A Phonetic Model of the Human Pharynx

by

Lynn Marie Heap
B.Sc., University of Victoria, 1994

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF ARTS

in the Department of Linguistics

University of Victoria

0-612-36614-6

Supervisor: Dr. J. H. Esling

Abstract

Esling (1996) has proposed that the range of speech sounds that are articulated behind the velum can be organized according to one place of articulation with four possible manners of articulation. This thesis provides acoustic and video-endoscopic evidence for the re-classification of post-velar sounds by describing pharyngeal speech sounds with criteria used in existing phonetic taxonomies and theories of linguistics. According to the International Phonetic Alphabet (IPA) taxonomy, there are sounds that have been identified as pharyngeal as well as sounds classified as epiglottal. Esling has proposed that the epiglottal sounds can be re-classified as varying manners of articulation that are produced in the pharynx.

This thesis uses acoustic and video-endoscopic evidence to show that there is one place in the pharynx that has major constriction capabilities, and that the larynx as a whole is able to move independently beneath the sphincter that is created by the aryepiglottic folds and the epiglottis. It is shown that the larynx can rise and lower vertically beneath this pharyngeal constriction, which contributes to the auditory confusion of post-velar sounds. Acoustic features that are associated with raised larynx and pharyngealized voices in Laver's taxonomy of Voice Quality Settings are used to provide support for larynx height as a distinctive feature in pharyngeal sounds.

Along with the correlation of acoustical information with linguistic taxonomies, other acoustical correlates including pitch, shimmer, jitter and harmonic-to-noise ratio are described for cardinal pharyngeal sounds as well as for natural language data that have pharyngeal sounds. Based on the description of acoustical effects that identify pharyngeal sounds there needs to be a modification of the traditional acoustical model of speech production.

The existing model of speech production, the source-filter theory, requires modification, since pharyngeal constriction is shown to have an effect on the source aspect

of the model by decreasing vocal fold vibration. The ability of the aryepiglottic folds to vibrate simultaneously is shown to provide another source of sound that can occur in conjunction with vocal fold vibration. The filtering acoustic effects of pharyngeal constriction and changes in larynx height have only been generally described in the past. The closer acoustical examination of cardinal pharyngeal sounds in this thesis suggests that it is energy in the upper frequency range that distinguishes between changes in the back part of the vocal tract.

Also developed in this thesis is a technique for measuring video images. By normalizing video images to a common reference in the vocal tract, it is possible to compare pictures between subjects and recording sessions. The results of applying the video measurement technique to images of cardinal pharyngeal sounds suggest that this technique will be more successful in describing voice quality settings when applied to a sequence of video pictures.

This thesis, in describing pharyngeal speech segments, has provided a range of acoustical measurements that are found to be salient in describing pharyngeal sounds and manners of articulations. The result of this broad acoustical analysis has shown that the present source-filter theory of speech production needs to be modified in order to accommodate the observed constriction capabilities in the pharynx and vertical movements of the larynx. The video measurement technique developed in this thesis allows video pictures to be compared so that quantitative details between subjects and recording sessions can be stated.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Thank you to my supervisor, John Esling who has encouraged me to do a good job even though I felt I couldn't do anymore, and to the rest of my committee, Ewa Czaykowska-Higgins and Andy Schloss, for comments and contributions along the way. I am grateful to Peter Dreissen for the encouragement to apply what I have done to the 'real' world.

Acknowledgements are also due to Craig Dickson and Roy Snell at Speech Technology Research Ltd. for help with the technical bits, and to the BC Science Council for granting us the GREAT award.

I am forever indebted to my family for their patience and help even though it is hard to understand what I am doing and the effort that is involved. Thank you Veronica Garner, Lily Heap, Dave, Kyloe and Rowen. I owe my sanity to Jocelyn Clayards who is keen and interested and has provided plenty of insightful discussions and been very generous with her time and the machines!

Chapter 1 - Introduction

1.0 Introduction

The goal of this thesis is to learn as much as possible about what happens in the pharynx when we talk so that speech can be more accurately modeled. This goal is accomplished by using linguistic phonetic taxonomies as a guide to organize the movements observed in the pharynx. On the basis of auditory analysis, it is proposed that the Pharyngeal column on the International Phonetic Alphabet (IPA) chart (see page 4) should be modified to include four manners of articulation. The hypothesis proposed in this thesis is that there is one place in the pharynx where it is optimally possible to produce sounds that correspond with the manners of articulation on the IPA chart. This hypothesis is motivated by Esling (1996), who proposes this place and manner of articulation organization of pharyngeal sounds based on historical and auditory evidence. This thesis provides support for the IPA interpretation of pharyngeal sounds by laying out a model of pharyngeal behaviour based on acoustic and video-endoscopic evidence. Also, this thesis provides acoustic evidence that larynx height serves as a distinctive phonetic feature used to distinguish speech sounds.

A distinctive feature is defined as part of a bundle of features that identifies a speech segment. A system of distinctive features can be defined by articulatory, acoustic or perceptual features (Trask, 1996). This thesis focuses on the acoustic features of pharyngeal stops, fricatives, trills and approximants; however, in discussing acoustic features, an outline of articulatory and auditory features is also developed.

A distinctive feature description of segments is a valuable method of investigating speech sounds because features tend to interact in ways that form a natural class of sounds within and among languages. A natural class of sounds 'appeals' to particular characteristics of segments. For example, in English, voiceless segments constitute a natural class that appeals to clustering with the word initial fricative [s]. This appeal gives words such as 'spin' rather than *sbin or *zpin. The sounds examined in this thesis are treated as a natural class of sounds that are produced behind the velum but otherwise have a rather vague interpretation.

McCarthy (1994) provides an outline of auditory and articulatory properties of pharyngeal sounds based on a phonological analysis of various languages. McCarthy notes that, in general, pharyngeal sounds tend to affect the vowels that they neighbour and pharyngeal sounds have co-occurrence restrictions that make them avoid syllable final position. Beyond noting these phonological characteristics of pharyngeal sounds, this thesis does not discuss the phonological arguments that McCarthy makes. The description in this thesis will provide a more precise phonetic description of the types of movements that take place in the pharynx.

The results of this research will describe the quantal nature of sounds that are made by movements in the pharynx. The term quantal was proposed by Stevens (1972) and accounts for the observation that languages do not distinguish between all possible speech articulations. The quantal theory suggests that in speech there are regions of acoustic stability that correlate with salient, or distinctive features. Speech output then, is mapped in a nonlinear way with the articulatory motion that produces speech; this makes it possible to produce speech segments rather sloppily yet the sound can still be perceived as the target sound.

Quantal theory implies that certain aspects of human speech, such as voicing, for example, offer a much wider range of acoustic variation than is used by humans for the purpose of conveying language. The example that Johnson (1997) cites illustrates how a speech sound uses only one of three different glottal widths; either voiceless (wide), voiced (not so wide) or glottal stop (closed). There is certainly a range of widths in between, but the perception of them falls into the category of phonation type and contributes more to voice quality rather than differentiating sound segments. With respect to post-velar sounds, the perceptually defined categories outlined in the IPA chart imply two places of articulation in the pharynx namely, pharyngeal and epiglottal. This thesis provides evidence which describes acoustically stable regions that suggest one place of articulation and four manners of articulation. This thesis shows that the term epiglottal is redundant since there is only one possible place of articulation in the pharynx. Also, larynx height is shown to be a distinctive feature that can be applied to pharyngeal sounds and may contribute to the auditory confusion between post-velar sounds.

