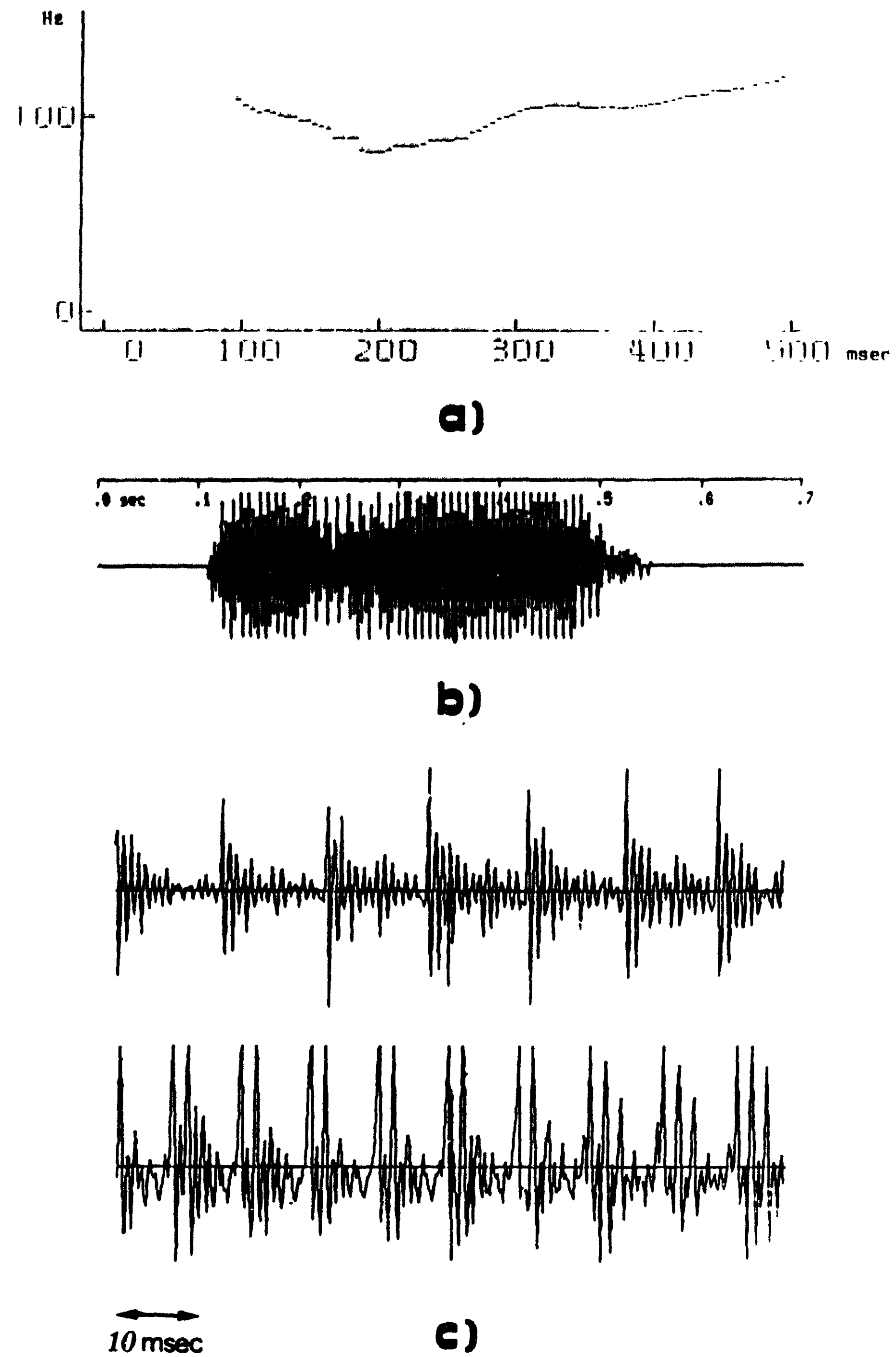


Figure 3.5: a) Fundamental frequency contour for the utterance /ʔaaʕaa/ by speaker MU. b) Time waveform of the utterance. c) Two 70 msec. sections of the waveform centered at the measurement points 250 msec., and 350 msec., respectively (see text for details).

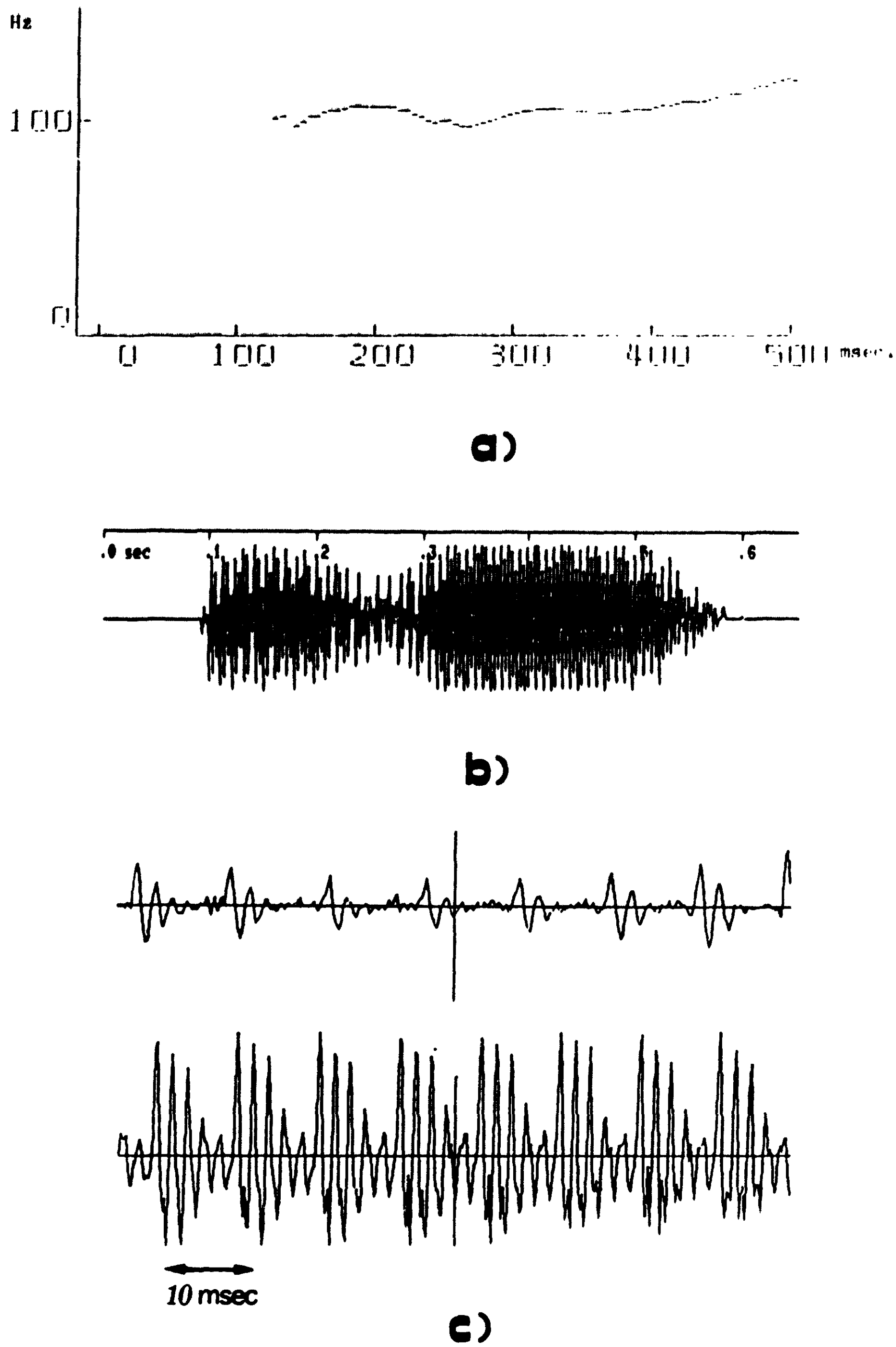


Figure 3.6: a) Fundamental frequency contour for the utterance /ʔaaʁaa/ by speaker MU. b) Time waveform of the utterance. c) Two 70 msec. sections of the waveform centered at the measurement points 260 msec., and 360 msec., respectively (see text for details).

Δf0 (Hz)				
C	ʔVCaa	ʔVCii	ʔVCuu	ʔVCV
ʕ	23 (9)	25 (10)	23 (8)	23 (8)
ʜ	12 (7)	14 (3)	16 (2)	14 (4)

Measurable tokens were 46 out of 72 for /ʕ/, and 60 out of 72 for /ʜ/

Table 3.2: Results of Δf0 measurements (defined as the difference in f0, in Hz, between the consonant and the following vowel). Average values, pooled across four speakers, are shown with standard deviations (in parentheses).

3.4.3 Spectral Analysis

The theoretical predictions, introduced in Chapter 2, regarding formant frequencies for pharyngeal and uvular consonants and the cavity affiliations of the formants can be summarized as follows: (1) the first formant for pharyngeal consonants should be higher than it is for uvular consonants, and F3 should be lower, and (2) the front-cavity resonances (which are strongly excited when there is a noise source near the constriction) should be F1 and F3 for the pharyngeals, and F2 and F4 for the uvulars; the degree of F1 excitation for the pharyngeal was predicted to be less than that of F3. These results were derived using an idealized model of the vocal tract during the production of these consonants, with a constriction located further back for pharyngeal consonants than for uvulars. In this section, we will investigate the spectral properties and formant patterns of the consonants and the vowels in natural CV utterances. The analysis attempts to examine the validity of the theoretical predictions when compared to natural speech, to quantify spectral properties attributed to each class of sounds, and to examine cross-speaker variability.

Method

Natural CV utterances were segmented and labeled manually. The criterion used for determining a consonant-vowel boundary was the same as that described in Section 3.3 and used in labeling the right boundaries of the consonantal intervals in ʔVCV utterances. Formant analysis of the voiced sounds /ʕ/ and /ʁ/, and the vowels was performed using an LPC-based technique with a 25.6 msec. analysis window (see Section 3.4.1 for a description of this analysis technique).

Discrete Fourier transform and critical-band spectra were used as a basis for describing the spectral properties of the voiceless fricatives /ħ/ and /χ/. Due to the random nature of the source for these consonants, a long analysis window (50 msec.) was used to capture the statistical characteristics of the spectra in that interval. The analysis window was centered carefully such that the formant structure in the consonant was steady, excluding the portion of the consonant where the formants undergo a transition due to the following vowel. A typical time interval in which the spectra were computed was 70 msec. prior to the onset of the following vowel (Figure 3.7). In the case of the voiceless uvular stop, analysis of the burst was achieved by placing a short analysis window of duration 12.8 msec. at the burst. The reason for choosing a short window duration was to capture the spectral properties of the burst alone, and to exclude any aspiration present during the time between the release and the onset of voicing.

Formant Trajectories in the Vowels

Formant frequencies at two points in the vowel are of particular interest: the onset and the midpoint of the steady-state. These two points illustrate the transition from the consonant to the vowel, which reflects changes in the vocal-tract configurations from the preceding to the following phoneme. The first three formant frequencies at these two points were measured using an LPC-based technique (described earlier). Figure 3.8 shows an example of an LPC spectrum sampled during the steady-state part of the vowel /aa/ in a /ʕaa/ utterance, from which formant frequencies were determined from the peaks in the spectrum envelope.

First, let us consider the steady-state part of the vowels, where the formant frequencies (especially F1 and F2) are indicative of vowel quality. For each vowel, these values were averaged for each speaker individually, and the results were then averaged across speakers (Tables 3.3 and 3.4). Large differences in F3 are apparent from one speaker to another. Hence, the vowels will be characterized, as has been traditionally the case, by the values of the first and second formant frequencies. Figure 3.9 is a plot of the averaged values of F1 and F2 for the three vowels. These results are acoustic manifestations of the vowel features: for /aa/, (+low, + back), for /ii/, (+high, −back), and for /uu/, (+high, +back).

Second, the formant trajectories in the vowels were examined. For each consonant, the first three formant frequencies at the onset of each trajectory were measured, and the average values (pooled across all subjects) are shown in Table 3.5.[7] Although exact

[7] The values of the onset frequencies for each speaker are shown in separate tables in the Appendix.

values of the formant frequencies vary from one speaker to another, due to different vocal-tract dimensions and/or pronunciations, there was a particular trajectory "pattern" associated with each class (i.e., pharyngeal or uvular) and context as seen in Figure 3.10. These trajectories can be described as follows:

- For the pharyngeals /ʕ,ħ/, the F1 trajectory falls from the consonant to the vowel, regardless of the context. For the uvulars /ʁ,χ,q/, the F1 trajectory falls into the vowels /ii/ and /uu/, and rises into /aa/. What differentiates the two classes in the context of high vowels is that the difference in F1 between the onset and the steady-state portions of the vowel is greater for pharyngeals than it is for uvulars. On average, this difference, with high vowels, is 135 Hz when the vowel is preceded by a pharyngeal, and 70 Hz when preceded by a uvular.

- If the voiced and voiceless consonants in each class are compared with one another (i.e., /ʕ/ vs. /ʁ/ and /ħ/ vs. /χ/ and /q/), then we notice that the onset values of F3 are lower when the vowel is preceded by a pharyngeal. F2 onset values for the vowel /uu/ are higher when following a pharyngeal. For the other two vowels, /aa/ and /ii/, no such statements could be made, since there were considerable differences from one speaker to another.

Exact values of the formant frequencies in the consonants will be discussed in the following sections.

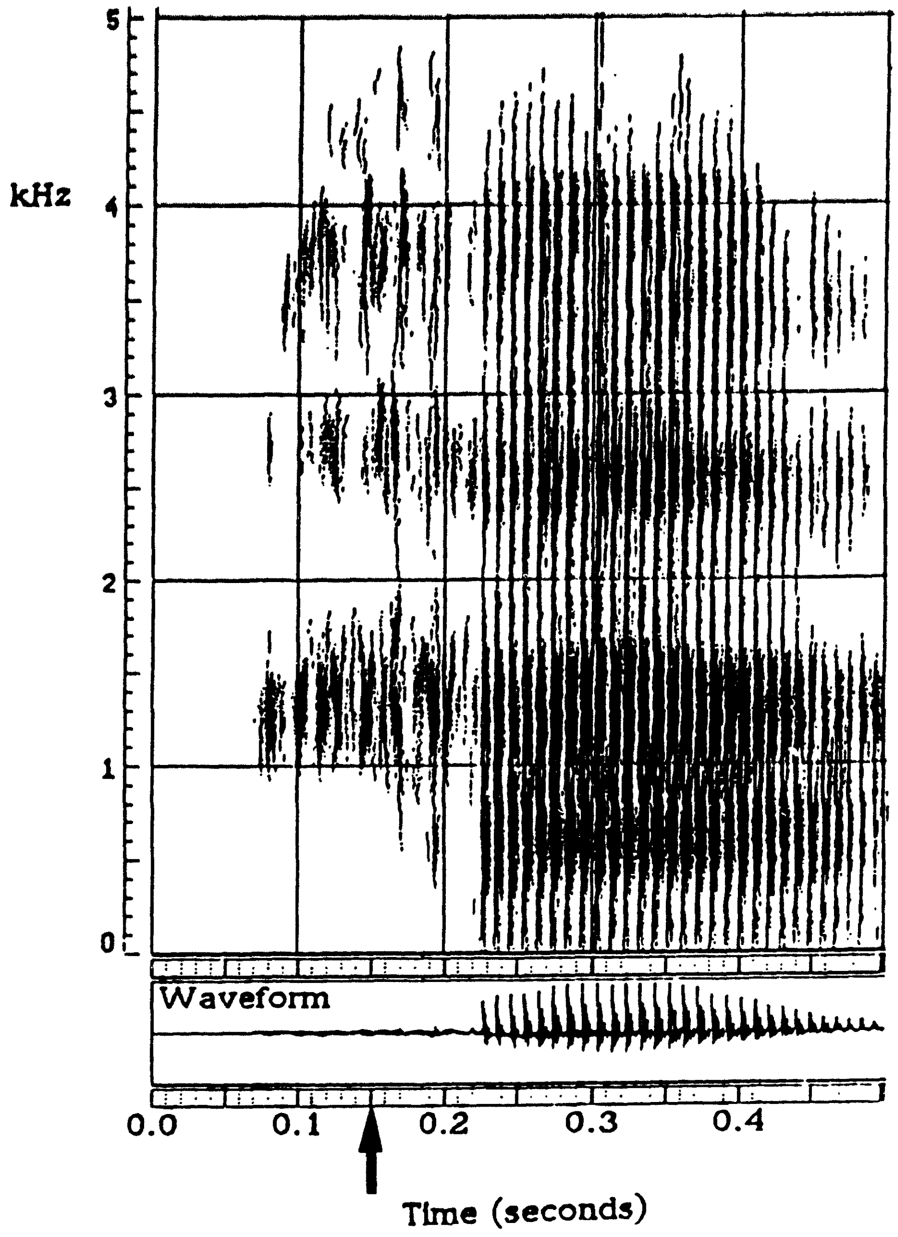


Figure 3.7: Spectrogram of the utterance /χaa/ by speaker JM. The arrow indicates the time where the spectrum of /χ/ was sampled.

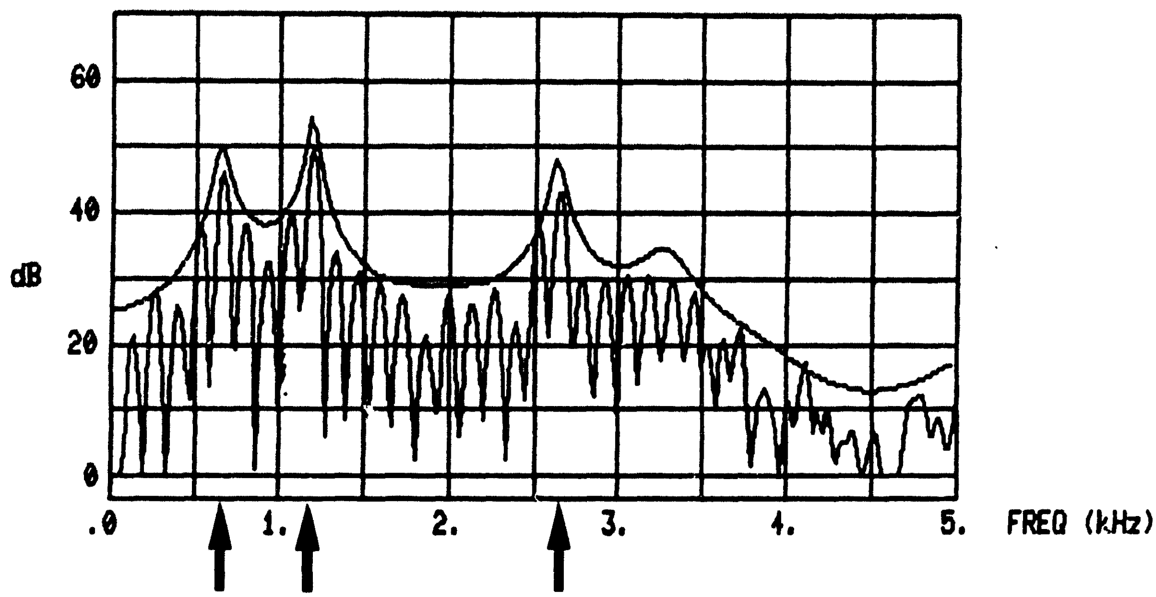


Figure 3.8: DFT and LPC spectra, sampled at the midpoint of the steady-state portion of the vowel /aa/ in /ʕaa/, produced by speaker HA. Formant frequencies are indicated on the spectra by arrows.

Formant Frequencies (Hz)			
	aa	ii	uu
Subj. JM			
F1	694	294	337
F2	1188	2188	812
F3	2408	2796	2442
Subj. MU			
F1	685	306	347
F2	1257	2180	768
F3	2404	2876	2279
Subj. EA			
F1	623	306	357
F2	1186	2230	808
F3	2628	2681	2558
Subj. LT			
F1	641	336	360
F2	1192	2254	756
F3	2438	2652	2379

Table 3.3: Average values in Hz of the first three formants for the vowels: /aa/, /ii/, and /uu/. The average, for each speaker, was based on measurements at the steady-state portion of the vowel in ten utterances.

Formant Frequencies (Hz)			
	aa	ii	uu
F1	661 (43)	311 (30)	350 (59)
F2	1206 (64)	2213 (77)	786 (57)
F3	2470 (163)	2751 (159)	2414 (172)

Table 3.4: Average values in Hz and standard deviations (in parentheses) of the first three formants for the three vowels. Average values, were pooled across four speakers, and were based on measurements made at the steady-state portion of the vowel.

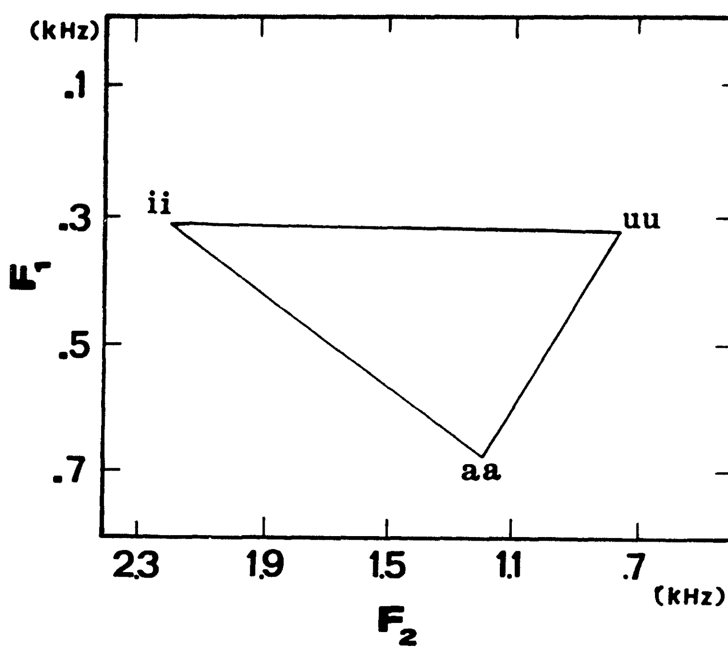


Figure 3.9: Vowel diagram.

Formant Frequencies (Hz)					
	ʕ	ħ	ʁ	χ	q
C/aa/					
F1	728 (70)	739 (45)	542 (46)	584 (30)	587 (31)
F2	1181 (80)	1266 (88)	1235 (122)	1294 (67)	1143 (105)
F3	2224 (71)	2321 (170)	2510 (134)	2553 (227)	2522 (189)
C/ii/					
F1	492 (48)	453 (43)	387 (33)	382 (41)	409 (58)
F2	1887 (129)	1937 (135)	1685 (170)	1987 (134)	1784 (244)
F3	2561 (177)	2592 (73)	2607 (179)	2604 (122)	2574 (95)
C/uu/					
F1	464 (72)	453 (23)	387 (23)	403 (28)	401 (34)
F2	1065 (126)	1181 (119)	709 (90)	790 (84)	782 (77)
F3	2022 (97)	2084 (88)	2580 (166)	2472 (179)	2562 (176)

Table 3.5: Average values with standard deviations (in parentheses) of the first three formants at the onsets of the vowels preceded by one of the five consonants. The average was taken across four speakers.

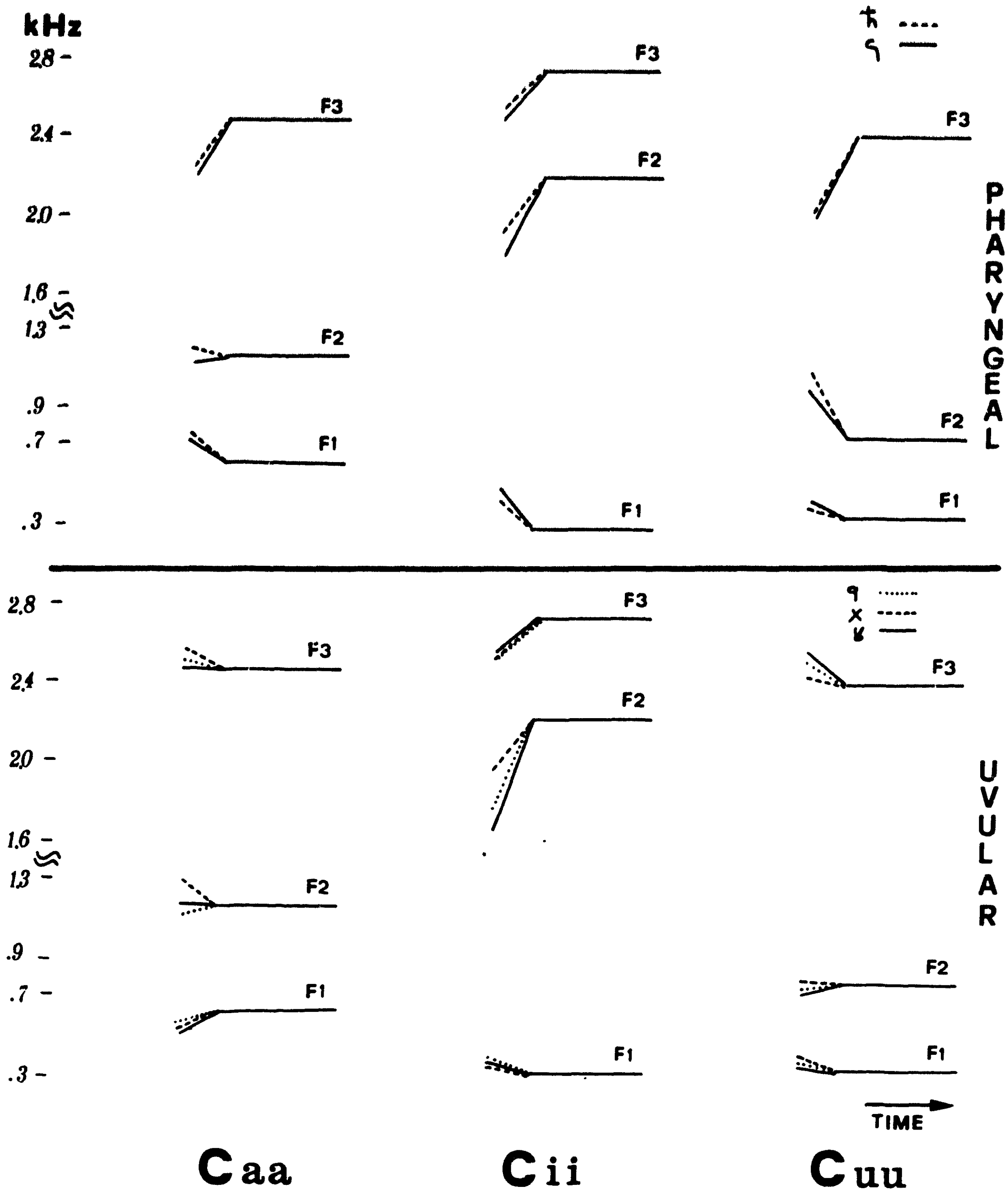


Figure 3.10: Formant trajectories of the five consonants in different contexts. Formant frequencies at the onset and steady-state portions of the vowels represent average values pooled across four speakers.

The Voiced Consonants /ʕ/ and /ʁ/

In Section 3.3 we mentioned that two realizations were observed for the voiced pharyngeal /ʕ/ intervocalically: a stop-like allophone and a continuant allophone. We also mentioned that the voiced uvular /ʁ/ was characterized, intervocalically, by a clear first formant structure.

Spectrograms of the voiced pharyngeal in /ʕV/ utterances also show differences, presumably reflecting various degrees of constriction used in producing this consonant. In Figure 3.11 we show three realizations of this consonant preceding the vowel /ii/: (a) stop-like, aspirated, (b) voiced fricative, and (c) sonorant-like (no evidence of noise). The most common realization (90% of the total number of tokens) was of type (c), that is, a sonorant-like with no evidence of noise. Similarly, spectrographic displays of the voiced uvular /ʁ/ show no evidence of noise except in one token (Figure 3.11d).

The values of the first three formant frequencies were measured at the onset of each consonant. Table 3.6 summarizes the results of these measurements. The measurements show that F1 for the pharyngeal are higher than that for the uvular and F3 is lower. The second formant is higher for the pharyngeal in /Cii/ and /Cuu/ contexts. Preceding the vowel /aa/, F2 does not seem to be distinguishable for the two sounds. This was predicted in Chapter 2, where the calculated value of F2 for the closed-glottis case (voiced) was approximately the same for both the pharyngeal and the uvular models.

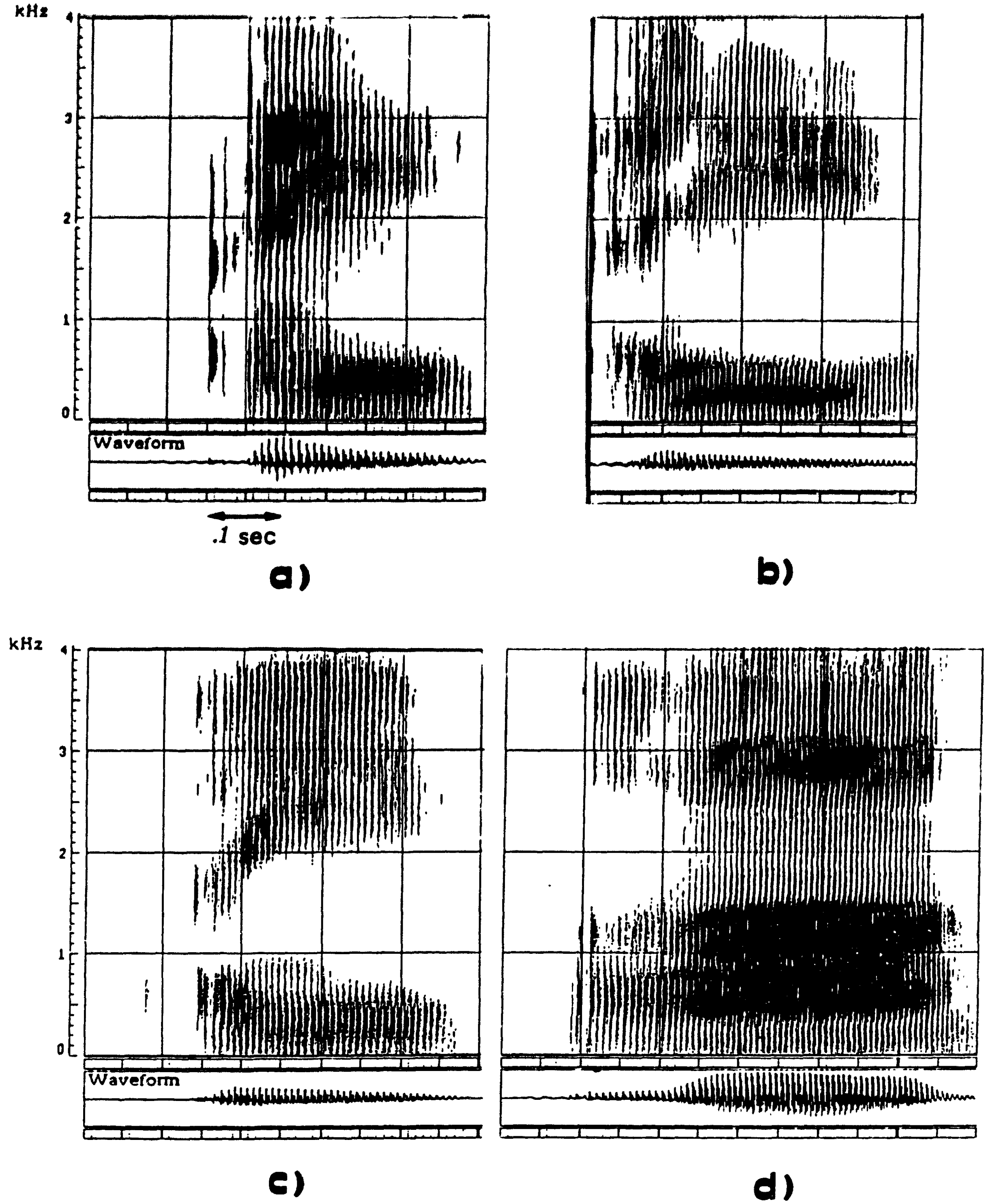


Figure 3.11: a), b), and c) are three realizations of the voiced pharyngeal /ʕ/ in /ʕii/ contexts; a) stop-like (speaker LT), b) voiced-fricative (speaker HA), c) sonorant-like (no evidence of noise) (speaker HA). d) The only token of /ʁ/ where there was evidence of noise (speaker HA).

FIRST THREE FORMANTS (Hz) AT THE ONSETS OF /ʕ/ and /ʁ/						
	ʕ			ʁ		
	F1	F2	F3	F1	F2	F3
/Caa/						
JM	750	1100	2200	478	1350	2470
HA	730	1240	2240	500	1000	2800
MU	830	1280	2170	520	1320	2380
LT	840	1148	2200	470	1230	2530
/Cii/						
JM	618	1670	2240	360	1250	2490
HA	640	1600	2500	470	1200	2800
MU	580	1800	2300	387	1588	2410
LT	530	1600	2233	448	1367	2470
/Cuu/						
JM	560	920	1950	372	821	2470
HA	500	1280	2060	480	?	2880
MU	440	1085	1900	440	700	2450
LT	497	1181	2000	400	690	2600

Table 3.6: Values of the first three formants in Hz, measured at the onset of the voiced consonants /ʕ/ and /ʁ/ in all vowel contexts, produced by the four speakers JM, HA, MU, and LT. The values shown are based on measurements from LPC spectra. A question mark indicates that the formant frequency was not measurable.

The Voiceless Fricatives /ħ/ and /χ/

The noise interval for the voiceless fricatives appears spectrographically to have clear formant structure (Figure 3.7). During a brief interval (10-15 msec.) immediately preceding the onset of the following vowel, the intensity of the noise appears to become weaker and in some cases a period of silence in that interval is observed. As mentioned earlier, DFT and CB spectra were sampled in the consonantal interval with a Hamming window of duration 50 msec. The window was carefully positioned to insure that the formants in the interval were steady.

First, we will explore the formant structure for these consonants. The peaks in the DFT spectra were associated with formant frequencies by comparison with the formant trajectories obtained earlier. The amplitudes of the spectral peaks were measured and normalized with respect to the highest peak in each spectrum, that is, the highest peak in the spectrum is assigned an amplitude of 0 dB. In some cases, none of the first three formants was assigned an amplitude of 0 dB; this indicates that the highest peak in the spectrum was at a higher formant than the third. The results of these measurements, one token for each speaker, are summarized in Tables 3.7 and 3.8.

The entries in Tables 3.7 and 3.8 are not complete; F1 could not be located for the uvular /χ/ for any speaker, and some of the values for F2 and F3 were not measurable (for example F2 in /ħuu/ for speaker JM). This indicates that these formants were not excited, and consequently, no peaks in the DFT spectra were observable.

The following conclusions could be drawn from these measurements:

- The second formant is higher for the pharyngeal /ħ/ than it is for the uvular /χ/ and F3 is lower in all contexts (for the measurable cases). This was not true for the voiced consonants /ʕ,ʁ/ when preceding the vowel /aa/; there, F2 values for the two consonants were approximately the same. A similar result was shown in Chapter 2, where we showed that when the glottis is closed (voiced case), F2 for both the pharyngeal and the uvular models is approximately the same (recall that our models were appropriate for the consonants preceding a low vowel). However, an open glottis (voiceless case) causes F2 for the pharyngeal to be shifted upward, due to the reactive part of the glottal impedance, resulting in a higher F2 for the pharyngeal than for the uvular.

For each speaker, the value of F2 for the two consonants is higher with /ii/ than it is with the two back vowels. Furthermore, the F2 values for /χ/ and F1 and F3 for /ħ/ are lower significantly when the consonant precedes the vowel /uu/ than they are when preceding /aa/. This was predicted in Chapter 2, since anticipatory lip rounding, which presumably occurs for the vowel /uu/, would lower mainly the front-cavity resonances (predicted to be F2 and F4 for the uvulars, and F1 and F3 for the pharyngeals).

- The second formant for the uvular /χ/ was strongly excited in all contexts, which verifies our prediction that F2 is a front-cavity resonance for this consonant. The third formant appeared to be a back-cavity resonance, with a few exceptions (for example, /χaa/ by speaker JM). The spectra of /χ/ for all speakers showed no evidence for F1 excitation in any context.

- The second and the third formants for /ħ/ were strongly excited (highest peaks in the spectrum) in the three contexts by all speakers except JM. For this speaker only F3 was strongly excited with the back vowels /aa/ and /uu/, and both F2 and F3 were excited with /ii/. In Chapter 2, we predicted that F2 for pharyngeal consonants should be a Helmholtz resonance which may or may not be cancelled by a zero depending on the pressure-source location. In addition, we predicted that this formant (F2) would be damped mainly by the glottal losses. These two factors (a possible zero nearby F2 and a widened bandwidth) led us to speculate that this formant would not be strongly excited by the noise source. However, the analysis of the natural utterances revealed that F2, in most cases, is excited. One explanation for F2 excitation is that F2 is not cancelled by a zero, and its amplitude is boosted due to F3 strong excitation, since F2 and F3 for this consonant are relatively close together (on average, within 500 Hz). The degree of F1 excitation was low in comparison with that of F2 and F3, and in some cases (for example, /ħuu/ by speakers MU and LT) F1 was not excited. This is yet another instance of agreement with the predictions made in Chapter 2, namely, that F1 for the pharyngeal is not expected to be as strongly excited as F3 is, because F1 amplitude is lowered due to the presence of a low-frequency zero, and by its relatively wide bandwidth (see Sections 2.3.1, 2.4.2).

Now we turn to a different representation of the spectral properties for these consonants, namely, critical-band spectra. This representation provides us with measures of the gross spectral shape of the consonants.