The International Phonetic Alphabet (IPA) is the most widely used and recognized phonetic alphabet which provides symbols for speech sounds. The IPA chart provides a taxonomy of speech sounds that is arranged according to the 'place' in the vocal tract that the sound occurs and the 'manner' in which it is articulated. That is, the leftmost column in the IPA chart corresponds to sounds produced at the lips and the rightmost corresponds to the glottis, where voice is initiated by the vocal folds. The rows correlate with the amount of constriction required to produce the manner of articulation, for example, the most constricted consonant is a plosive, which blocks the airstream completely. Page 4 contains a modified version of the IPA chart. The area of discussion for this thesis is the pharyngeal column and the section below the chart called 'other symbols'.

Along with organizing sounds according to their articulatory features, the IPA chart classifies sounds dynamically with respect to frequency and time domains. Features of a sound that have temporal characteristics describe manner of articulation and features that

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993, updated 1996)

CONSONANTS (PULMONIC)

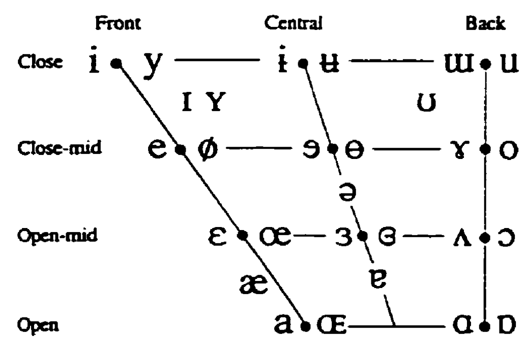
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks		Voiced implosives		Ejectives	
ʘ	Bilabial	ɓ	Bilabial	ʼ	Examples:
ǀ	Dental	ɗ	Dental/alveolar	pʼ	Bilabial
ǃ	(Post)alveolar	ʄ	Palatal	tʼ	Dental/alveolar
ǂ	Palatoalveolar	ɠ	Velar	kʼ	Velar
ǁ	Alveolar lateral	ʛ	Uvular	sʼ	Alveolar fricative

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

OTHER SYMBOLS

ʍ Voiceless labial-velar fricative

w Voiced labial-velar approximant

ɥ Voiced labial-palatal approximant

ʜ Voiceless epiglottal fricative

ʢ Voiced epiglottal fricative

ʡ Epiglottal plosive

ɕ ʑ Alveolo-palatal fricatives

ɺ Alveolar lateral flap

ɧ Simultaneous ʃ and x

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary. k͡p t͡s

DIACRITICS Diacritics may be placed above a symbol with a descender, e.g. ŋ̊

̥ Voiceless	n̥ d̥	̤ Breathy voiced	b̤ a̤	̪ Dental	t̪ d̪
̬ Voiced	s̬ t̬	̰ Creaky voiced	b̰ a̰	̺ Apical	t̺ d̺
ʰ Aspirated	tʰ dʰ	̼ Linguolabial	t̼ d̼	̻ Laminal	t̻ d̻
̹ More rounded	ɔ̹	ʷ Labialized	tʷ dʷ	̃ Nasalized	ẽ
̜ Less rounded	ɔ̜	ʲ Palatalized	tʲ dʲ	ⁿ Nasal release	dⁿ
̟ Advanced	u̟	ˠ Velarized	tˠ dˠ	ˡ Lateral release	dˡ
̠ Retracted	e̠	ˤ Pharyngealized	tˤ dˤ	̚ No audible release	d̚
̈ Centralized	ë	̴ Velarized or pharyngealized	ɫ		
̽ Mid-centralized	e̽	̝ Raised	e̝ (ɹ̝ = voiced alveolar fricative)		
̩ Syllabic	n̩	̞ Lowered	e̞ (β̞ = voiced bilabial approximant)		
̯ Non-syllabic	e̯	̘ Advanced Tongue Root	e̘		
˞ Rhoticity	ɚ a˞	̙ Retracted Tongue Root	e̙		

SUPRASEGMENTALS

ˈ Primary stress

ˌ Secondary stress ˌfoʊnəˈtɪʃən

ː Long eː

ˑ Half-long eˑ

̆ Extra-short ĕ

| Minor (foot) group

‖ Major (intonation) group

. Syllable break ɹi.ækt

‿ Linking (absence of a break)

TONES AND WORD ACCENTS

LEVEL			CONTOUR		
e̋ or	˥	Extra high	ě or	˩˥	Rising
é	˦	High	ê	˥˩	Falling
ē	˧	Mid	e᷄	˦˥	High rising
è	˨	Low	e᷅	˩˨	Low rising
ȅ	˩	Extra low	e᷈	˧˦˧	Rising-falling
↓		Downstep	↗		Global rise
↑		Upstep	↘		Global fall

have frequency characteristics define place of articulation. For example, the manner of articulation known as stop has a certain amount of 'time' where there is no sound, andwhether the stop is a [t] or a [k] depends on the energy at certain 'frequencies' at the release of the stop. These auditory distinctions lend themselves to acoustic interpretation.

Articulatorily, when it comes to speech sounds past the velum, the literature suggests that there can be lower, mid and upper pharyngeal constriction (Ladefoged and Maddieson, 1996). Upper pharyngeal constriction generally corresponds to sounds that are classed as uvular. According to Ladefoged and Maddieson (1996), mid pharyngeal constriction correlates with pharyngeal consonants while epiglottal consonants, found in the 'other symbols' section of the IPA chart, involve some lower constriction. In contrast, Bessell (1993) suggests that pharyngeals can be produced either "by (i) contraction of the upper pharynx where the palatoglossus and palatopharyngeus muscles (front and rear faucal pillars) intersect with the oropharynx and/or (ii) rearward movement of the tongue root, carrying with it the epiglottis and thereby achieving a constriction lower in the pharynx" (Bessell 1993, p. 41). The evidence in this thesis suggests that the upper pharynx constriction described in (i) by Bessell has the acoustic consequence of raising formants as in a raised larynx voice quality setting. The rearward movement of the tongue described in (ii) is found to be only part of what causes constriction in the pharynx. This thesis provides a more detailed account of constriction in the lower pharynx by comparing video and acoustic evidence with auditory phonetic definitions.

Yanagisawa et al. (1989) provide a physical description of pharyngeal constriction, whereby the epiglottis and aryepiglottic folds that are attached to the epiglottis can sphincter. This is the only 'constriction' available in the mid to lower pharynx. Below the aryepiglottic folds, the larynx tube does not offer constriction capabilities in the same way as other parts of the vocal tract, and neither does the tip of the epiglottis itself. It is the sphinctering of the aryepiglottic folds that has been identified by Esling (1996) to produce the pharyngeal consonants examined in this thesis and which is used to produce good singing styles observed in Yanagisawa et al.