In Figures 3.12 through 3.14 we show critical-band spectra for the two consonants preceding the three vowels, one token for each speaker. The spectra for the uvular /χ/ could be characterized as *compact* with narrow peaks centered at frequencies between 1100-1500 Hz preceding /aa/, 1500-1800 Hz preceding /ii/, and 700-1000 Hz preceding /uu/. For speakers HA and MU, the peak in the /χuu/ spectrum was less pronounced than it was for the other speakers. The spectra for the pharyngeal /ħ/ are characterized by two peaks: a broad peak centered at high frequencies (2000-2500 Hz preceding /aa/ and /ii/, and 1500-2000 Hz preceding /uu/), and a narrow peak of lower amplitude at low frequencies (1000 Hz with /aa/, 750 Hz with /ii/ and 500-600 Hz with /uu/). The exact frequencies of the peaks vary depending on the speaker. What is consistent in the spectra for all speakers is that the highest peak in each spectrum is located in the region of F3 for /ħ/ and F2 for /χ/. The peak in the pharyngeal case is broad since both F2 and F3 are generally excited (Table 3.8), whereas for /χ/ F2 is the only resonance excited in most cases. Analysis of the second tokens shows results similar to those reported here, and will not be discussed.

The Voiceless Stop /q/

Critical-band spectra sampled at the burst of the voiceless uvular stop /q/ for the four speakers in different vowel contexts are shown in Figures 3.15 and 3.16. The spectral shape is quite similar to that of the voiceless uvular /χ/, in terms of having a compact spectrum peak centered at the second formant frequency, which varies according to the following vowel: 1000-1300 Hz preceding /aa/, 1500 Hz preceding /ii/, and 800 Hz preceding /uu/.

F1 (Hz) and normalized amplitude (dB)			
	ħaa	ħii	ħuu
JM			
F1	1000 (-10)	750 (-7)	560 (-6)
HA			
F1	960 (-14)	780 (-15)	600 (-20)
MU			
F1	960 (-10)	600 (-16)	?
LT			
F1	?	650 (-20)	?

Table 3.7: Values of the first formant in Hz, and its normalized amplitude in dB (shown in parentheses), for the voiceless consonant /ħ/ in all vowel contexts, produced by the four speakers JM, HA, MU, and LT. The values shown are based on measurements from DFT spectra. A question mark indicates that the formant frequency was not measurable.

F2 and F3 (Hz) and normalized amplitude (dB)						
	/Caa/		/Cii/		/Cuu/	
	ħ	χ	ħ	χ	ħ	χ
JM						
F2	?	1328 (0)	1726 (-3)	1497 (0)	?	990 (0)
F3	2285 (0)	2500 (-4)	2118 (0)	?	2000 (0)	2647 (-10)
HA						
F2	1667 (0)	1177 (0)	1784 (-4)	1420 (-2)	1650 (-2)	700 (-3)
F3	2578 (-2)	?	2686 (0)	?	2340 (0)	?
MU						
F2	1600 (0)	1500 (0)	1800 (-1)	1783 (0)	1445 (-2)	1000 (0)
F3	2122 (0)	?	2100 (0)	2675 (-2)	1920 (0)	?
LT						
F2	1588 (-3)	1294 (0)	1870 (-5)	1650 (0)	1500 (-2)	1000 (0)
F3	2056 (0)	?	2500 (0)	?	1963 (0)	2413 (-5)

Table 3.8: Values of the second and third formants in Hz, and their normalized amplitudes in dB (shown in parentheses) for the voiceless consonants /ħ/ and /χ/ in all vowel contexts, produced by the four speakers JM, HA, MU, and LT. The values shown are based on measurements from DFT spectra. A question mark indicates that the formant frequency was not measurable.

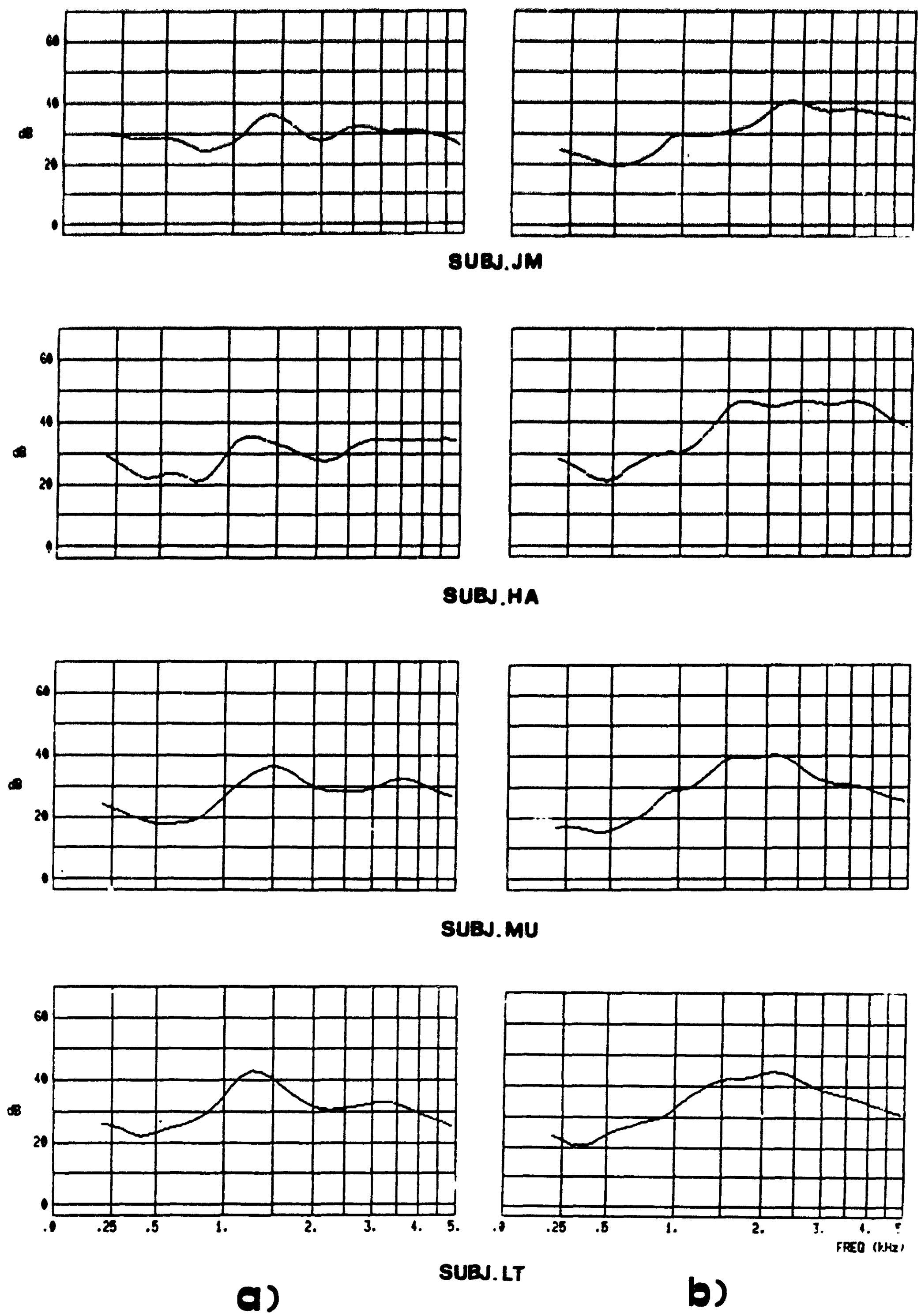


Figure 3.12: Critical-band spectra sampled during the consonantal intervals of a) /χaa/ and b) /ħaa/, for the four speakers JM, HA, MU, and LT.

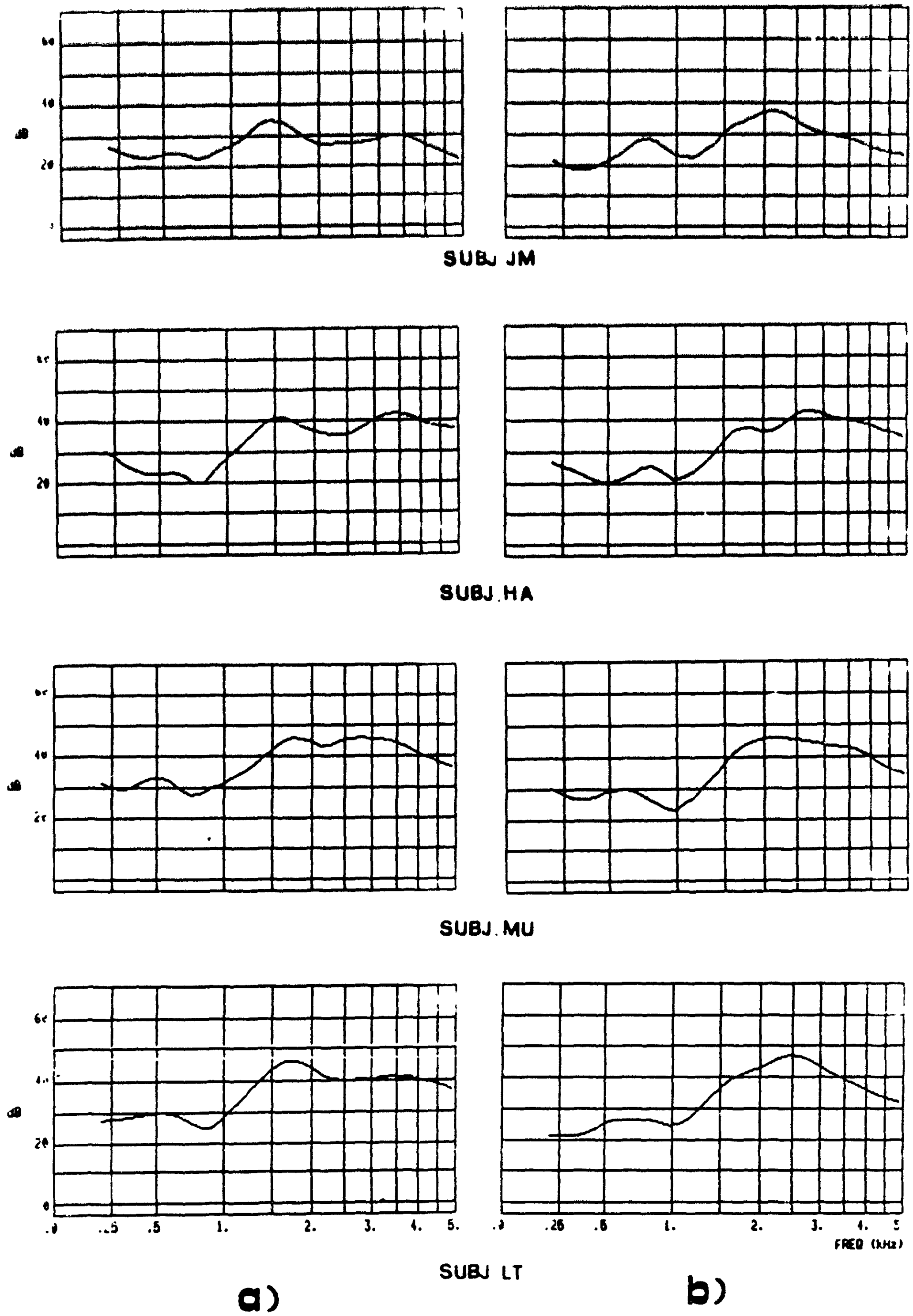


Figure 3.13: Critical-band spectra sampled during the consonantal intervals of a) /χii/ and b) /ħii/, for the four speakers JM, HA, MU, and LT.

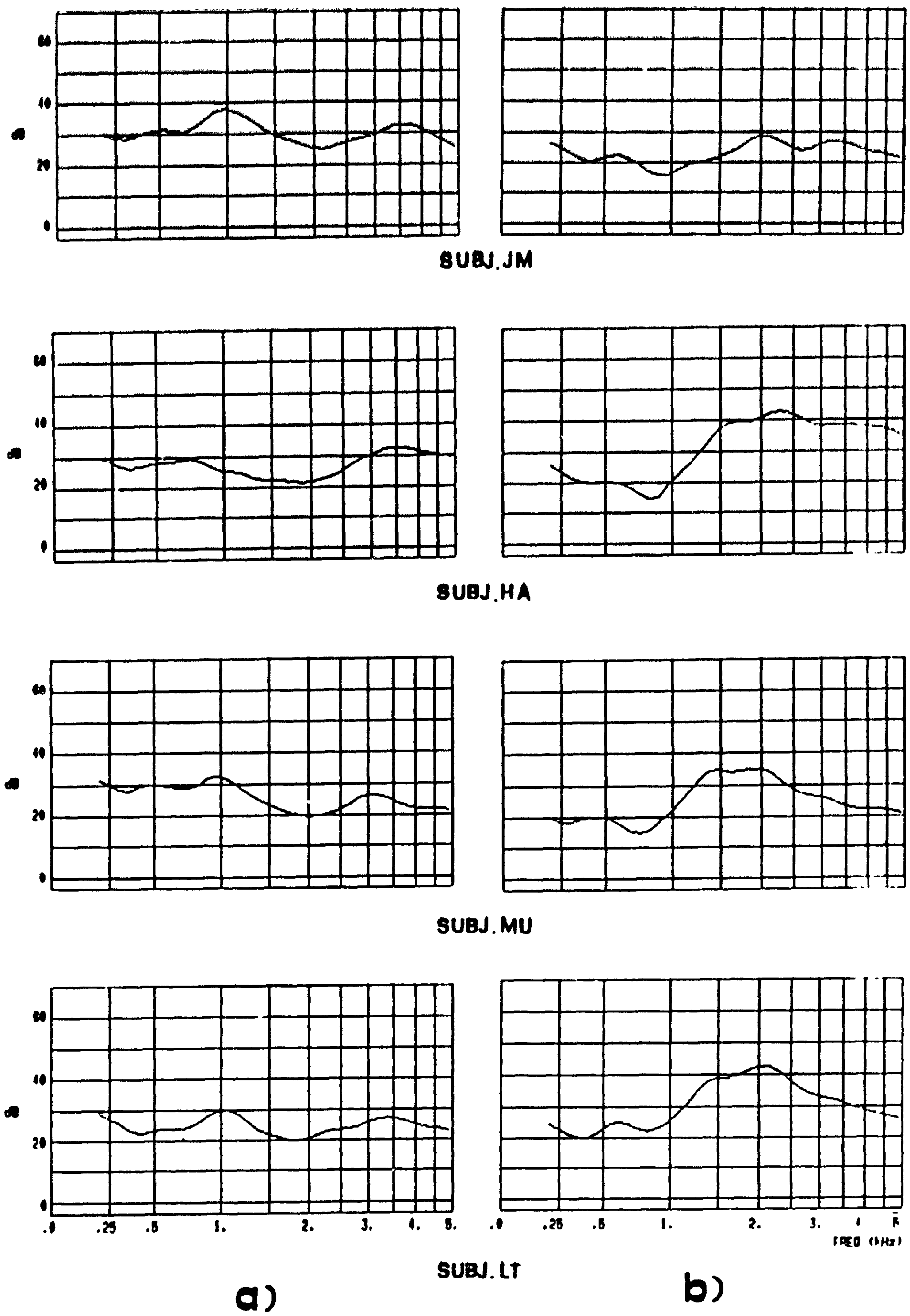


Figure 3.14: Critical-band spectra sampled during the consonantal intervals of a) /χuu/ and b) /ħuu/, for the four speakers JM, HA, MU, and LT.

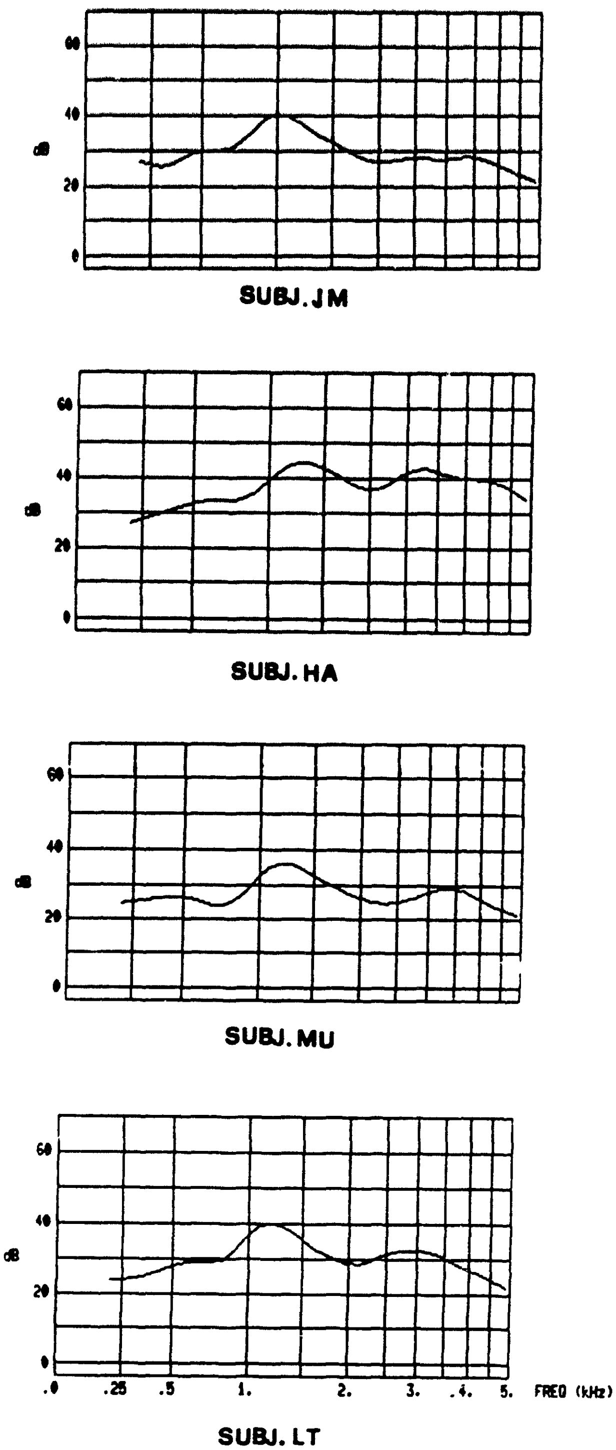


Figure 3.15: Critical-band spectra sampled at the burst of the voiceless uvular stop /q/ in a /qaa/ context. The four spectra represent one token for the speakers JM, HA, MU, and LT.

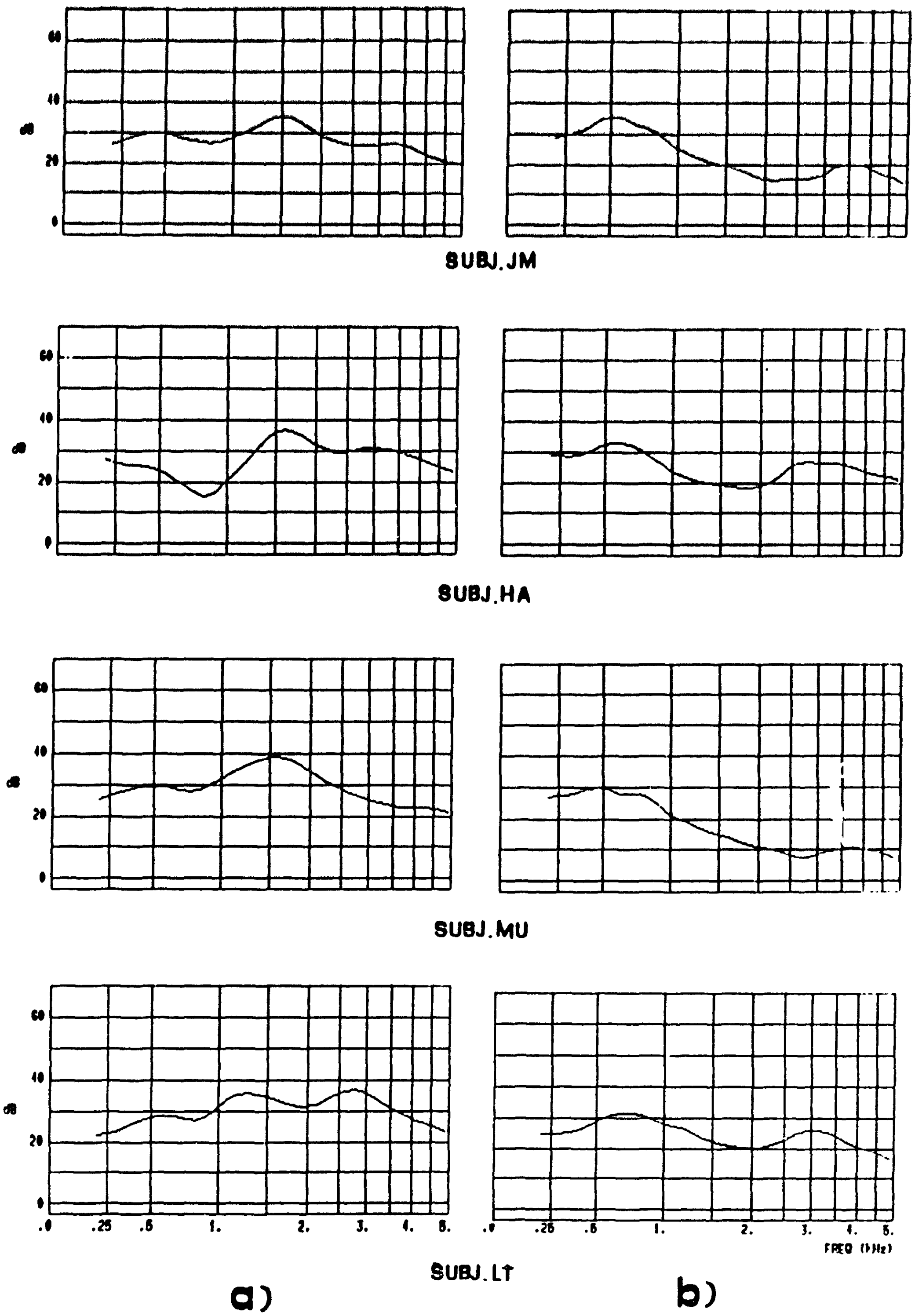


Figure 3.16: Critical-band spectra sampled at the burst of the voiceless uvular stop /q/ in: a) /qii/, b) /quu/ contexts. The four spectra represent one token for the speakers JM, HA, MU, and LT.

The *F* Pattern

In Chapter 2, we showed how different vocal-tract configurations result in different resonances (formant frequencies). Ideally, one might want to associate each sound with a particular configuration, regardless of the context, that is, to associate a set of "target" frequencies for each sound/configuration. We will adopt the notion of acoustic "loci" to describe these frequency targets.[1] Let us define these targets to be the average formant frequencies at the onset of the consonants if voiced, at the steady-state part of the voiceless fricatives, and at the burst for the stop. If a formant frequency is not measurable then the frequencies at the onset of the following vowel will be approximated as targets. We can then use the results obtained earlier in describing the target frequencies associated with each sound and context. Table 3.9 shows the average target frequencies, pooled across four speakers, for each consonant in different vowel contexts.

The results of the acoustic analysis do not show evidence for the existence of loci that are independent of the context, because of coarticulatory effects of the following vowels. What these results show instead is that (1) the F1 locus for both classes varies depending on the height feature of the following vowel, where the locus is higher preceding the low vowel /aa/, (2) the F2 locus for both classes and F3 locus for the pharyngeals varies significantly depending on the backness feature of the following vowel (that is, the locus in both cases is lower preceding back vowels), and (3) with back vowels, the F2 locus for the uvulars and F3 locus for the pharyngeals is lower if the following vowel is rounded (i.e., /uu/ vs. /aa/).

[1]At this point, the term acoustic "loci" has no perceptual justification, and hence is different from that adopted by Delattre, et al. (1955).

FREQUENCY TARGETS					
	Pharyngeal		Uvular		
	ʕ	ħ	ʁ	χ	q
F1					
/Caa/	**787**	**970**	**492**	**584**	**587**
/Cii/	**592**	**695**	**416**	**382**	**409**
/Cuu/	**500**	**580**	**423**	**403**	**401**
F2					
/Caa/	**1192**	**1618**	**1225**	**1324**	**1176**
/Cii/	**1667**	**1795**	**1351**	**1587**	**1478**
/Cuu/	**1116**	**1531**	**737**	**922**	**636**
F3					
/Caa/	**2202**	**2260**	**2545**	**2500**	**2522**
/Cii/	**2318**	**2351**	**2540**	**2600**	**2574**
/Cuu/	**1977**	**2055**	**2600**	**2562**	**2562**

Table 3.9: **Targets of the first three formants in Hz, defined as the average formant frequencies measured at the onset of the voiced consonants, at the steady state part of the voiceless fricatives, and at the burst for the stop. The values are averaged across four speakers (see text for details).**

3.4.4 Helmholtz Resonance Bandwidth for /ʕ/ and /ʁ/

In Chapter 2, it was predicted that when the glottis is closed (voiced) then the contribution of the constriction losses to the bandwidth of the Helmholtz resonance (F2 for the pharyngeal and F1 for the uvular) is significant. This prediction was derived from a low-frequency model of the vocal tract. In this section, we will examine the formant structure (frequencies and amplitudes) of the two voiced consonants /ʕ, ʁ/ in natural utterances, and compare the measurements with values predicted for the no-constriction-loss case to determine whether there is any evidence supporting the prediction.

Method

As stated earlier, we are primarily concerned in this section with examining the bandwidth of the Helmholtz resonance for the voiced pharyngeal and uvular consonants to determine whether there is evidence for the bandwidth contribution of the *constriction* losses. The CV utterances were excluded from the analysis in this part because of possible additional losses at the initial phase of the glottal opening, which could contribute to an additional widening of the bandwidths of the formants. We restricted the analysis to ʔVCV utterances in symmetric context only, i.e., /ʔaaCaa/, /ʔiiCii/ and /ʔuuCuu/.

Discrete Fourier transform (see Section 3.4.1) spectra were computed at two points in each utterance: at the midpoint of the consonantal interval, and in the following vowel at 50 msec. after the consonant/vowel boundary. From each spectrum, the first harmonic (F0), the first three formant frequencies (F1, F2, and F3) (in Hz) and their amplitudes A0, A1, A2, and A3 (in dB) were measured.

We then predicted the changes in the amplitudes of F0 and the first two formant frequencies between the consonant and the adjacent vowel by the following procedure.

The transfer function of the volume velocity at the lips to that at the glottis ($U_o(s)/U_g(s)$) is an all-pole transfer function (Fant, 1960), and could be expressed by the following equation:

$$\frac{U_o(s)}{U_g(s)} = \prod_{i=1}^{n} \frac{(s_i)(s_i^*)}{(s-s_i)(s-s_i^*)} \tag{3.1}$$

where each complex frequency (s_i) is described by an imaginary part ($w_i = 2\pi F_i$) and a real part (σ_i). The imaginary part refers to a formant frequency, and the real part to its bandwidth. If we assume that the bandwidth of each formant is much less than the formant-frequency location ($\sigma_i \ll w_i$), then, the amplitude of each formant (A_i) in dB could be approximated by the following expression:

$$A_i \cong 20 \log \frac{\pi F_i}{\sigma_i} \prod_{k=1, k\neq i}^{n} \frac{F_k^2}{(F_i - F_k)(F_i + F_k)} \tag{3.2}$$

The formant frequencies for the pharyngeal and uvular consonants are different from those in the vowels (see Section 3.4.3). In accord with Equation 3.3, we would expect that the amplitude of each formant frequency (A_i) will differ between these consonants and the adjacent vowels in the ?VCV utterances. Furthermore, if we assume that the formant bandwidths (σ_i) are the same for both the consonants and the vowels then we can predict the change in the formant-frequency amplitudes by the following equation:

$$\frac{A_{iv}}{A_{ic}} \cong 20 \log \frac{F_{iv}}{F_{ic}} \prod_{k=1, k\neq i}^{n} \frac{F_{kv}^2 (F_{ic} - F_{kc})(F_{ic} + F_{kc})}{F_{kc}^2 (F_{iv} - F_{kv})(F_{iv} + F_{kv})} \tag{3.3}$$

The subscripts in the above equation (v,c) refer to the vowel and the consonant, respectively.

Equation 3.4 was used to predict the changes in A1 and A2 due to the shift in the first three formant frequencies (n=3) between the consonant and the vowel. The change in the first-harmonic amplitude (A0) due to this shift was predicted using the following equation:

$$\frac{A_{0v}}{A_{0c}} \cong 20 \log \prod_{k=1}^{3} \frac{F_{kv}^2 (F_{0c} - F_{kc})(F_{0c} + F_{kc})}{F_{kc}^2 (F_{0v} - F_{kv})(F_{0v} + F_{kv})} \tag{3.4}$$

In arriving at these equations, we have made an important assumption, that is, each formant bandwidth (σ_i) is the same for both the vowel and the consonant. This assumption implies that the bandwidth contributions of the different losses are the same for both. Specifically, the constriction-loss contribution for the consonants is not being considered. Hence, the predicted values are appropriate for a no-constriction-loss case.

If the predicted changes in the amplitude of the Helmholtz resonance, for the no-constriction-loss case, are less than those measured in natural utterances, this difference could be an evidence for the bandwidth contribution of the constriction loss to the Helmholtz bandwidth.

The measured and predicted changes in A0, A1, and A2 were compared, and the differences between the two values (measured and predicted) were labeled $\triangle$A0, $\triangle$A1, and $\triangle$A2, respectively.

The calculations performed in this section were carried out with the assumption that the glottal spectrum has a slope of -12dB/octave (corrected to 0 slope with radiation and preemphasis) at all frequencies of interest (above 250 Hz). The results of these calculations are shown in Table 3.10.

Results and Discussion

Table 3.10 summarizes the results obtained for $\triangle$A0, $\triangle$A1, and $\triangle$A2 for the pharyngeal and the uvular consonants with the three vowels /aa/, /ii/, and /uu/. The values are listed for each speaker individually and average values across speakers are also shown. Positive/negative values of $\triangle$A0, $\triangle$A1, and $\triangle$A2 indicate a decrease/increase in these amplitudes which are not explained by the shift in formant frequency alone. Explanations for the observed reductions are summarized as follows:

- The reduction in A0, observed for both consonants, is related to a change in the glottal source; specifically this reduction indicates a decrease in the glottal-pulse area during the consonantal interval in comparison with that during the vowel.

- The reduction in A1 could be ascribed mainly to two factors: (1) a reduction in the glottal-pulse area (as indicated by $\triangle$A0), and (2) an increase in F1 bandwidth. The change in F1 amplitude, above that due to the change in the glottal-pulse area (ΔA1$-\Delta$A0) is, on average, -1, -2, and 2 dB for the pharyngeal, and 9, 8, and 4 dB for the uvular, with the three vowels /aa/, /ii/, and /uu/, respectively. We notice that there is a significant reduction in the F1 amplitude (positive value for the difference between ΔA1 and ΔA0) for the uvular which is primarily attributed to an increased F1 bandwidth. This suggests that F1 bandwidth in the uvular consonantal interval is at least one and a half times (4 dB difference) the F1 bandwidth for the no-constriction-loss case. For the pharyngeal, there was no evidence for such a reduction in F1 amplitude, except with the vowel /uu/.

- The factors that cause a reduction in A2 are similar to those affecting A1, namely a reduction in the glottal-pulse area, and formant bandwidth changes. In addition, a less abrupt closure of the glottal pulse would cause a further reduction in the amplitude of F2 and higher formants. The reduction in A2 above that caused by the reduced glottal pulse (ΔA2$-\Delta$A0) is, on average, 4, 17, and 6 dB for the pharyngeal, and 3, 19, and 0 dB for the uvular. These values show that F2 amplitude is decreased due to a less abrupt closing slope of the glottal pulse and/or an increase in F2 bandwidth. We cannot precisely estimate which of the two factors is more significant, since we lack detailed information regarding the glottal-pulse shape.

In summary, the results introduced in this section show that the amplitudes of the Helmholtz resonances (F2 for the pharyngeal and F1 for the uvular) are decreased in comparison with those for the no-constriction-loss case. The reduction in the case of the uvular was attributed to a widened F1 bandwidth, supporting earlier theoretical predictions. For the pharyngeal, it was not clear whether the reduction in A2 was mainly due to an increased bandwidth or to a less abrupt glottal-pulse closure, since we lack information regarding the glottal-pulse shape for this consonant.

	/ʔaaʕaa/			/ʔiiʕii/			/ʔuuʕuu/		
	ΔA0	ΔA1	ΔA2	ΔA0	ΔA1	ΔA2	ΔA0	ΔA1	ΔA2
Subj.									
JM	7	12	10	8	8	25	7	10	17
MU	9	10	19	13	10	37	3	12	11
HA	11	11	14	14	6	25	5	6	6
LT	5	-3	4	7	10	20	2	-5	5
Avg.	8	7	12	10	8	27	4	6	10

	/ʔaaʁaa/			/ʔiiʁii/			/ʔuuʁuu/		
	ΔA0	ΔA1	ΔA2	ΔA0	ΔA1	ΔA2	ΔA0	ΔA1	ΔA2
Subj.									
JM	5	20	12	4	16	20	12	15	-3
MU	7	15	7	10	22	31	7	10	9
HA	8	17	7	7	8	21	10	12	18
LT	5	10	10	6	15	33	5	10	10
Avg.	6	15	9	7	15	26	8	12	8

Table 3.10: ΔA0, ΔA1, and ΔA0 in dB computed for each speaker individually, and average values across speakers. The results are shown for both the pharyngeal /ʕ/ and the uvular /ʁ/.

3.5 Summary

In this chapter, we have quantified some of the acoustic properties of pharyngeal and uvular consonants. The properties examined were duration (Section 3.3), fundamental frequency of the voiced consonants (Section 3.4.2), the *F*-pattern associated with each consonant and context, and spectral shape of the noise for the voiceless consonants (Section 3.4.3). In addition, we examined the shift in the Helmholtz resonances (F2 for the pharyngeal and F1 for the uvular) locations and amplitudes from the voiced consonants /ʕ,ʁ/ to the vowel in /ʔVCV/ utterances and compared the results with changes in amplitude predicted theoretically (Section 3.4.4).

The results of the analysis show that:

- The duration of the voiceless consonants is longer than that of their voiced counterparts. The voiceless/voiced consonants in both classes have similar duration.
- The fundamental frequency for the voiced consonants /ʕ/ and /ʁ/ is lower than it is for vowels. The change in the fundamental was greatest for the pharyngeal; the explanation given for this change was that the lower pharyngeal region is constricted during the production of this consonant. For the uvular, we attributed the lower fundamental frequency to the acoustic and/or aerodynamic effects of a supraglottal constriction on the vocal-fold vibrational pattern.
- Three allophones were observed for the voiced pharyngeal /ʕ/: stop-like, sonorant-like (no evidence of noise) and a continuant with noise. The voiced uvular /ʁ/ was realized as a continuant or as a sonorant. For both consonants, the most likely realization was as a sonorant. Thus, the feature (***continuant***) is not distinctive for these voiced consonants.
- The acoustic loci for the pharyngeal and uvular consonants are different in each vowel context. This result is not in disagreement with the theoretical predictions in Chapter 2, since the idealized models were appropriate for the production of the consonants preceding a low vowel (like /aa/), and it ignored perturbations in the models, which occur in anticipation of the other two vowels /ii, uu/.

- The first formant is consistently higher for the pharyngeals than it is for the uvulars and the third formant is lower. The second formant is higher for the pharyngeals than it is for uvulars, except with /aa/ where F2 in that case does not distinguish the voiced consonants in the two classes. These results agree with the theoretical predictions introduced in Chapter 2.

- The noise spectrum for the voiceless pharyngeal /ħ/ is characterized by a broad peak centered at F3, and the spectra for the voiceless uvulars /χ/ and /q/ are characterized by a narrow peak centered at F2. Since F2 location varies depending on the following vowel, the peak location in the spectra is different accordingly. The broad peak in the pharyngeal case is attributed to the excitation of F2 and F3, and the narrow peak in the uvular case is due to the excitation of F2 alone. There were few exceptions to this general trend: for the pharyngeal, one speaker did not excite F2 with the back vowels, F1 was excited in some cases, and for the uvular /χ/, F3 was excited for some speakers. The peak locations correspond mainly to front-cavity formants which are strongly excited by the noise source in the vicinity of the supraglottal constriction.

- The amplitudes of the Helmholtz resonances (F2 for the pharyngeal and F1 for the uvular) are less than those predicted for the no-constriction-loss case. The reduction in the case of the uvular was attributed to a widened F1 bandwidth, whereas for the pharyngeal, it was not clear whether the reduction in F2 amplitude was due to an increased F2 bandwidth or to a less abrupt glottal-pulse closure, since we lack information regarding the glottal-pulse shape for these consonants.

In the next chapter we will conduct perceptual experiments to determine which of the acoustic properties predicted in Chapter 2 and analyzed in this chapter are important perceptually.

Chapter 4

Perceptual Experiments

In this chapter we will attempt to find which properties listeners use as cues in identifying the place of articulation for pharyngeal and uvular consonants through perceptual tests. We will restrict the choice of consonants to the voiced pharyngeal and uvular (/ʕ/ and /ʁ/) in /Caa/ context. The properties examined were the trajectories of the first two formant frequencies, and the bandwidths of F2 for the pharyngeal and F1 for the uvular.

4.1 Stimuli

Synthetic stimuli used for both experiments were generated using the Klatt formant synthesizer (Klatt, D.H., 1980) implemented on the MIT-SPEECHVAX. A block diagram of this synthesizer is shown in Figure 4.1. Only the cascade part of the synthesizer was used in generating the synthetic stimuli.

The duration of the stimuli was 465 msec. For each stimulus, the amplitude of voicing (av) rose from 50 to 60 dB linearly, over 20 msec., remained at 60 dB for 425 msec., and then fell, linearly, from 60 to 50 dB in 20 msec. The fundamental frequency (f0) rose from 89 to 111 Hz in 250 msec., and remained fixed at 111 Hz until the end of the stimulus. Plots of the parameters av and f0 as functions of time are shown in Figure 4.2.