Having only one place of articulation available in the pharynx makes it possible to reclassify the IPA sounds so that the epiglottal sounds can be included in the pharyngeal column. Lieberman and Blumstein (1991) state that "scientific theories... relate things that were thought to be unrelated" (p. 163). The theory developed in this thesis relates the terms pharyngeal and epiglottal with acoustic data and suggests that perceptually, pharyngeals and epiglottals are produced at the same place in the vocal tract. Further to the constriction capabilities in the pharynx, the larynx, which contains the vocal folds, can rise and lower beneath the constriction and may contribute to the perception of another place of articulation for post-velar sounds. This thesis examines acoustically, using language data, the plausibility of larynx height as a distinctive feature.

In order to describe pharyngeal movements, it is useful to consider some phonetic labels that are used to describe voice quality settings, since some of the characteristics observed for pharyngeal consonants have been observed in voice quality settings. The taxonomy of voice qualities proposed by Laver (1980) is important to the discussion of pharyngeal consonants since research in this area can provide evidence to help describe what types of behaviour are possible in the pharynx.

Laver's taxonomy of voice quality settings is organized according to settings that are created by modifying the source of sound, and are referred to as phonation types. Phonation types are further classified into simple types such as whisper, creak, modal voice and falsetto; and compound types, which are composed of more than one simple type. Settings that involve filtering the source sound are known as supralaryngeal settings and refer to habitual constriction in one area of the vocal tract or changes in overall length. For example, a raised larynx setting, which is auditorily associated with a Jim Hensen 'Muppet' voice quality, implies that the entire vocal tract is shortened because the larynx is raised.

Pharyngealization is another voice quality and is associated with the tongue being positioned back in the pharynx, more or less all the time, during speech. It has however, been found that a pharyngealized voice quality setting is the lower pitch version of raised larynx voice (Esling et al., 1994). This 1994 study suggested that raised larynx voice and pharyngealized voice are on an auditory continuum with respect to pitch. This implies that these voice quality settings are produced in the same way; the difference is that raised larynx voice is high in pitch and pharyngealized voice is low in pitch. Titze and Story (1996) provide a model that shows that constriction in the pharynx has the effect of slowing vocal fold vibration and therefore, lowering pitch. The constriction they observed is believed to be the same as for pharyngealized voice quality settings and thus explains the lowered pitch that is perceptually associated with pharyngealized voice. On the other hand, the raised larynx setting eliminates the impedance matching cavity found in low-pitched pharyngealized settings because the larynx rises and is located at the constriction; this causes pitch to rise, and the resulting shortened tube adds to the higher frequency energy.

In an effort to categorize what happens in the pharynx for speech segments, Laver's taxonomy of voice quality settings helps to describe the physical vocal tract setting for cardinal pharyngeal sounds examined in this thesis; and this voice quality setting taxonomy provides salient acoustic features to look for in the pharyngeal sounds. The voice quality setting for cardinal pharyngeal sounds is 'pharyngeal', which has the acoustic feature of a raised F1 and lowered F2 compared with sounds made with a neutral setting. Since a raised larynx setting is on the same perceptual continuum as a pharyngealized setting, features that define raised larynx settings may also be associated with pharyngeal sounds.

Comparing pharyngeal behaviour with the also controversial place of articulation known as glottal is useful since languages with pharyngeal sounds tend to have a series of glottals as well. The glottal/ pharyngeal distinction shows that glottals are affected by neighbouring vowels, that is, they mirror the vowel they are beside, and pharyngeals affect neighbouring vowels, or change the vowel they are beside. Acoustic evidence that would motivate including glottals in the post-velar class phonologically is still in debate. With respect to this thesis, acoustic evidence that classes glottals as post-velars is proposed to be the existence of energy across the entire frequency spectrum, and perhaps upper formant frequency information.

The majority of the acoustic investigation used for this thesis is performed on one subject who is producing 'ideal' or cardinal pharyngeal sounds with the intent that these sounds represent an example that can be used for comparison with language data. The subject is a trained phonetician who can produce all of the sounds on the IPA chart as well as Laver's Voice Quality Settings. It is believed that analyzing the phonetician's pharyngeal sounds will illustrate the capabilities of the human vocal tract and provide a set of model, or cardinal physical and acoustic parameters that can be compared with natural language data.

The acoustical study is enhanced by a video observation of the cardinal pharyngeal sounds produced by the same subject. The subject used for the video examination is not a native speaker of a language that uses pharyngeal sounds, but it is shown in chapter 5 that a native speaker may not be an ideal candidate for producing cardinal pharyngeal sounds because of their habitual pharyngealized voice quality setting. Thus, a native Hebrew speaker may inadvertently 'hide' pertinent articulatory information because the vocal tract is constricted in the pharynx more or less all the time.

Along with providing analytical support for the proposed description of pharyngeal sounds, the video examination develops a framework for video measurement. One problem with using video data for phonetic research is that it is difficult to discuss quantifiable measurements. This thesis modifies a technique outlined by Painter (1986 and 1991) to make a comparison between the cardinal pharyngeal sounds. The problem of obtaining a depth measurement in the video pictures still remains; however, the technique used in this thesis normalizes the video pictures so that each picture's measurements come from the same virtual depth and thus can be compared across subjects and sessions.

The results of this research also show the need to modify the existing source-filter theory of speech production. This model considers sound that is produced by the vocal folds to be the source, and the shape of the vocal tract above the source provides a filtering effect. A more detailed description of how pharyngeal constriction affects vocal fold vibration and resonance characteristics is of importance to the improvement of digital speech coding in such applications as digital cellphones and internet telephony.

1.1 Goal of the Thesis

The goal of this research is to systematically describe the role of the pharynx in speech. The study is comprehensive, since a variety of measurement techniques have been employed in order to observe this part of the vocal tract. Since the human pharynx is the back part of the throat connecting the nasal, oral, middle ear and glottal passages, it is difficult to observe this area non-invasively. This study employs acoustic, electroglottographic and laryngoscopic video measurement techniques in order to observe pharyngeal movements that are important in speech. The phonetic taxonomies are used to provide a framework for examining pharyngeal movements for this thesis. This survey of analysis techniques used to examine pharyngeal sounds offers an alternative interpretation of an existing linguistic taxonomy.

The existing IPA interpretation of the pharyngeal area of the vocal tract implies that there is only one linguistically relevant manner of articulation that is utilized in the pharynx, namely the pharyngeal fricative. It is proposed by Esling (1996) and here, that some members of the 'other symbol' set on the IPA chart could also be interpreted as pharyngeal in place of articulation. More specifically, the epiglottal plosive could be considered a pharyngeal plosive and the voiced and voiceless epiglottal fricatives appear to be pharyngeal trills. This re-arrangement of symbols would embellish the pharyngeal column on the IPA chart and may reflect more accurately the way languages use these sounds.

Another goal of this research is to aid in determining where and how the pharyngeal configuration changes in order to produce the cardinal sounds. This is accomplished by using video data. In describing the pharyngeal sounds using the nasendoscope, a measurement technique for analyzing video pictures is developed. Further to an acoustic and video description of cardinal pharyngeal sounds, the goal of this thesis is to provide an acoustic analysis of language data in order to determine if languages make phonetic use of the proposed sounds.

1.2 Thesis Outline

The outline of this thesis is as follows: An articulatory to acoustic mapping of cardinal pharyngeal sounds is made by describing each pharyngeal sound in terms of four manners of articulation. This description is also compared with general acoustic data to support the manners of articulation classification. Other acoustic cues for place of articulation are discussed with respect to general theories of speech production.

A qualitative examination of video data is discussed in chapter 3. A normalization technique is used to measure degree of pharyngeal constriction between the four manners of articulation which are discussed in chapter 2. Along with this is an attempt to correlate a sequence of video data with the acoustic output for the pharyngeal approximant.

An analysis of pharyngeal constriction on vocal fold behaviour is covered in chapter 4. Since the formant structure change is so subtle between the common vowel [ɑ] and the pharyngeal approximant, and because this constriction is very near the vocal folds, it is hypothesized that the salient feature of the approximant is the constriction effect on vocal fold behaviour. A description of changes in voicing between vowels and voiced pharyngeal articulation in the cardinal examples along with real language examples is discussed in support of the constriction in the pharynx having a marked effect on vocal fold vibration.

Acoustic language data taken from the University of Victoria Phonetic Database (PDB), and Sounds of the World's Languages (SOWL) database from the University of California at Los Angeles are examined in chapter 5. The observations in chapters 2 through 4 are compared with real language data in order to provide support for the taxonomy of four manners of articulation in the pharynx proposed by Esling (1996). Language data from Agul support the four manners of articulation, or an alternative interpretation which could involve larynx height being the distinctive feature.

The concluding chapter summarizes the theory of pharyngeal behaviour that has evolved from this thesis. The IPA taxonomy provides a framework which allows for an acoustical examination of pharyngeal sounds and supports the manner of articulation interpretation. General acoustical theory is pressed into service in order to support one place of articulation for the pharynx proper. The video examination also supports this interpretation of place of articulation. A detailed observation of vocal fold behaviour suggests that the salient feature that distinguishes between the pharyngeal vowel [ɑ] and a pharyngeal consonant is the effect of constriction on vocal fold behaviour. These discoveries combine to support the theory that, from an acoustic and articulatory point of view, the cardinal pharyngeal sounds produced by the phonetician are a series of four manners of articulation produced at one place of articulation. The acoustic language data also support this, however, further research from the phonologists who study these languages would be needed to provide evidence to show that languages use the sounds in this manner.

The motivation for larynx height as a distinctive feature comes from the acoustic characteristics that are observed in the larynx height voice quality settings. That is, there is a relationship between larynx height and pharyngeal settings, and the distinctive feature for these settings is pitch. The acoustic quality of a raised larynx setting, in general, is a rise in all formants. These features are all considered with respect to natural language data in chapter 5.

Chapter 2

Articulation to Acoustic Mapping of Cardinal Pharyngeals

2.0 Introduction

The purpose of this chapter is to give acoustic definition to Esling's (1996) proposed pharyngeal consonants. The pharyngeal consonants are comparable with the pharyngeal vowe [ɑ] and are defined acoustically as having the first three formant frequencies occuring at about 700Hz for F1 (first formant), 1000Hz for F2 and 2500Hz for F3. Esling's production of cardinal pharyngeal consonants is examined here acoustically in order to determine if these sounds can be classed as four manners of articulation. Also, from this acoustical study will determine if pharyngeal sounds are produced in one place of articulation.

The value of analyzing a phonetician's cardinal production of pharyngeal consonants is best described in the following quote from Lieberman and Blumstein (1988):

> "In order for languages to use phonetic categories in speech, it is necessary to have distinctive acoustic patterns to define that category so that the listener can perceive it, and it is useful that the articulatory system can produce this pattern even with some articulatory imprecision" (p. 186).

Examining cardinal productions acoustically will determine what salient acoustical features are used to distinguish between sounds produced in the pharynx. This evidence will also show "that the articulatory system can produce this pattern even with some articulatory imprecision," as Lieberman and Blumstein suggest. The degree of imprecision is not directly measured here but is considered with the examination in following chapters of a video description and various language data discussed in chapter 5.

Lieberman and Blumstein's formulation alludes to Stevens' (1972) Quantal Theory. This theory suggests that there are regions of acoustic stability that correlate with salient features in speech. That is, some sounds can be produced rather sloppily and still be understood. For example, we seem to be able to understand sounds such as [**i**] (as in 'heat') even when the speaker varies their tongue position within 2 cm (Johnson, 1997).

In the acoustical examination here, token examples of the cardinal pharyngeal consonants are produced by Esling and examined for features that define the particular manner of articulation. Along with noting manner of articulation features, predictive information can be derived by describing the shape of the vocal tract in the pharynx for these sounds. The observed resonance effects for vocal tract shape are compared with concepts proposed and discussed in Fant (1960), Sundberg (1987) and Johnson (1997).

The work of Ingo Titze, Jo Estill and Johan Sundberg is pertinent to a discussion of pharyngeal behaviour. Each of these researchers is interested in pharyngeal behaviour because of their interest in professional singing, and pharyngeal constriction has been associated with good singing styles. The constriction described by these researchers involves the epiglottis and aryepiglottic folds approximating as is seen in the pharyngeal sounds examined here. The pharyngeal postures that are thought to contribute to good singing are described by Yanagisawa, et al. (1984), and observed acoustically by Sundberg (1987); with the effect on vocal fold vibration described in Titze and Story (1996).

Characterizing the range of motion possible in the pharynx, according to the IPA taxonomy, is useful to language researchers since the taxonomy allows for comparison with other areas of the vocal tract. The IPA taxonomy provides a symbol for every speech sound and does so by organizing the sounds based on articulatory, perceptual and acoustic features as discussed in chapter 1. Further motivation for distinctive features comes from the phonologies of languages that use certain groups of sounds. The groups of sounds are motivated by their appeal to a sub-set of a segments distinctive features.

This acoustic discussion of features of pharyngeal sounds is organized as follows: Each of the proposed cardinal consonants is presented together with a description of what is observed articulatorily with a video nasendoscope. Following is a description of the general manners of articulation, and then a discussion of the qualities that define the cardinal pharyngeals. First is a description of the methodology.

Method: Prior to the acoustic analysis, production of the pharyngeal consonants was studied with the aid of a flexible fibreoptic nasendoscope. This device is inserted through the nose and is positioned just behind the velum (see Figure 1). However, the subject is so comfortable with the procedure that he is able to insert the scope lower in the throat, beneath the apex of the epiglottis. This method allows unobstructed viewing of the larynx and pharynx and gives the view seen in Figure 2. The pertinent structures are labeled and discussed with respect to the manners of articulation in which they participate.

In general, Figure 2 represents the vocal tract during normal voicing. The *larynx* contains the vocal folds within the laryngeal tube and it is situated at the bottom of the *pharynx*. The larynx is bound by the epiglottis, just above the tubercle at the base, and the aryepiglottic folds which attach to the cuneiform and arytenoid cartilages. The *epiglottis* is attached to the base of the tongue, and the *pharyngeal wall* forms the back of the throat. It may be useful to examine this figure along with Figure 1 in order to better understand the position of structures within the larynx and pharynx.