Two "standard" stimuli were used in the experiments: type I (pharyngeal) and type II (uvular). For type I, the onset values of F1 and F2 were set at 780 and 1010 Hz, respectively, and for type II, F1 onset value was at 450 Hz and F2, at 1360 Hz. The first two formant frequencies remained fixed at these values for 25 msec. and then traversed to a final value of 680 Hz for F1 and 1160 Hz for F2 in 100 msec., for both types. The onset and final values of F1 and F2 were chosen to be similar to values measured in natural /Caa/ utterances for one of the tokens by speaker JM. The values of F3, F4, and F5 were held fixed for all stimuli at 2400, 3250, and 3700 Hz, respectively.

Plots of the formant-frequency trajectories, varied in a piecewise-linear fashion, used for types I and II are shown in Figure 4.3a and 4.3b, respectively.

The bandwidths of the formant frequencies for the standard stimuli were set at their default values of: B1=60 Hz, B2=90 Hz, B3=150 Hz, B4 and B5=200 Hz.

Three subjects listened informally to a randomized set of six repetitions of the two standard stimuli, without prior knowledge of the goal of the listening test. They were asked to identify the consonant in each stimulus. The subjects identified the consonant in type I as the voiced pharyngeal (95% correct), and in type II as the voiced uvular (93% correct). Hence, these two stimuli were considered to be adequate as reference stimuli for both experiments.

4.2 Subjects

Six male native speakers of Arabic, ranging in age between 21 and 30 years, participated as subjects for the two experiments. Three subjects (MU, AM, MAH) were from the city of Baghdad, Iraq; the other three were from the cities of Kuwait, Kuwait (BA), Beirut, Lebanon (MR), and Khartoum, Sudan (MAL). They have been living in the United States for the last 1-5 years. Only two of the subjects were paid (BA and MAL) for participating in these experiments. None of the subjects had known speech or hearing impairments, previous phonetic training, or exposure to synthetic speech sounds.

4.3 Procedure

Stimuli were repeated six times, randomized, and recorded onto TDK D60 cassettes. The stimuli were presented to the subjects binaurally over headphones in the sound-treated room described in Chapter 3. Four one-hour sessions were conducted, two for each experiment.

Instructions for each experiment were written on the response sheets provided to the subjects, and were explained orally prior to each session. A brief training period preceded each listening session to familiarize the subjects with the computer-generated sounds.

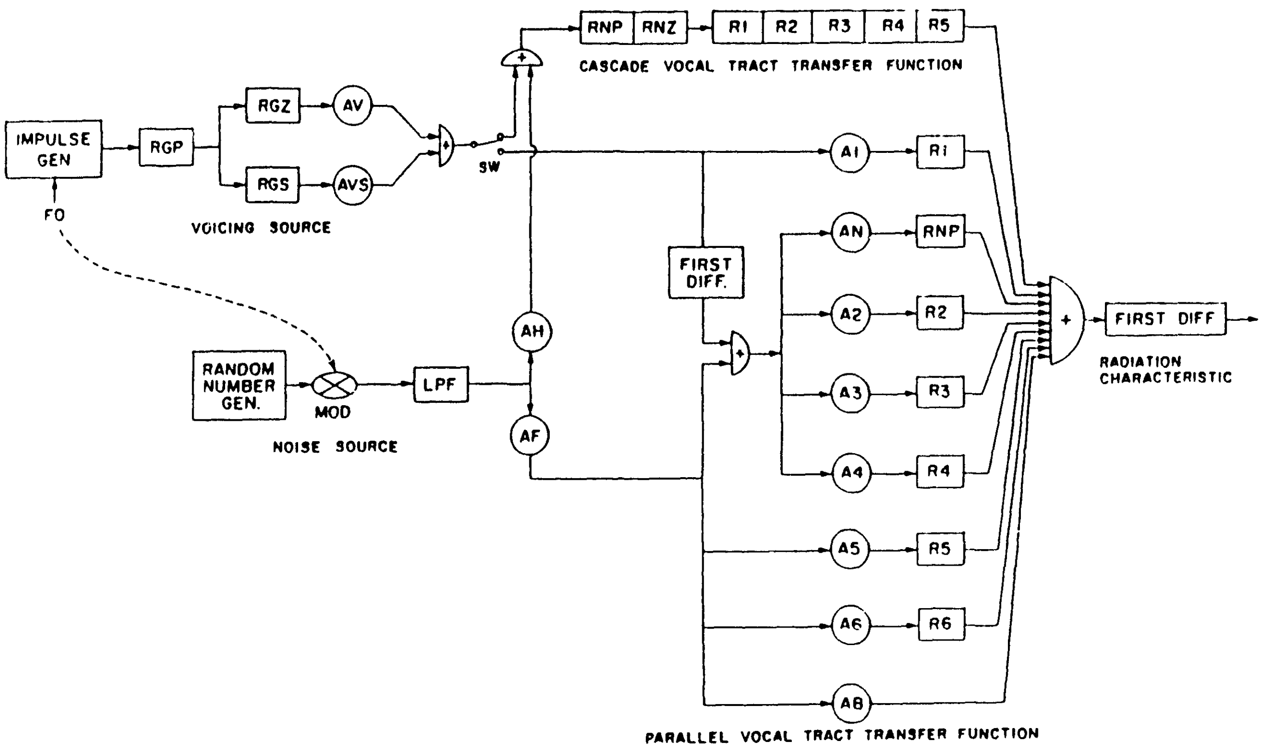


Figure 4.1: A block diagram of the Klatt cascade/parallel formant synthesizer (Klatt, D.H., 1980). Only the cascade part of the synthesizer was used in generating the stimuli for both experiments.

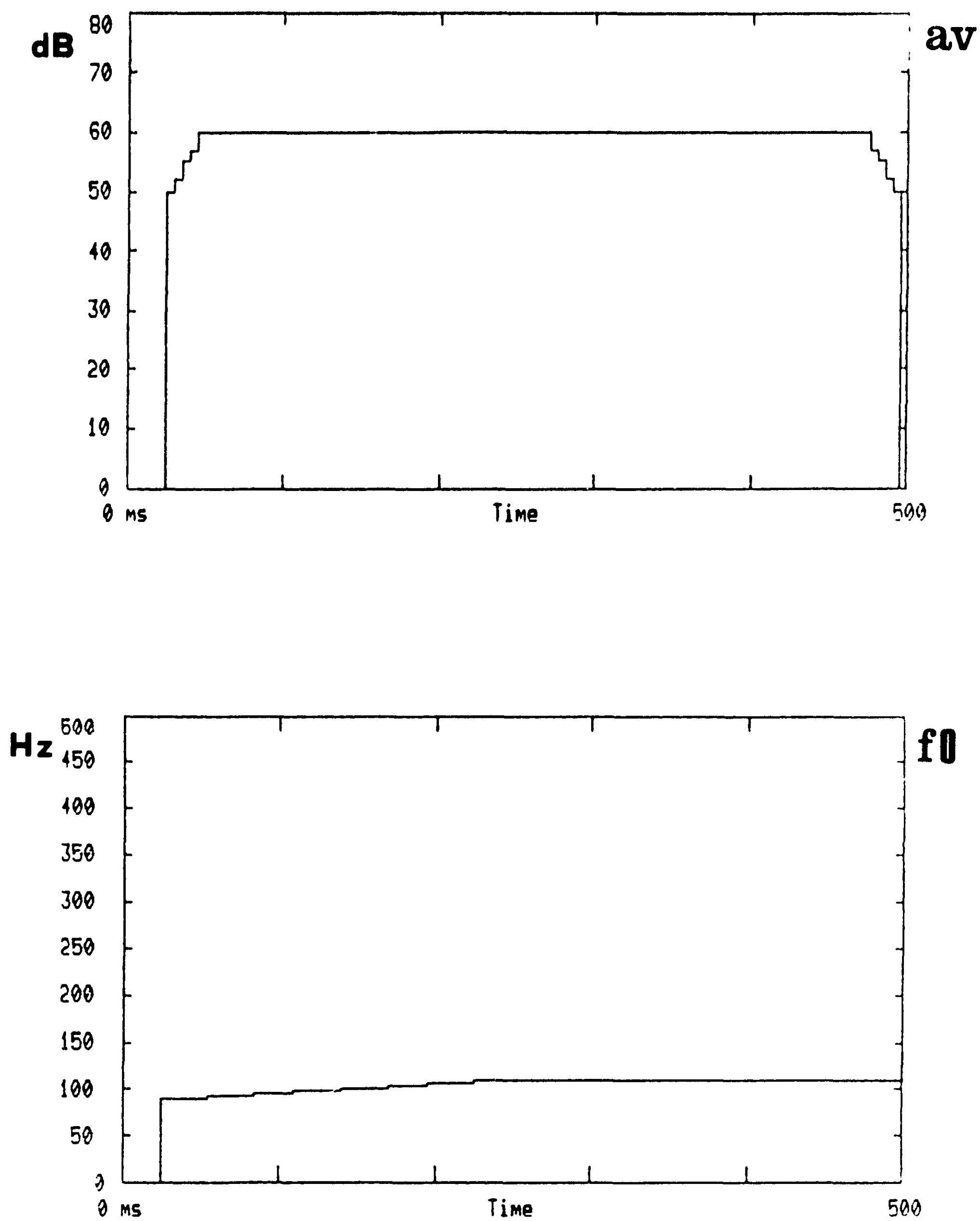


Figure 4.2: Contours of the amplitude of voicing (av) and the fundamental frequency (f0) used for all stimuli.

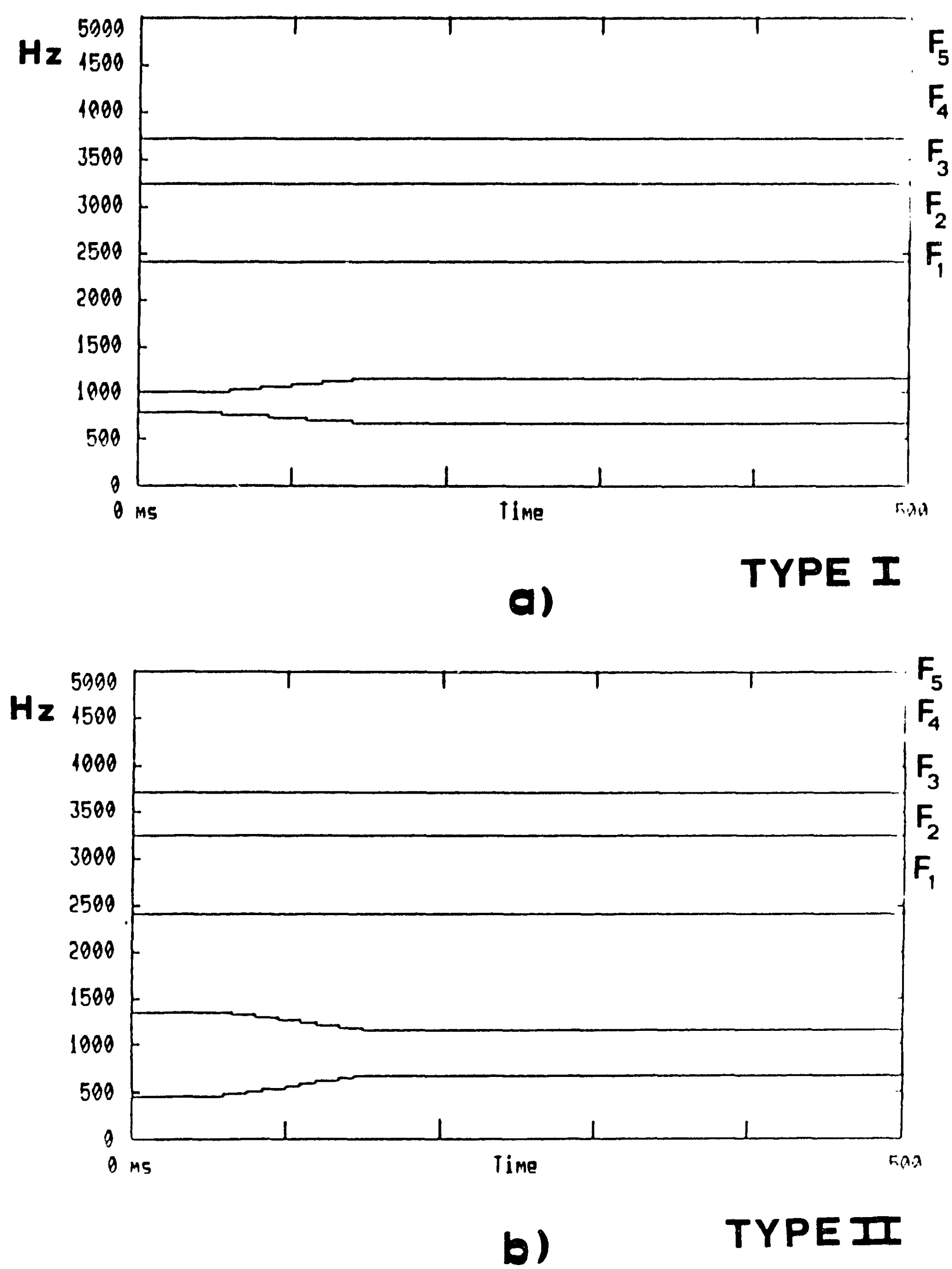


Figure 4.3: Plots of the first five formant frequencies as functions of time for the two "standard" stimuli of a) type I (pharyngeal) and b) type II (uvular).

4.4 Experiment I

4.4.1 Procedure

The goal of the first experiment was to investigate the role of the trajectories of the first two formant frequencies in the perception of the voiced consonants /ʕ/ and /ʁ/ in initial position, preceding the vowel /aa/.

Two continua were formed for each type of stimulus: an F1 continuum and an F2 continuum. The F1 and F2 continua were formed by systematically varying the onset values of these formant frequencies in steps of 50 Hz for F1 and 60 Hz for F2. There were seven stimuli in each continuum, resulting in a total number of 28 stimuli. Schematized spectrograms for the two continua for each type appear in Figures 4.4 and 4.5. In both figures, the trajectories for the reference stimuli are represented by solid lines. An additional stimulus (ss) was synthesized with straight formant frequencies, that is, F1 and F2 were held fixed at their steady-state values (680 and 1160 Hz, respectively).

Prior to the experiment, three subjects (different from those who participated in the experiment) were asked to identify the consonant in all the stimuli. The subjects identified the consonant as either the voiced pharyngeal /ʕ/, the voiced uvular /ʁ/, or the glottal stop /ʔ/.[1] Based on the preliminary results, subjects participating in this experiment were instructed to identify the consonant in the nonsense /Caa/ stimuli as one of three consonants: /ʕ/, /ʁ/, or /ʔ/.

The forced-choice paradigm provides us with the information concerning the cues for place of articulation. However, it does not give us information regarding the quality of the synthetic stimuli. Hence, we decided to use a subjective rating scheme along with the forced-choice paradigm. The subjects were asked to rate the speechlike quality or naturalness of the synthetic consonants on a 0 to 2 scale (0, 1 or 2).

Two one-hour sessions were held for this experiment on different days.

[1] A /ʔaa/ utterance is characterized acoustically by straight formant frequencies.

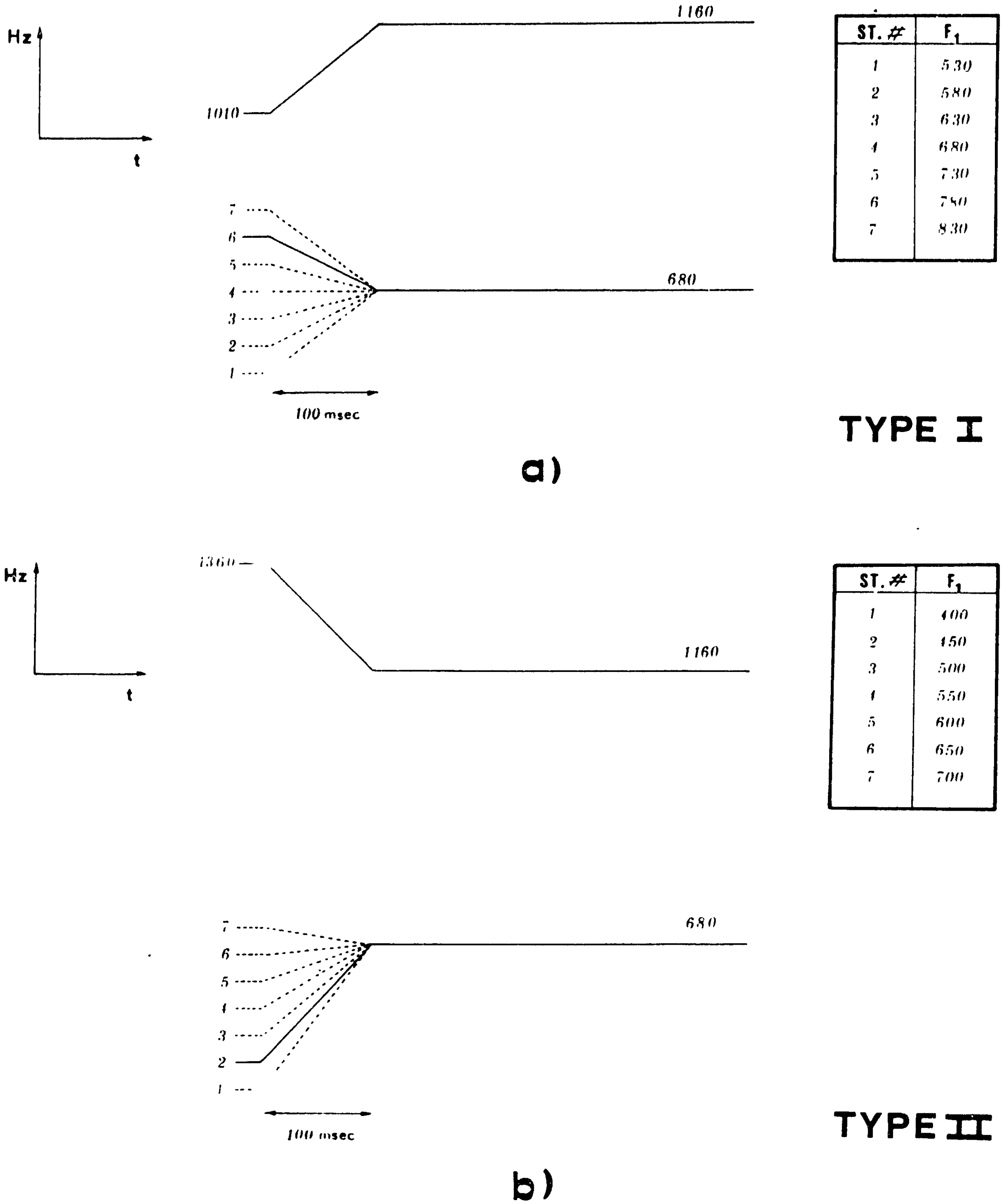


Figure 4.4: Schematized spectrograms of the first two formant frequencies used in stimulus 1 through 7 in the F1 continua for a) type I, and b) type II. Solid lines represent the formant trajectories for the "reference" stimuli.