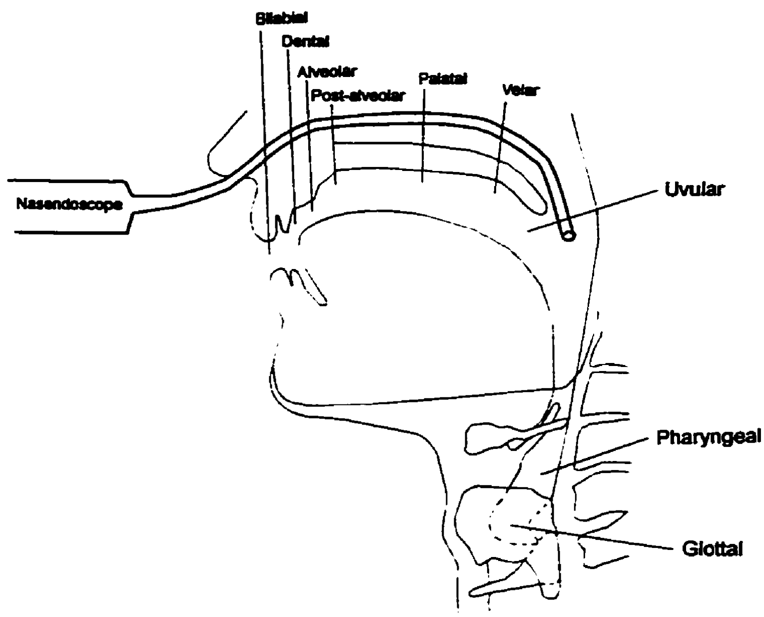


Figure 1: Positioning of the nasendoscope for viewing of the pharynx and larynx. Labeled are the places of articulation according to the International Phonetic Alphabet (IPA).

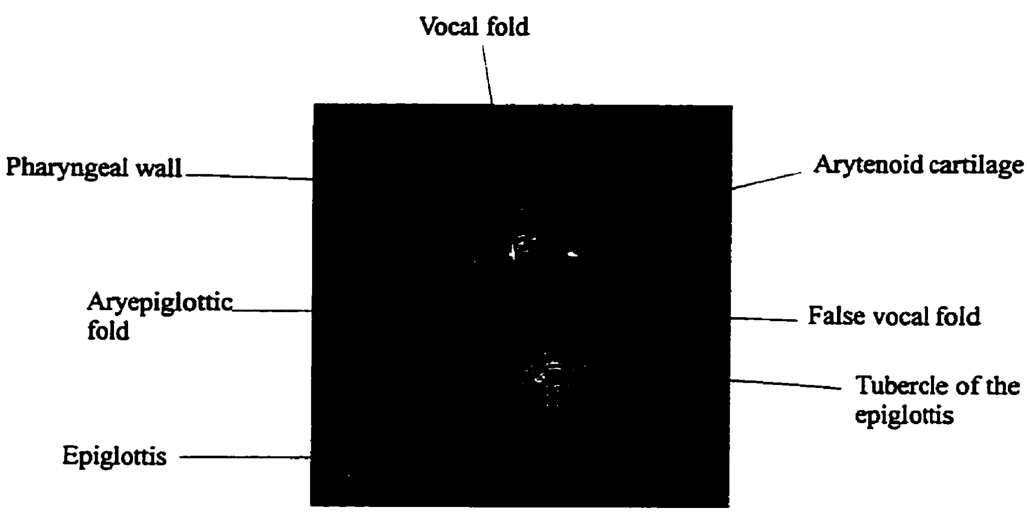


Figure 2: The view from the nasendoscope positioned as in Figure 1.

The acoustic output that correlates with the video was captured onto DAT from the video tape and from separate recording sessions. The microphone signal was enhanced and conditioned by a Nakamichi mixer MX-100 before going to DAT via line input to ensure the signal was not overdriven and to eliminate noise that could be picked up by going in direct. The data were analyzed using the Computerized Speech Laboratory (CSL) 4300A, Multi-Speech 3700 version 1.20, and Analysis by Synthesis Laboratory (ASL) from Kay Elemetrics Corp.

2.1 - The Pharyngeal Plosive [ʡ].

The initial stage of a pharyngeal stop is hypothesized to be the same as a glottal stop. The production of a glottal stop involves the false vocal folds pressing onto the vocal folds enough to stop them vibrating. Other structures in the pharynx, such as the epiglottis, remain stationary. In the pharyngeal stop, the same damping of the vocal folds occurs, but the larynx then rises up to the epiglottis which moves back toward the pharyngeal wall.

General Acoustic Description: Stops appear to have a variety of acoustical measures, namely: Formant transitions, vowel length, closure length, and burst behaviour. These measures will be examined with respect to the pharyngeal plosive.

There are three intervals that are associated with stops in general: Closing, closed and opening. Formants are visible in both the closing and opening phases, thus providing information on the shape of the vocal tract going into and coming out of the stop. The burst is a measure of speed of the articulators during the opening phase. Physically speaking, the formants of an utterance provide organized acoustic energy in the frequency domain whereas burst noise is the organization of energy in the time domain (Halle, Hughes, Radley, 1991, p. 224).

The three intervals of a stop help to provide information on the place of articulation and voicing distinction. The closing phase of a stop involves the vocal tract movement and configuration shaping the formants in a way that indicates the place of articulation of the stop in phonetic terms. The closed phase can be relatively longer or shorter and contributes to the perception of voicing. Such is the case in the contrast between the English [**b**] and [**p**] where the voiceless [**p**] has a longer closed phase than its voiced counterpart. This closed portion, along with vowel length, has been suggested as a cue for voicing (Lisker 1978).

Production of a pharyngeal plosive creates one large front oral cavity because the larynx rises up to the level of the epiglottis. The large oral cavity description would be the 'normal' case of production for the plosive; however, Esling (personal communication) has evidence to suggest that a lowered or raised larynx setting can be applied to the stop. It is the author's hypothesis that this lowered larynx setting involves the same rising of the larynx to the epiglottis to create the pharyngeal stop, but the unit (the epiglottis and arytenoid cartilages) is *tilted* so that the epiglottis and tongue are pushed toward the pharyngeal wall. This may give the opportunity for the pyriform sinuses to be expanded and consequently resonate.

Stops in the oral region are distinguished according to their formant transition to the following segment. These transitions are predictable and visible in the spectrogram. Further support for formant transitions as a cue for distinguishing stops has been obtained by synthesizing these transitions and then presenting them to listeners (Cooper et al. 1952). Cited in Johnson (1997, p. 136) the F2 locus, or ideal start frequency for bilabials is around 700Hz, alveolars, 1800Hz, velars, 300Hz. Burst and F2 locus are associated with front cavity resonance for velars, and in the case of pharyngeals, it is likely the same since, the back cavity is uncoupled from the front at the epiglottis, making a large oral cavity. With this in mind, the F2 locus for pharyngeals is likely to be slightly lower than for velars as a consequence of the front cavity increasing.

<u>The Waveform:</u> To illustrate the features of a pharyngeal stop it is compared with a glottal stop, as seen in Figure 3. In all figures, the IPA transcription of the word is centered over the acoustic event. The top waveform of Figure 3 is an intervocalic glottal stop. The first part of the waveform corresponds to the vowel [ɑ] and the straight horizontal line portion is the stop and the last bit is another [ɑ] vowel. The bottom waveform is the pharyngeal stop combined with the same vowels.

Comparing these two waveforms, the vowel preceding the stop is shorter in the pharyngeal stop [ʡ]. The closed portion is longer than the glottal stop [ʔ]. These observations are predictable since more effort is required to raise the larynx to the epiglottis and this should cause a longer stop portion and larger amplitude upon release due to increased lung pressure.

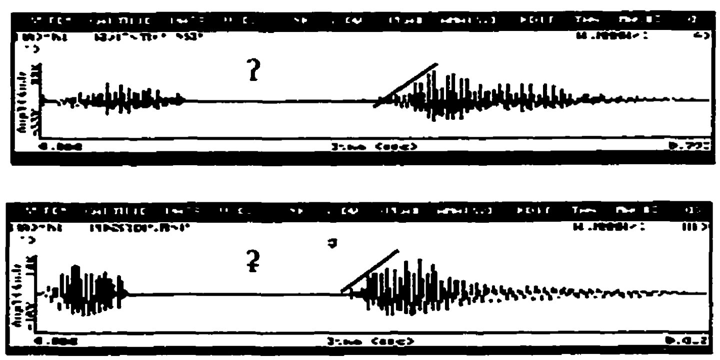


Figure 3: Waveforms of the pharyngeal stop (bottom) compared with the glottal stop (top). The stop gap for the pharyngeal is longer and the preceding vowel is shorter compared with the glottal. Also note the steep amplitude of the following vowel.

The relationship between glottal and pharyngeal stop intervals is comparable with a lenis/fortis type distinction as described in Ladefoged and Maddieson (1996) in Zapotec languages of Mexico. That is, the pharyngeal stop requires more articulatory force than the [ʡ] and is analogous to fortis production. The distinction of lenis/fortis has also been applied to voiced/voiceless distinction, in English. In the case of post-velar sounds, if the pharyngeal stop is fortis, then it would be analogous to a voiceless segment and the glottal stop would be considered the voiced stop.

The implications of a pharyngeal component in the lenis/fortis description of stops in languages is not expanded here, but is certainly worth investigating. It is predicted the results would better define fortis stops as having a pharyngeal component, eliminating the less precise terminology of lenis/fortis.

The steepness of the amplitude burst noted in Figure 3 has been associated with the fricative/affricate distinction (Johnson 1997, p. 138). The short rise time that is associated with affricates, in this case, would correlate with the pharyngeal stop. The longer rise time then, associated with fricatives would correlate with the glottal stop. This is illustrated in Figure 3 by the angle marked on the following vowel in the waveform. There is motivation for using a fricative/affricate interpretation of these sounds from some Native North American languages discussed later, however, it is questionable whether this feature is salient as is seen in the Hebrew data in chapter 5.

Pitch: Pitch of the vowels preceding and following the pharyngeal stop is higher than in the environment of a glottal stop. Increased pitch is sometimes achieved by raising the larynx (Esling, Heap, Snell and Dickson, 1994). Therefore, it is reasonable to assume that rising of the larynx observed in the video data for the [ʡ] would correlate with an increase in pitch in vowels surrounding a stop.

The Spectrogram: The effect of complete pharyngeal occlusion on the following vowel is seen in Figure 4. The pharyngeal stop in the environment of an [ɑ]-vowel (Figure 4) shows no obvious formant transitions. However, in the environment of [i] (Figure 5) there is marked formant behaviour visible in the spectrogram for the following [i]. The preceding vowel in Figure 4, does not show a transition because the sequence of events for the pharyngeal stop starts with vocal folds closing which stops the noise source, then the larynx rises causing the more massive stop. Comparison of these spectrograms with the glottal stop (Figure 6 and 7) shows that there is no formant transition at all for the glottal stop in either the [i] or the [ɑ] environment.

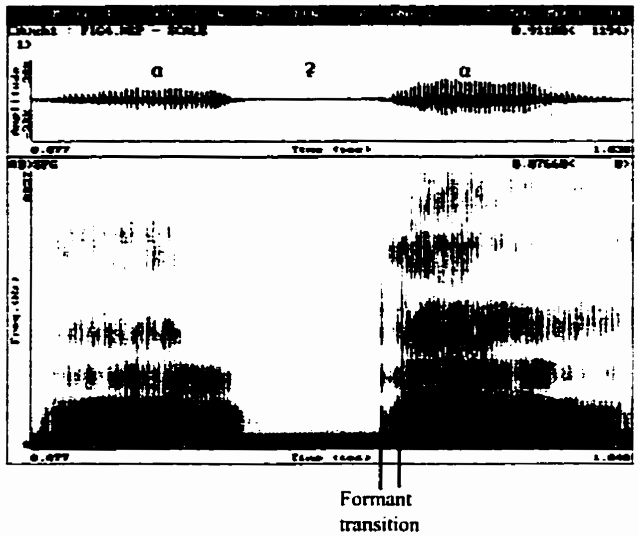


Figure 4: Note the formant transition in the following vowel for this sequence of [ɑʔɑ].

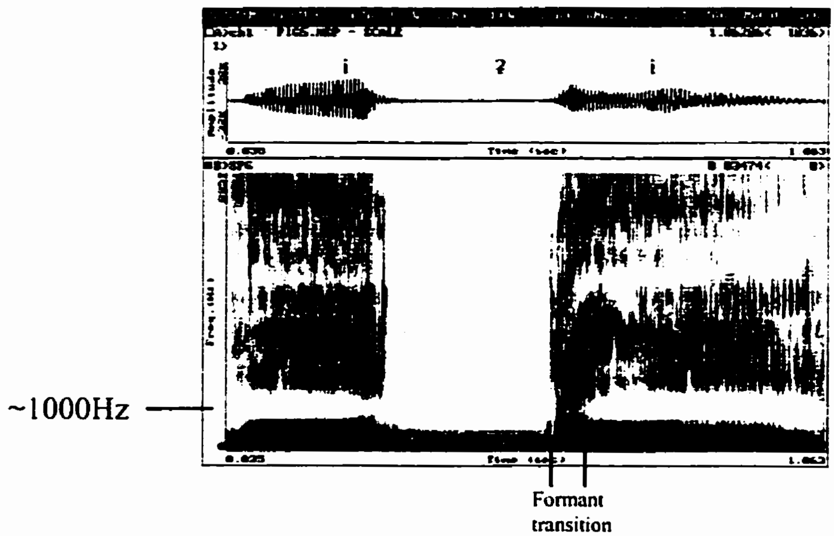


Figure 5: Note the formant transition in the following vowel for this sequence of [iʔi].

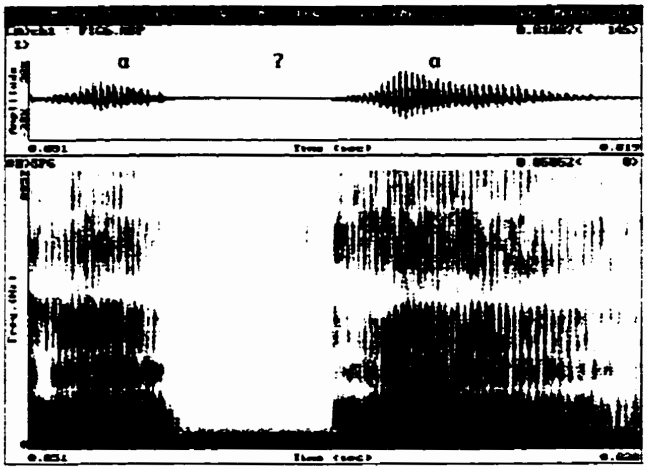


Figure 6: Formant behaviour remains stable throughout the transition regardless of the vowel, in the case of the glottal stop [ɑʔɑ].