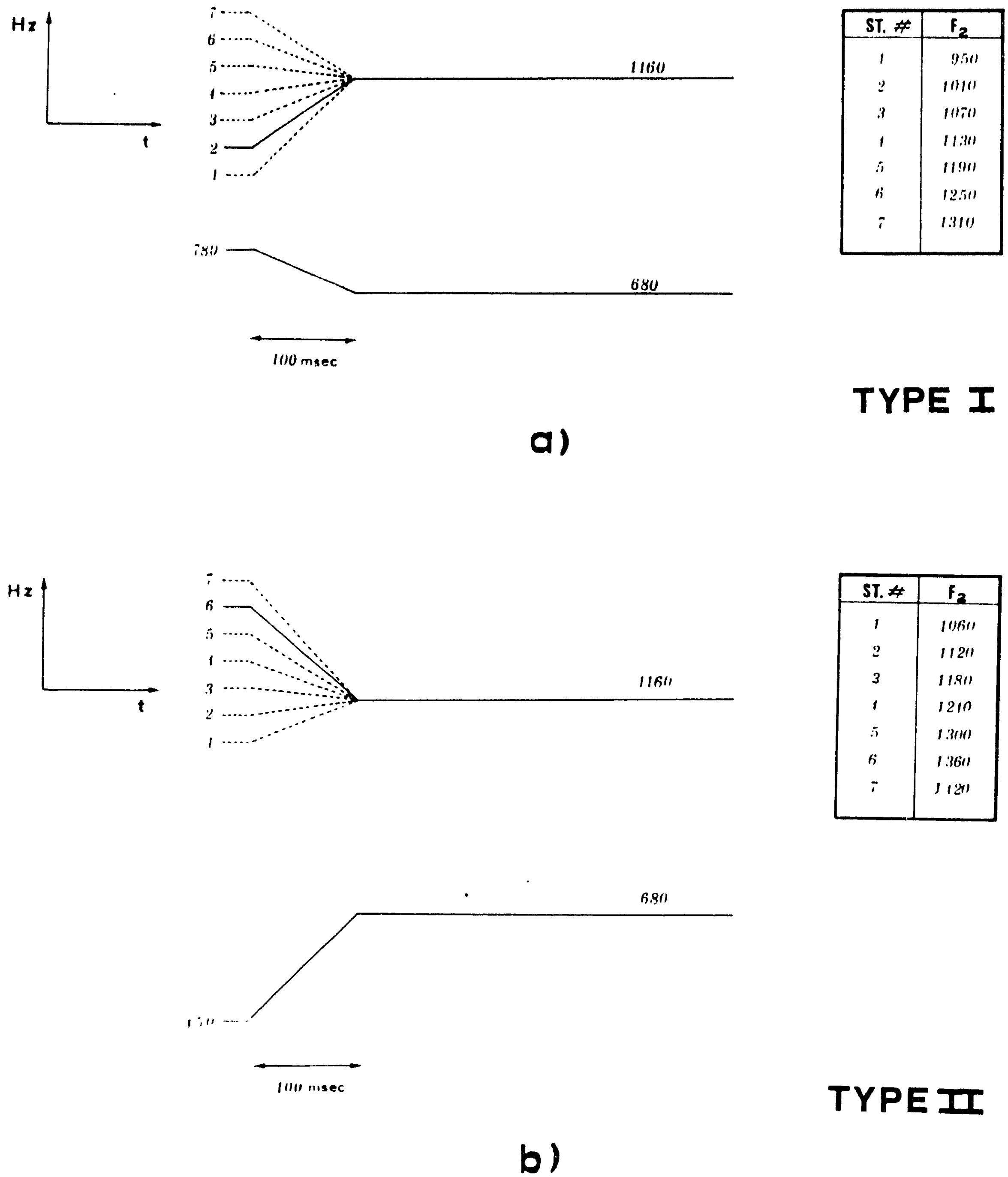


Figure 4.5: Schematized spectrograms of the first two formant frequencies used in stimulus 1 through 7 in the F2 continua for a) type I, and b) type II. Solid lines represent the formant trajectories for the "reference" stimuli.

4.4.2 Results and Discussion

F1 Continua

Identification functions for the F1 continua for types I and II are shown for each subject individually in Figures 4.6 and 4.7 (the name of each subject appears in the top left-hand corner of each plot), and the average values across the six subjects for both types are shown in Figure 4.8a and 4.8b. Also shown in Figure 4.8a and 4.8b are percentage values (averaged across all subjects) of the "naturalness" ratings of the consonant that was identified above chance and received the highest ratings. For example, in the F1 continuum for type I, the consonant in stimulus 7 was perceived as a pharyngeal with the highest rating (80%) among all other consonants perceived as pharyngeal above chance. Likewise, the consonant in stimulus 1, in the same continuum, was perceived as a uvular with highest rating (50%) among all other consonants perceived as uvular above chance. As explained earlier, the F1 continua were formed by varying the onset values of F1 keeping F2 onset value constant. Stimulus ss, formed by straight formant frequencies, was perceived as a glottal stop 98% of the time, and its speechlike quality was judged by the listeners to be 68%.

A similar trend is observed for all subjects and both continua, namely, that subjects labeled the consonants at the lower end of the continua as the uvular /ʁ/, at the higher end as the pharyngeal /ʕ/, and in between as the glottal stop /ʔ/.

The /ʁ/-/ʔ/ boundary occurred, on average, at 580 Hz for type I, and 550 Hz for type II. The /ʔ/-/ʕ/ boundary occurred at 684 Hz for type I and 680 Hz for type II. Although the onset values of F2 for both types differ by 350 Hz, the boundaries are strikingly similar.

These results suggest that, preceding the vowel /aa/, if F2 is held constant at a value appropriate for either the pharyngeal or the uvular, then F1 position is the essential cue for identifying the place of articulation. An onset value of F1 which is at least as high as that in the steady state of the vowel (in this case, 680 Hz), results in the perception of the pharyngeal, and its perception increases with increasing F1 onset value. The uvular, on the other hand, is perceived when the onset value of F1 is at least 130 Hz less than that during the steady state of the vowel (corresponding to 550 Hz for our stimuli), and its perception increases as the onset of F1 decreases in value. A glottal stop is perceived when the F1 onset value does not satisfy the criteria described above for identifying either the pharyngeal or the uvular.

The naturalness responses were correlated with the identification of the consonant, that is, the stimulus with the highest percentage identification was also perceived as being the most natural.

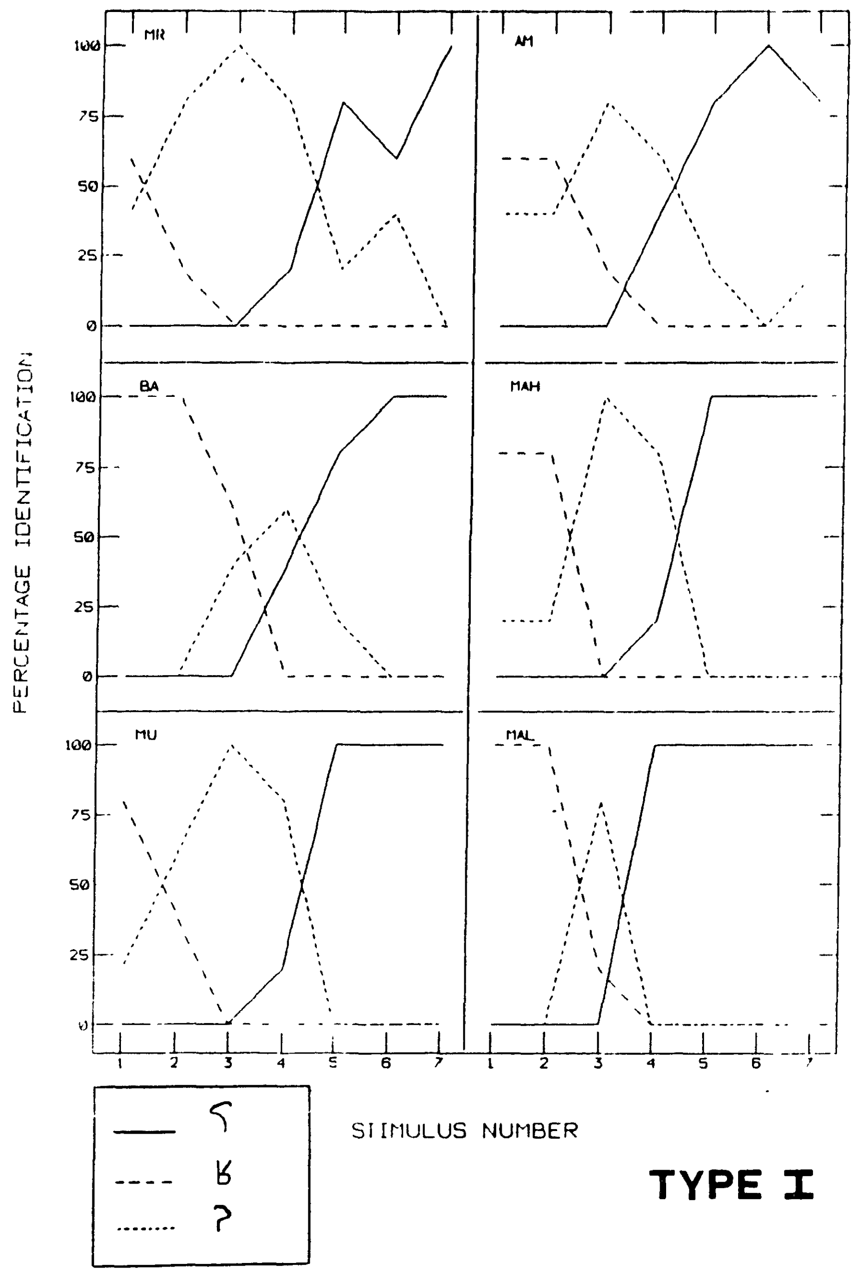


Figure 4.6: Plots of percentage identification versus stimulus number for the F1 continuum of type I. The plots are shown for each of the six subjects individually.

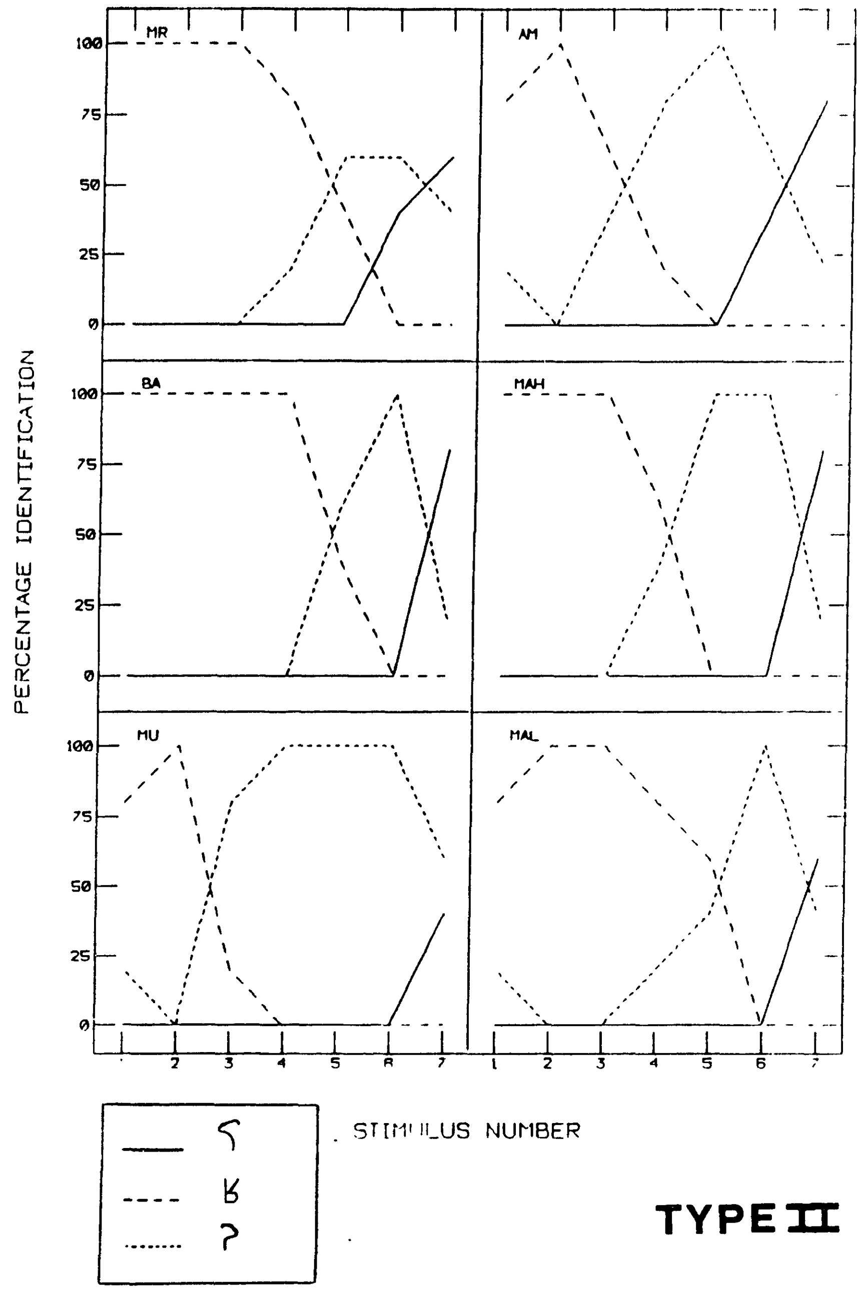


Figure 4.7: Plots of percentage identification versus stimulus number for the F1 continuum of type II. The plots are shown for each of the six subjects individually.

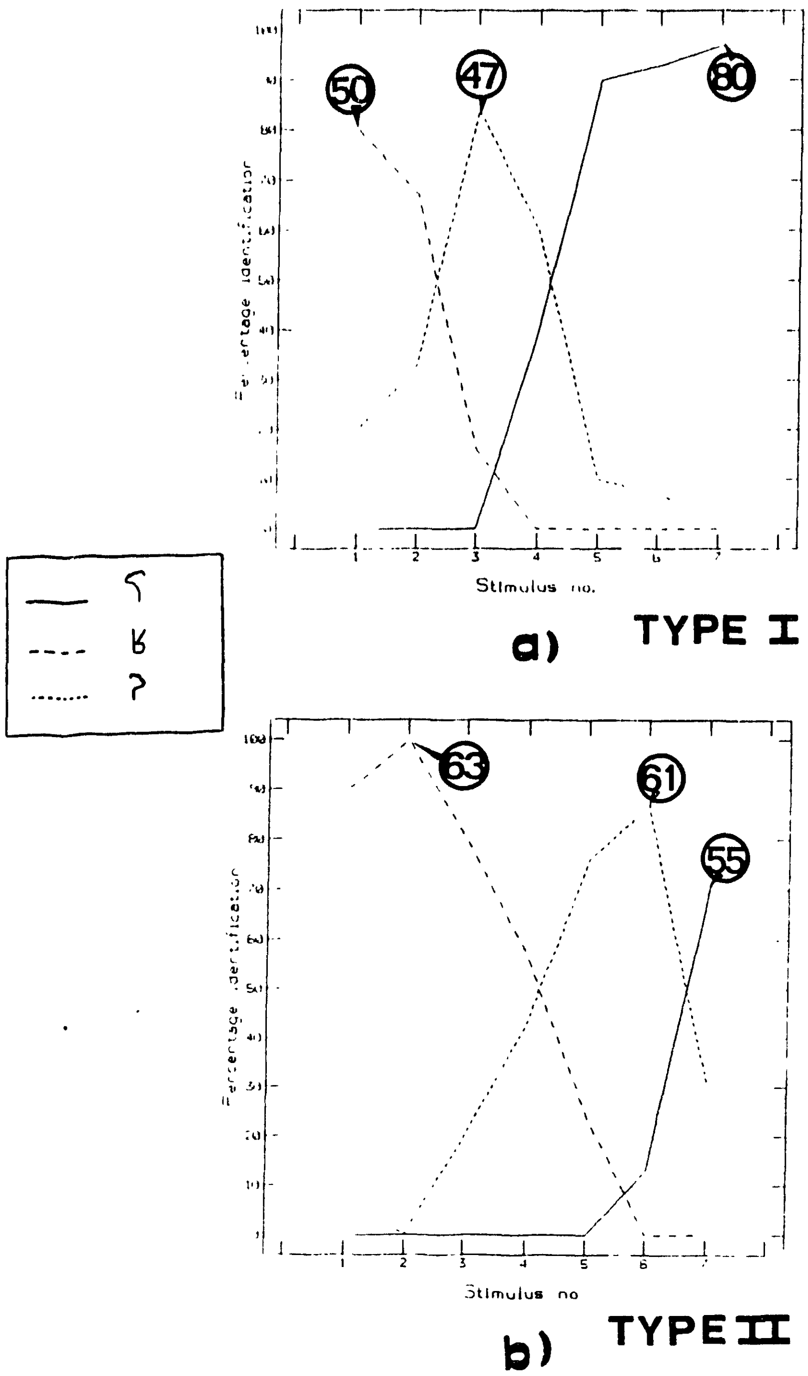


Figure 4.8: Plots of the average percentage identification versus stimulus number for the F1 continuum of a) type I, and b) type II. Encircled numbers represent percentage "naturalness" of the stimuli (see text for details).

F2 Continua

Identification functions for the F2 continua for both types are shown in Figure 4.9a and 4.9b. The functions reflect average values across all subjects, along with the rating for the most "natural" consonant perceived above chance. Plots of these functions for each subject individually are shown in the Appendix.

Unlike the results for the F1 continua, changing the onset values of F2 had insignificant effects on the identification of either consonant. All stimuli in type I and type II were perceived as pharyngeal and uvular, respectively, above chance. This could have been predicted from the results of the F1 continua, since subjects responded to both types of stimuli (I and II) in a similar fashion. However, the results provide further evidence to the hypothesis proposed earlier that F1 is the important cue for place for these consonants.

The "naturalness" responses show that the stimulus with the lowest F2 onset value (stimulus 1 in type I) received the the highest pharyngeal ratings, while an F2 onset value of 1300 Hz (stimulus 5 in type II) was perceived as the most "natural" uvular stimulus. It could be that listeners prefer a lower (F2−F1) value for the pharyngeal than they do for the uvulars, which means that they favor an enhancement of the feature *back* for the pharyngeal, since it is produced with a more posterior constriction location.