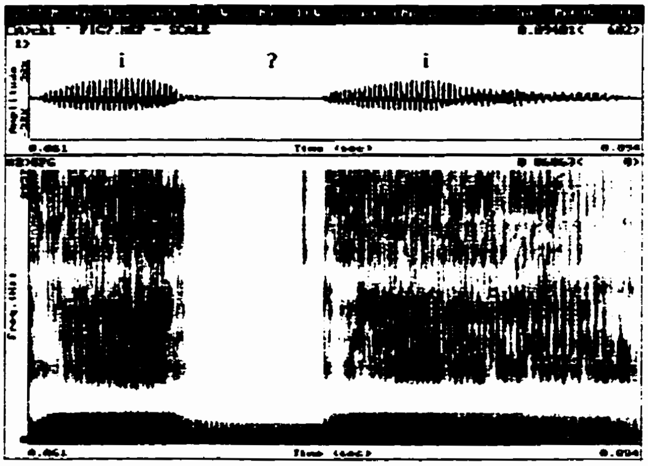


Figure 7: Formant behaviour remains stable throughout the transition regardless of the vowel, in the case of the glottal stop [iʔi].

Synthesis: To test further if pitch, vowel length, closure length and amplitude rise time characteristics are pertinent for pharyngeal stops, a glottal stop token was manipulated and synthesized using ASL. The following parameters were altered: Pitch was increased in the vowels preceding and following closure, the initial vowel was shortened, and the stop closure was lengthened. These adjustments result in a sound, judged by a phonetician and by a native speaker of Farsi, who is familar with Arabic, to be similar to the original production of a pharyngeal stop.

Conclusion: The pharyngeal plosive has the general acoustic feature of a plosive in that there is a stop portion in the waveform. The effect on formants surrounding vowels suggests that for a pharyngeal spoken in the environment of an [i]-vowel there is marked formant transition, in that it looks like the formants are 'coming' from an [ɑ] like articulation. The glottal stop shows no formant transitions, suggesting that there is no change in vocal tract area function for the production of a glottal stop, regardless of the neighbouring vowel.

The characteristics of the three intervals of the pharyngeal stop, that is, the closing, closed and opening stages, shows similar characteristics as with the voicing distinction in other areas of the vocal tract. The pharyngeal stop behaves acoustically like a 'voiceless' glottal stop in that the initial vowel is shorter compared with the glottal stop, the closure time is longer and the amplitude rise time in the following vowel is shorter for the pharyngeal stop. Synthesis of these features provides enough of a cue for identification of a pharyngeal stop for at least two subjects. This implies the glottal and pharyngeal plosive could act as a voiced/voiceless pair respectively, based only on the acoustic description.

The acoustic characteristic of amplitude rise time for an affricate contrast (Johnson, 1997) offers an alternate description for pharyngeal stop behaviour. Evidence to support this interpretation in found in an auditory language description provided by Bessell (1993). She notes that a dialect of the Native North American language, called Haida, has a segment that has been described as a pharyngeal affricate. Acoustic data is not documented for this claim to date.

The acoustic description of a pharyngeal stop comparison offers two alternatives for a phonologist faced with pharyngeal stop-like segments in a language. One is that the contrasts between the segments mirror a voiced/voiceless contrast and the other is that the segment may be an affricate. It is up to the phonologist to test these distinctions in order to determine if languages with a glottal/pharyngeal plosive contrast prefer a voicing distinction or a manner-of-articulation distinction between the segments.

2.2 Pharyngeal trills [ʜ, ʢ]

The production of a pharyngeal trill involves the larynx rising enough to allow the aryepiglottic folds to vibrate against the epiglottis.

General Acoustic Description: The waveform in Figure 8 illustrates the quick, regularly spaced pulses associated with trilling of something other than the vocal folds: This type of pulsing is characteristic of trills in general. Note that outside the marked portion is the vowel which has regular pulsing that is faster and is associated with vocal fold behaviour.

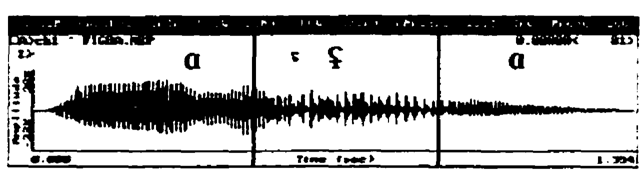


Figure 8: The waveform is of the sequence [ɑʢɑ]; the marked portion shows slower vibration since the pulses are spaced farther apart.

The trilling that occurs during the consonant portion pulses at a rate of 50 times a second, which is faster than lingual trills at about 25 Hz, but slower than the vocal fold rate of about 100Hz for this speaker.

It is plausible that the aryepiglottic folds would vibrate faster than the tongue but slower than the vocal folds because these folds are less massive and dense than the tongue but more massive than the vocal folds. Though the 50Hz vibration is a plausible rate of vibration for the aryepiglottic folds, the mechanism of trilling is likely influenced by the glottal source below the pharyngeal constriction. This is also suggested by the reduction in rate of trilling observed in the voiceless version, to about 40Hz.

<u>The Spectrogram:</u> The spectrogram in Figure 9 shows that formant behaviour does not change markedly from the [ɑ] vowel through the [ʕ] to the following [ɑ] vowel. This suggests that the vocal tract configuration does not radically change throughout the [ʕ] segment. A comparison between the [i] vowel in the approximant in Figure 10 illustrates the effect of change vocal tract posture.

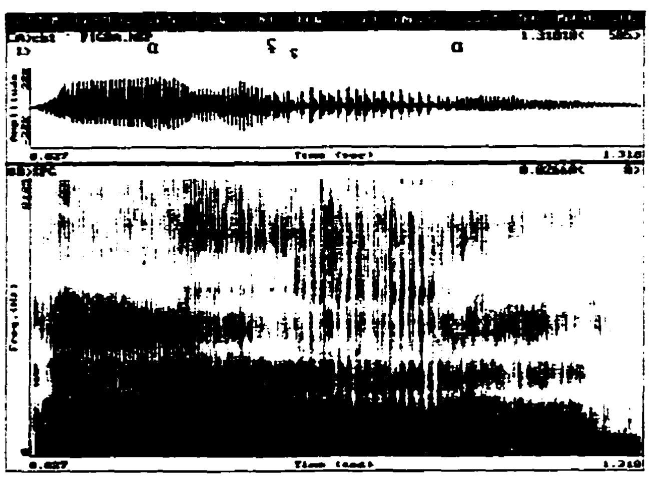


Figure 9: Spectrogram of the sequence [ɑʕɑ] showing no marked formant changes.