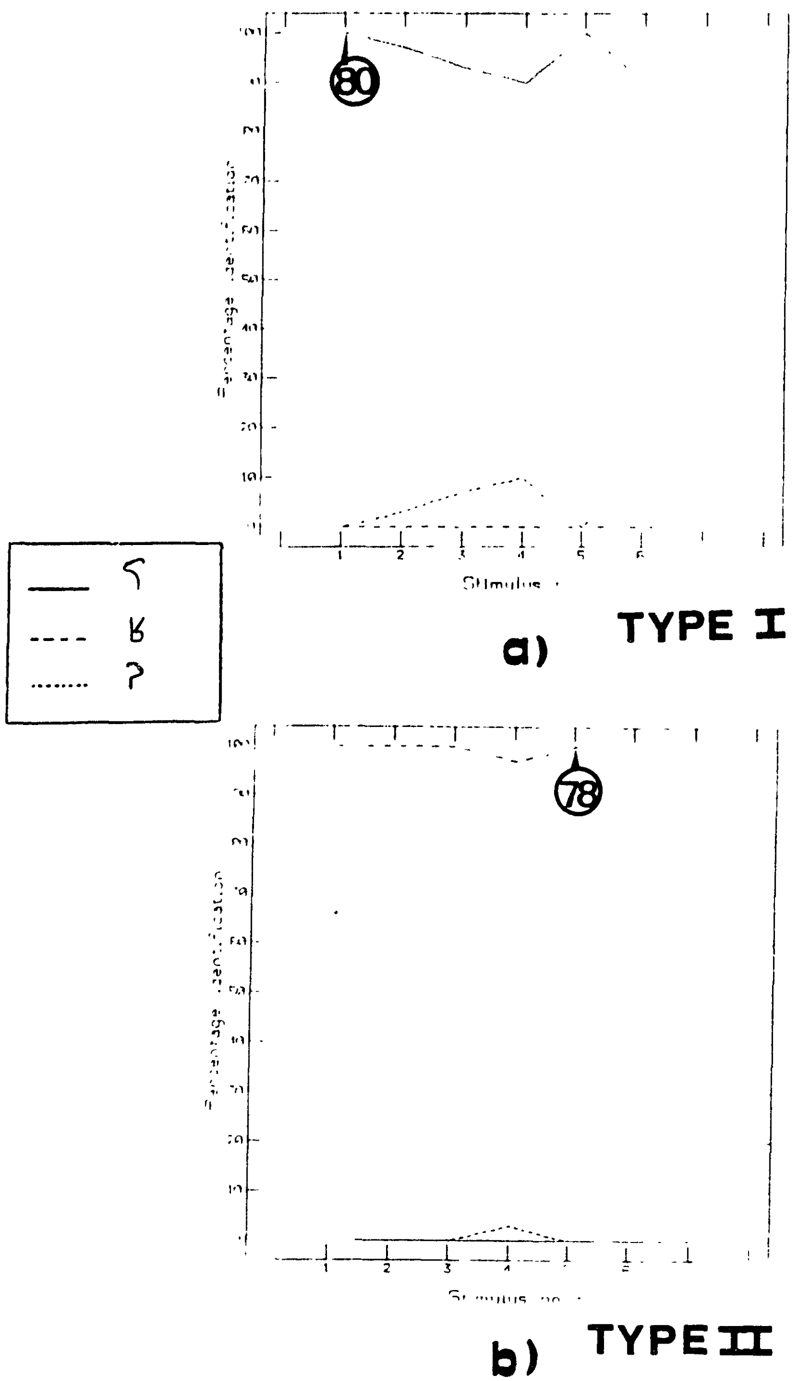


Figure 4.9: Plots of the average percentage identification versus stimulus number for the F2 continuum of a) type I, and b) type II. Encircled numbers represent percentage "naturalness" of the stimuli (see text for details).

4.5 Experiment II

In Chapter 2 we predicted that the bandwidths of the Helmholtz resonances (F2 for the pharyngeal and F1 for the uvular) would be wider, compared to a no-constriction-loss case, due to constriction losses. Results of the acoustic analysis in Chapter 3 showed evidence for this prediction for the uvular case, while for the pharyngeal such evidence was not clear.

The goal of the second experiment was to investigate whether a widened F1 bandwidth for the uvular and a widened F2 bandwidth for the pharyngeal would have any effect on the perception of these sounds.

4.5.1 Procedure

The same procedure and instructions as for Experiment I were used for this experiment. The stimuli for Experiment II had characteristics similar to those used in Experiment I except that the bandwidth of F2 for the F2 continuum of type I and of F1 for the F1 continuum of type II, were increased from their default values (90 and 60 Hz, respectively) to 200 Hz. In addition, the stimulus with the highest percentage identification as a pharyngeal in the F1 continuum of type I (stimulus 7), and stimulus ss (which was perceived as a glottal stop) were resynthesized with an increased F1 bandwidth (200 Hz).

The formant-bandwidth increase occurred over a time interval of 125 msec. and dropped linearly to its default value in 50 msec. A schematized plot of the formant bandwidth of F2, for type I, and of F1, for type II, is shown in Figure 4.10.

4.5.2 Results and Discussion

F1 Continuum

Identification functions for the test stimuli in the F1 continuum of type II are shown in Figure 4.11 for each subject individually, and averaged identification functions are shown in Figure 4.12. If we compare these results with those of Experiment I (Figures

4.7 and 4.8b) we notice several differences: (1) a greater cross-subject variability than that in Experiment I (while there is a clear /ʁ/-/ʔ/ boundary for subjects AM and MU, no such boundaries are observed for the other subjects), (2) greater uvular responses in Experiment II than in Experiment I (this is most noticeable for subjects BA and MAH, where the two perceived *only* the uvular for all the stimuli in Experiment II), (3) no pharyngeal responses were given for these stimuli (only for subject MR, where the pharyngeal was perceived below chance), and (4) the "naturalness" rating for the best perceived uvular increased from 63% in Experiment I to 93% in Experiment II.

As mentioned earlier, we had resynthesized the best perceived pharyngeal in Experiment I, and the glottal-stop stimulus (ss) with a widened F1 bandwidth, and presented them to the subjects as part of the stimuli in this experiment. The subjects identified the consonant in the pharyngeal stimulus as a uvular (67%). Similarly, the glottal-stop responses decreased from 98% in Experiment I to 36% in Experiment II. These results indicate that a widened F1 bandwidth decreases substantially the pharyngeal and glottal-stop responses, even if the formant transitions are appropriate for either sound.

F2 Continuum

The responses for type II continuum with an increased F2 bandwidth showed results (identification and naturalness) similar to those in Experiment I (Figure 4.9b). These results indicate that the subjects were insensitive to a change in F2 bandwidth, as they were to F2 location, when identifying the place of articulation for the voiced pharyngeal.

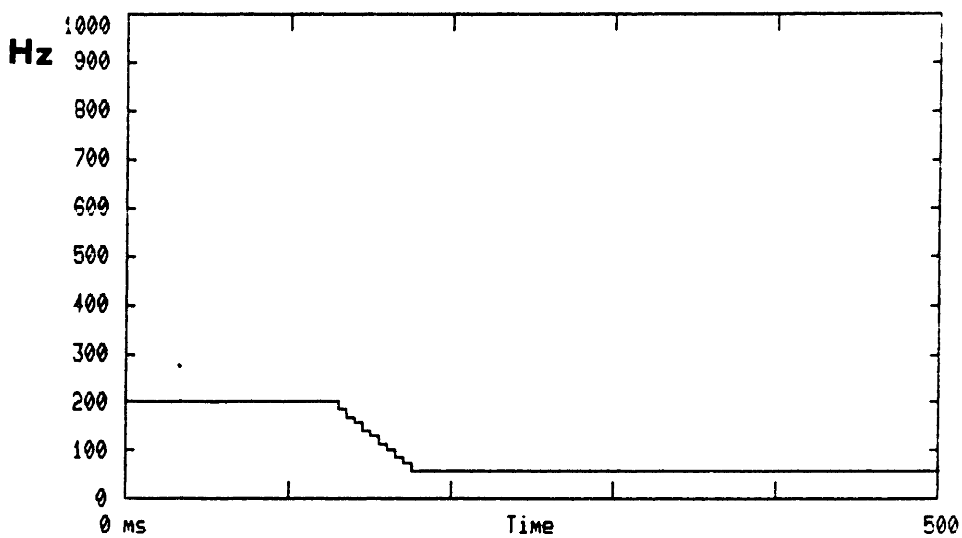


Figure 4.10: Schematic plot of the bandwidth of F2 for type I, and of F1 for type II as a function of time (Experiment II).

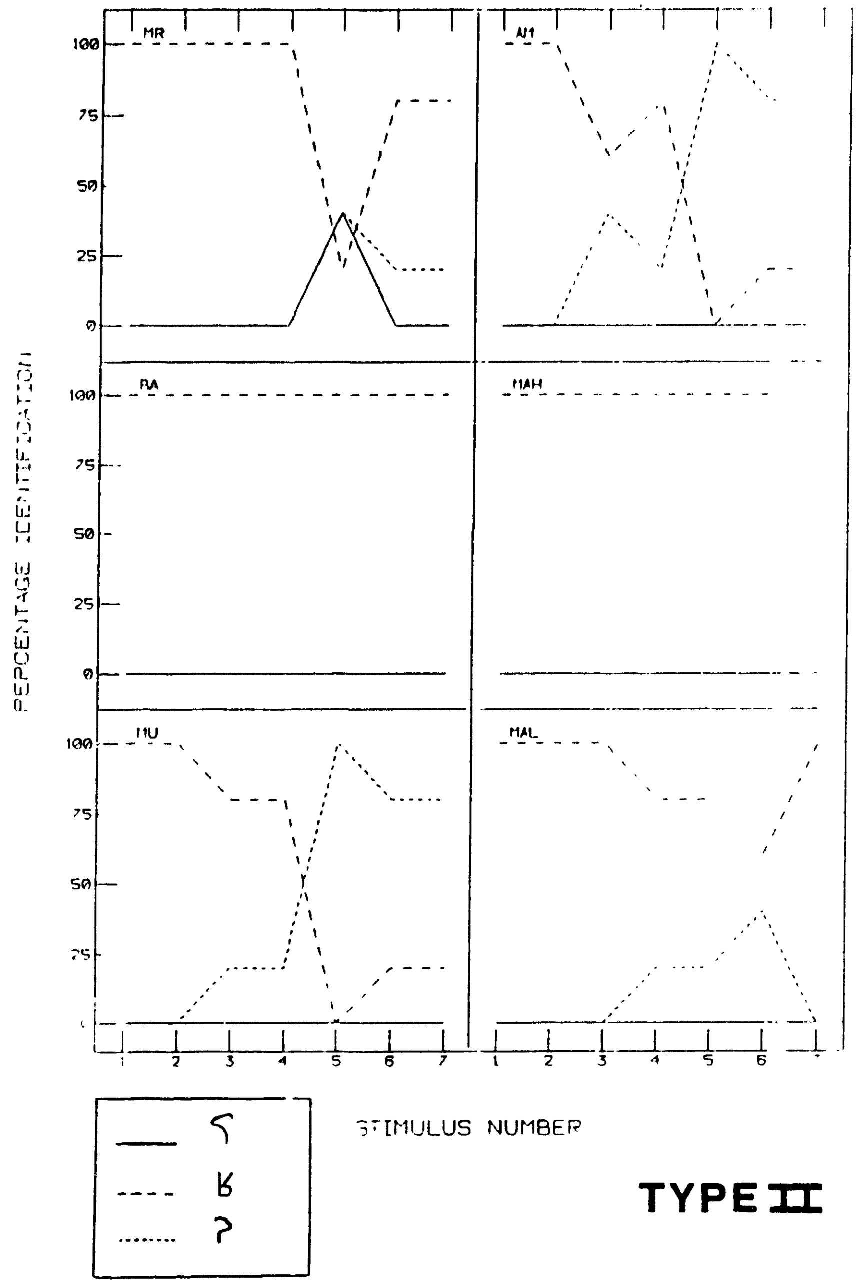


Figure 4.11: Plots of percentage identification versus stimulus number for the F1 continuum of type II with an increased F1 bandwidth (Experiment II). The plots are shown for each of the six subjects individually.

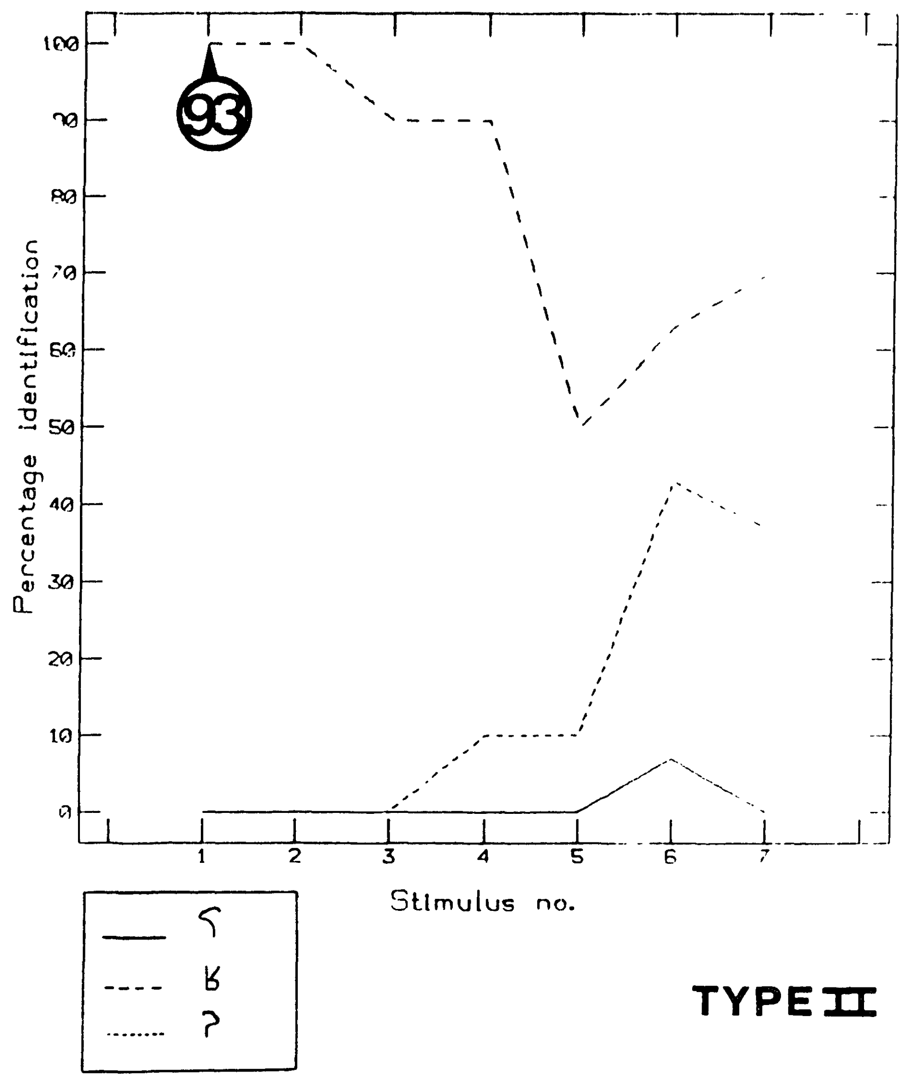


Figure 4.12: Plots of the average percentage identification versus stimulus number for the F1 continuum of type II with an increased bandwidth. Encircled numbers represent percentage "naturalness" of the stimuli (see text for details).

4.6 Summary

In this chapter we have investigated the perceptual cues essential for the identification of the voiced pharyngeal and uvular consonants in initial position preceding the vowel /aa/. Results of the first experiment show that the onset value of F1 ($F1_o$) is the essential cue in discriminating between the two sounds; the pharyngeal is identified when $F1_o$ is at least that during the steady state of the vowel ($F1_s$). The perception of the consonant increases as $F1_o$ becomes higher. The uvular is perceived when the difference ($F1_s - F1_o$) is at least 130 Hz, and its perception improves as F1 becomes lower in value. A glottal stop is perceived when ($F1_o$) is in between the values appropriate for the pharyngeal and the uvular.

We also found that F2 position and bandwidth have an insignificant effect on the perception of these sounds. This result is quite interesting, since several researchers have shown that one of the important cues used in identifying the place of articulation for English consonants is the F2 transition from the consonant to the vowel (Delattre, et al., 1955). Their results are valid for consonants with a low F1 (for example, labial and alveolar) where the F1 transition is approximately the same for all consonants considered, and the F2 transition provides the main cue for place. In contrast, pharyngeal consonants have a higher F1 than do uvulars, and this difference seems to be quite essential in discriminating between the two, at least with the vowel /aa/.

Results of the second experiment show that a widened F1 bandwidth results in higher responses for the uvular consonant. A widened F1 bandwidth improves the naturalness of the uvular stimuli (from 63% in the first experiment to 93% in the second). While a widened F1 bandwidth does not seem to be the essential cue in perceiving the uvular /ʁ/, it improved a great deal the speechlike quality of the uvular stimuli. For the pharyngeal, on the other hand, a widened F1 bandwidth degraded the perception of this consonant substantially.

Chapter 5

Conclusion

5.1 Summary and Discussion

This thesis has investigated the production mechanisms and the acoustic and perceptual correlates of pharyngeal and uvular consonants. In the first part of the study, we proposed simplified models of the vocal-tract area function during the production of these consonants. The models were appropriate for the consonants preceding a low vowel (for example, /aa/). From these models, we calculated the formant frequencies and determined the formant-cavity affiliations, we examined the acoustic effects of introducing a noise source in the vicinity of a supraglottal constriction, and we calculated the contributions of the different vocal-tract losses to the bandwidths of the formants. In addition, we examined the effects of the impedances of the glottal and the supraglottal constrictions on the formant-frequency locations and bandwidths. Tapering the inlets of the supraglottal constriction resulted in formant frequencies comparable to those measured by Ghazeli (1977) and Al-Ani (1970) from natural utterances, suggesting that these simplified models are reasonable models for the pharyngeals and the uvulars.

From these calculations, we predicted that, for the pharyngeals, F2 should be a Helmholtz resonance, and F1 and F3, front-cavity resonances, and for the uvulars, F1 should be a Helmholtz resonance, F2 and F4, front-cavity resonances, and F3, a back-cavity resonance. In terms of relative values of the formant frequencies, we predicted that F1 for pharyngeals should be higher than that for uvulars, and F3 should be lower.

F2 could not be distinguished for the two classes of sounds for the closed-glottis case. For the open-glottis case, however, the reactive part of the glottal impedance causes F2 for pharyngeals (a Helmholtz resonance) to be shifted upward, resulting in a higher value of F2 than that for uvulars.

The presence of a noise source, modeled as a series pressure source, near the supraglottal constriction introduced zeros to the transfer function from the volume velocity at the lips to the pressure source near the constriction (U_o/p_s). These zeros were in the vicinity of the back-cavity resonances, including the Helmholtz resonance. The zero-frequency location which was in the vicinity of the Helmholtz resonance was found to be highly sensitive to the pressure-source location, which led to the conclusion that the Helmholtz resonance for these consonants may or may not be cancelled by this zero. Other back-cavity resonances, on the other hand, were cancelled by zeros regardless of the pressure-source location. Accordingly, we predicted that when a noise source is present in the vicinity of the supraglottal constriction, the front-cavity resonances should be strongly excited, and the Helmholtz resonance may or may not be excited, depending on the noise-source location.

For the closed-glottis case, the bandwidth contribution of the constriction losses to the bandwidth of the Helmholtz resonance was found to be significant. For the open-glottis case, glottal losses accounted for the damping of the Helmholtz resonance.

Results of the acoustic analysis of two pharyngeal consonants /ʕ,ħ/ and three uvular consonants /ʁ,χ,q/ prevocalically with /aa,ii,uu/ in Arabic were in general agreement with the theoretical predictions. One of the important findings of the acoustic analysis was that the spectral shape of the voiceless uvular fricative and stop /χ,q/ was *compact.* The spectra for these consonants had a narrow peak at frequencies corresponding to the second formant, and the peak (formant) location varied depending on the following vowel. The spectral shape of the voiceless pharyngeal /ħ/ was *grave* with broad peaks at the second- and third-formant frequencies. The noise interval for both fricatives /ħ,χ/ had a clear formant structure. For the voiced consonants /ʕ,ʁ/ two allophones were found: a continuant and a non-continuant allophone. We concluded that the feature *continuant* is not distinctive for the voiced consonants. The "targets" of the formant frequencies were found to be different depending on the context; the F1 target was influenced by the height of the following vowel, whereas the F2 and F3 targets were influenced by the backness and rounding of the following vowel. These targets were higher for the voiceless consonants than they were for the voiced. The higher targets for the voiceless consonants are partly due to the upward shift in formant frequencies caused by the reactive part of the glottal impedance. This shift was predicted in the theoretical study. Coupling to the subglottal system could be another factor contributing to the increase in the formant-frequency locations. This factor, however, was not predicted theoretically since the subglottal system was not represented in our models.

As is the case with consonants in other languages, the voiceless consonants /χ,q,ħ/ were longer than their voiced counterparts /ʁ,ʕ/. We also found that duration does not contrast the two classes. The fundamental frequency (f0) for the voiced consonants /ʕ,ʁ/ was lower than that of the adjacent vowel, suggesting a source-filter interaction during the production of these sounds which affects the vibrational pattern of the vocal folds.

In the last part of the study, we conducted perceptual experiments, using synthetic /Caa/ stimuli, and found that the onset value of F1 is essential in discriminating between the voiced pharyngeal /ʕ/ and uvular /ʁ/. A falling trajectory of F1 resulted in the perception of the pharyngeal consonant, while a rising trajectory of F1 with an onset value at least 130 Hz less than that at the steady state of the vowel was perceived as the uvular. Onset values of F1 which did not satisfy the criteria for identifying either consonant resulted in the perception of the glottal stop /ʔ/. Widening the bandwidth of F1 increased the /ʁ/ responses, and improved the naturalness of the synthetic stimuli perceived as /ʁ/, whereas it decreased the /ʕ/ responses substantially. F2 position and bandwidth had no effect on the perception of either consonant. The increased bandwidths were predicted from the theoretical study.

The results of this thesis are in agreement with the phonetic description of pharyngeal and uvular consonants given by Chomsky and Halle (1968). That is, both classes are (*−anterior*, *−coronal*, *−high*, *+back*) with the *low* feature distinguishing the two classes; pharyngeals have the feature (*+low*) and uvulars, (*−low*). However, this description fails to describe the spectral shape specific for the voiceless consonants. One suggestion would be to assign the features *compact* and *grave* to the voiceless uvular consonants, and *compact* and *−grave* to the voiceless pharyngeal. We are using the feature *compact* to describe well-defined peaks in the spectra and the feature *grave* to indicate that the locations of these peaks are in the mid-frequency region. Table 5.1 summarizes the feature matrix proposed for pharyngeal and uvular consonants. We notice in the table that the *continuant* feature is not distinctive for the voiced consonants; this is concluded from the results of the acoustic analysis.

At this point we cannot comment on Jakobson's (Jakobson, 1957) description of the pharyngeals as being *−consonantal* and we suggest that further studies of Arabic phonology would help in clarifying this issue.

Feature Matrix					
	Pharyngeal		Uvular		
	ʕ	ħ	ʁ	χ	q
voiced	+	−	+	−	−
low	+	+	−	−	−
high	−	−	−	−	−
back	+	+	+	+	+
continuant	+/−	+	+/−	+	−
compact		+		+	+
grave		−		+	+

Table 5.1: A suggested feature matrix for the pharyngeals and the uvulars.

5.2 Future Work

The theoretical models presented in this study are appropriate for the pharyngeals and the uvulars preceding the vowel /aa/. Results of the acoustic analysis showed that the formant targets for these sounds are highly influenced by the context. Hence, more complex models of the vocal-tract area function which take into account perturbations, in anticipation of the following sound, need to be developed. These models should also account for coupling to the subglottal system, when the glottis is open, a feature not present in our models. Development of these models requires more reliable articulatory data. For example, X-ray data of the consonants in different contexts would help in estimating the area function of the vocal tract for each case, measurements of the subglottal pressure and the volume-velocity flow would facilitate the estimation of the cross-sectional area of the constriction, and electroglottograph (EGG) data of the voiced consonants would aid in understanding the source-filter interaction occurring during the production of these consonants.

In our models we represented the noise source at the constriction by a localized pressure source. It would be of interest to examine the acoustic effects of modeling the noise source as a distributed pressure source, rather than localized.

Quantifying the acoustic properties of these sounds in a wide range of contexts and in continuous speech would assist in developing algorithms for practical applications, such as speech recognition and synthesis.

A natural extension to the perceptual study would be identifying the perceptual cues for place of articulation for the voiced consonants with the vowels /ii,uu/, and for the voiceless consonants in different vowel contexts.

Appendix A

Tables and Figures

Formant Frequencies (Hz)					
	ʕ	ħ	ʁ	χ	q
C/aa/					
F1	744	759	496	604	604
F2	1116	1162	1333	1317	1023
F3	2247	2278	2464	2480	2681
C/ii/					
F1	527	449	356	325	325
F2	1798	1984	1798	1891	2061
F3	2449	2573	2495	2573	2542
C/uu/					
F1	558	480	372	434	372
F2	914	1007	821	899	775
F3	2015	2139	2464	2542	2639

Table A.1: Average values of the first three formants measured at the onsets of the vowels preceded by one of the five consonants. The average values were based on measurements from two tokens for speaker JM.

Formant Frequencies (Hz)					
	ʕ	ħ	ʁ	χ	q
C/aa/					
F1	821	790	588	604	604
F2	1162	1379	1317	1317	1282
F3	2170	2387	2371	2387	2265
C/ii/					
F1	496	403	387	391	422
F2	1829	2061	1627	2077	1906
F3	2402	2573	2495	2495	2655
C/uu/					
F1	387	434	403	372	386
F2	1085	1209	700	682	758
F3	1937	1984	2449	2263	2443

Table A.2: Average values of the first three formants measured at the onsets of the vowels preceded by one of the five consonants. The average values were based on measurements from two tokens for speaker MU.

Formant Frequencies (Hz)					
	ʕ	ħ	ʁ	χ	q
C/aa/					
F1	682	713	511	542	561
F2	1240	1302	1069	1271	1118
F3	2247	2387	2681	2883	2647
C/ii/					
F1	449	480	403	391	422
F2	1891	1798	1472	2015	1628
F3	2728	2635	2852	2774	2635
C/uu/					
F1	480	449	387	403	424
F2	1193	1302	635	806	758
F3	2123	2123	2821	2681	2780

Table A.3: Average values of the first three formants measured at the onsets of the vowels preceded by one of the five consonants. The average values were based on measurements from two tokens for speaker HA.

Formant Frequencies (Hz)					
	ʕ	ħ	ʁ	χ	q
C/aa/					
F1	666	697	573	589	580
F2	1209	1224	1224	1271	1149
F3	2232	2232	2526	2464	2498
C/ii/					
F1	497	480	403	424	467
F2	2030	1906	1844	1968	1543
F3	2666	2588	2588	2573	2467
C/uu/					
F1	434	449	387	403	422
F2	1069	1209	682	775	837
F3	2015	2092	2588	2402	2387

Table A.4: Average values of the first three formants measured at the onsets of the vowels preceding by one of the five consonants. The average values were based on measurements from two tokens for speaker LT.

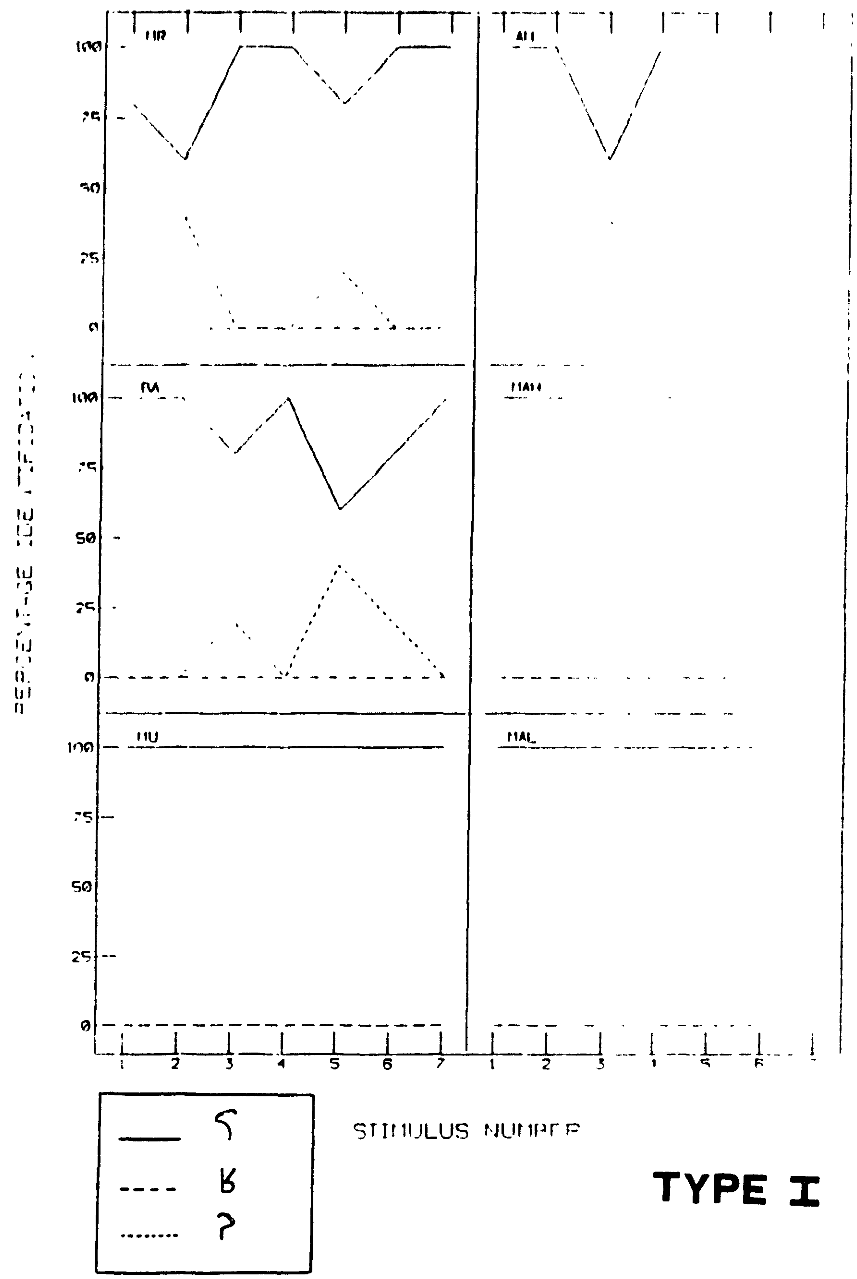


Figure A.1: Plots of percentage identification versus stimulus number for the F2 continuum of type I. The plots are shown for each of the six subjects individually.

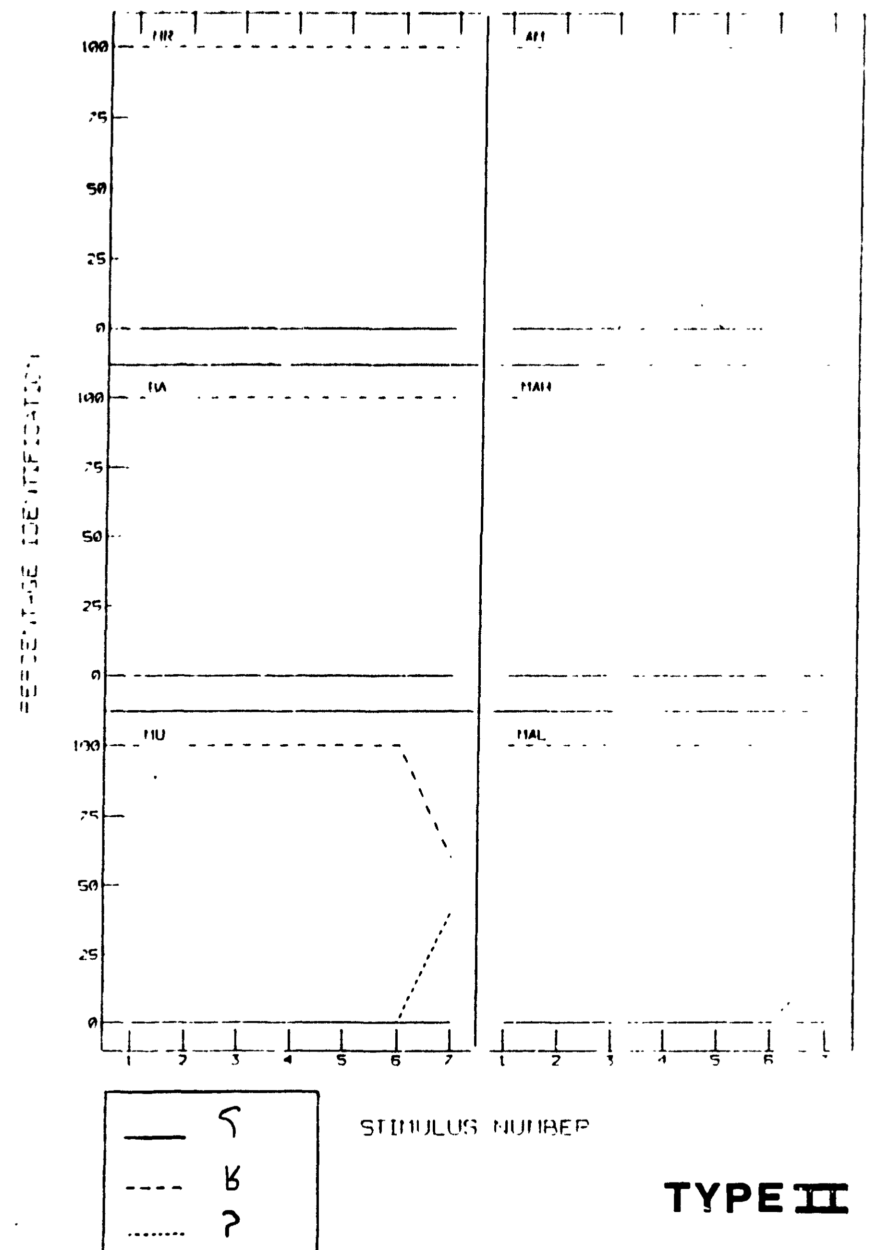


Figure A.2: Plots of percentage identification versus stimulus number for the F2 continuum of type II. The plots are shown for each of the six subjects individually.

Bibliography

[1] Adamson, B. (1981). "A spectrographic analysis of pharyngeal consonants in Sudanese Arabic", *Work in Progress no. 3*, Phonetics Laboratory, Univ. of Reading, 81-96.

[2] AL-Ani, S. (1970). *Arabic Phonology.* Mouton, the Hague.

[3] AL-Ani, S. (1985). Personal communication.

[4] Avicenna (1333/1916). *Makharij al-Huruf.* University of Tehran Press, Tehran.

[5] Bickley, C. and K.N. Stevens (in press). "Effects of a vocal-tract constriction on the glottal source: Data from voiced consonants", to appear in K.S. Harris, T. Baer, and C. Sasaki (Eds.), *Vocal Fold Physiology: Laryngeal Function in Phonation and Respiration,* San Diego: College-Hill Press.

[6] Catford, J.C. (1968). "The articulatory possibilities of man", in *Manual of Phonetics*, Amsterdam:North Holland publishing Company, 309-334.

[7] Chomsky, N., and M. Halle (1968). *The Sound Pattern of English.* Harber & Row Publishers, New York, 305-307.

[8] Delattre, P. (1971). "Pharyngeal features in the consonants of Arabic, German, French, Spanish and American English ", *Phonetics* 23, 129-155.

[9] Delattre, P.C., Liberman, A.M., and Franklin S. Cooper (1955). "Acoustic Loci and Transitional Cues for Consonants". *J. Acoust. Soc. Am.* 27 no. 4, 769-773.

[10] Fant, G. (1960). *Acoustic Theory of Speech Production.* Mouton, The Hague.

[11] Gardiner, W.H.T. (1925). *Phonetics of Arabic.* Oxford University Press, London.

[12] Ghazeli, S. (1977). "Back consonants and backing coarticulation in Arabic", unpublished Ph.D. thesis, Univ. of Texas at Austin.

[13] Heinz, J.M., and K.N. Stevens (1961). "On the properties of voiceless fricative consonants", *J. Acoust. Soc. Am.* 33 no. 5, 589-596.

[14] Henke, W. (1966). "Dynamic Articulatory Model of Speech Production Using Computer Simulation", unpublished Ph.D. thesis, M.I.T..

[15] Hosein, P. (1983). "Computer Simulation of a Model for the Vocal Tract", unpublished B.S. thesis, M.I.T..

[16] Jakobson, R. (1957). "Mufaxxama, The 'Emphatic' Phonemes in Arabic ", in *Studies presented to Joshua Whatmough on his 60th Birthday*, Mouton, The Hague, 105-115.

[17] Jakobson, R., Fant, C.G.M. and M. Halle (1963). *Preliminaries to Speech Analysis.* M.I.T. Press, Cambridge, Mass.

[18] Klatt D.H. and Stevens K.N. (1969)."Pharyngeal consonants ", *QPR* no. 93, RLE, M.I.T., 207-215.

[19] Klatt D.H. (1980). "Software for a cascade/parallel formant synthesizer ", *J. Acoust. Soc. Am.* 67 no. 3, 971-995.

[20] Klatt, D.H. (1983). *M.I.T. Speechvax User's Guide.*

[21] Ladefoged, P. (1975). *A course in phonetics.* HBJ Inc., New York, 143.

[22] Lisker, L. and A. Abramson (1964). "A cross-language study of voicing in initial stops: Acoustical Measurements", *Word 20.*

[23] Maddieson, I. (1984). *Patterns of Sounds.* Cambridge University Press.

[24] Meyer-Eppler, W. (1953). "Zum Erzeugungsmechanisms der Gerauschlaute", *Z. Phontik Vol.7, issue 3/4*

[25] Obrecht, D.H. (1968). *Effects of the Second Formant on the Perception of Velarization Consonants in Arabic.* Mouton, The Hague.

[26] Shadle, C.H. (1985). "The Acoustics of Fricative Consonants", M.I.T. Ph.D. thesis, published as Research Laboratory of Electronics Technical Report no. 506.

[27] Sibawayhi (750/1975). *al-Kitab.* Aalam-Alkutub, Beirut.

[28] Stevens, K.N. (in press). *Acoustic Phonetics.*

[29] Umeda, N. (1977). "Consonant Duration in English", *J. Acoust. Soc. Am.* 61 no. 3, 846-858.