The vowel [i] is described as having a low first formant and high second formant compared with perfectly evenly spaced formants in a totally open vocal tract. The vowel [ɑ], on the other hand, shows a merging of the first and second formants towards the 1000 Hz mark. Figure 10 shows this comparison rather dramatically. This figure contains the spectrogram of the approximant sequence [iʕi], in order to illustrate the formant effects.

The uvular trill shown in Figure 11 illustrates the slower trilling rate (~25Hz) compared with the pharyngeal in Figure 9. With respect to formant behaviour, the difference

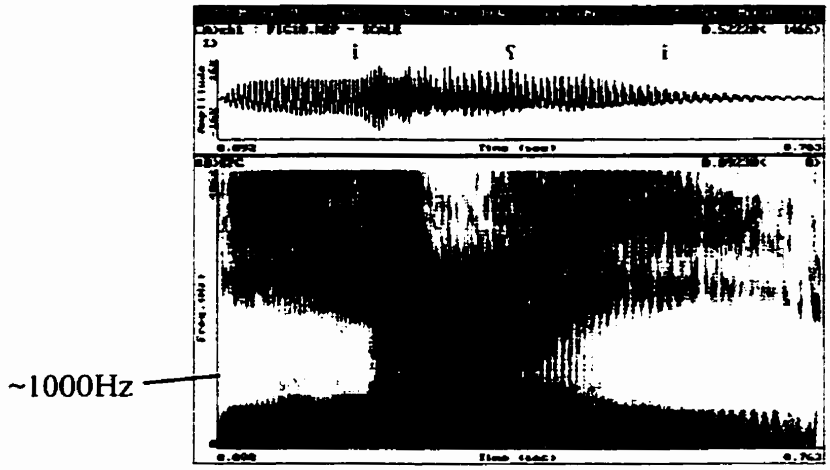


Figure 10: The sequence [iʕi] showing marked formant movement. The approximant is used to illustrate the formant behaviour.

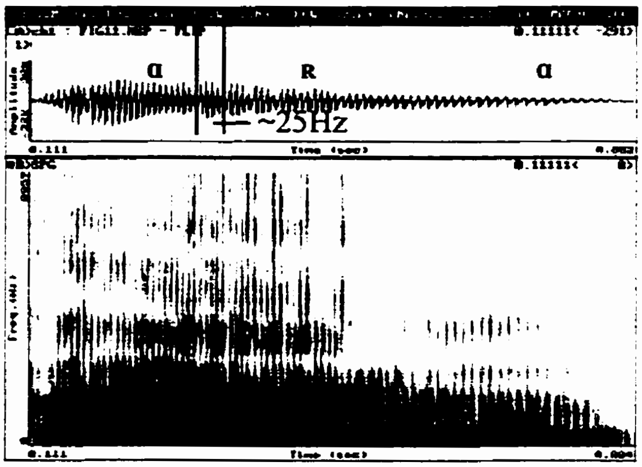


Figure 11: The uvular trill in the [ɑʀɑ] sequence.

between the uvular and pharyngeal appears to be in the upper frequencies and is marked by a weakening of energy in the case of the uvular.

Conclusion: The pharyngeal trill has the general acoustic feature of successive quick stops resulting in pulses in the waveform that do not correlate with vocal fold vibration, since the frequency is much lower. Compared with trills in other areas of the vocal tract, the pharyngeal trill vibrates at a faster rate. Trilling that is faster than the uvular trill and slower than vocal fold vibration is explainable if the trilling is associated with the thinner aryepiglottic folds.

The formant structure seen in the spectrographic evidence suggests that the vocal tract remains relatively stationary throughout the VCV sequence when the vowel is [ɑ]; this suggests that there is no need to class the sound as epiglottal rather than pharyngeal. Spectrographic comparison with a uvular trill suggests that pharyngeal movement is marked by changes in the upper frequency range. As noted the pharyngeal trill can be produced with or without voicing which poses a logistical problem since no other trills in the IPA are voiceless.

2.3 The Voiceless Pharyngeal Fricative [ħ].

The production of a pharyngeal fricative involves the larynx rising to meet the epiglottis. The arytenoid cartilages are abducted to reveal the glottal opening throughout the production of the fricative.

General Acoustic Description: Fricatives are characterized auditorily and acoustically by noise or aperiodic waveform behaviour that is continuous. One quality of this noise is that there is energy across the entire frequency range which is visible in the spectrogram, in Figure 12. Lieberman and Blumstein (1991, p. 227) claim the overall amplitude of a fricative is associated with place of articulation, such that [s] and [ʃ] (as in ship), have greater amplitude than [f] and [θ] (as in thin), for example. The amplitude of the waveform then, diminishes the farther back the place of articulation is in the vocal tract. This would predict that a pharyngeal fricative would have the lowest amplitude. The degree of constriction for fricatives and the type of structure that causes the friction, is also associated with relative amplitude. That is, more constriction in the vocal tract, particularly if it is caused by a narrowed wall, results in lower amplitude (Johnson 1997). The aryepiglottic folds have the capability of forming a tight-walled constriction, thus contributing to lowered amplitude.

The Waveform: In the waveform above the spectrogram in Figure 12, there is marked low amplitude of the signal during the consonant compared with the [ɑ]-vowel portion. Considering the above information, Figure 12 suggests the pharyngeal fricative is produced relatively far back and with a great deal of constriction.

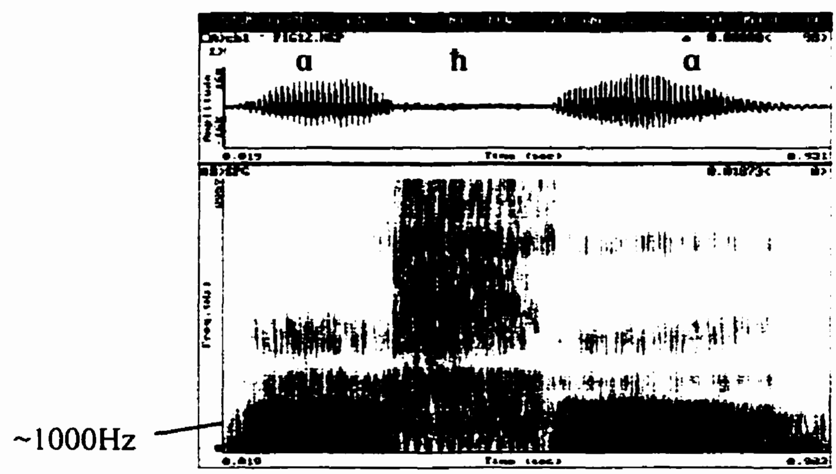


Figure 12: The voiceless pharyngeal fricative. Note the lack of periodicity (vertical striations) in the middle of the sequence and the low amplitude in the waveform in screen A (data is captured at 10kHz).

In Esling (1996, p. 77), the aryepiglottic folds are proposed as the active articulator for the consonants studied in this thesis. Support for this argument comes from the discussion of fricatives and whether the friction is created by a wall or obstacle. Johnson (1997) suggests that fricatives with lower amplitude are usually classed as wall fricatives, such as the bilabial [ɸ], or palatal [ç]. Considering the wall vs. obstacle description, the pharyngeal could be classed as a 'wall fricative' because of its low amplitude. And, based on video observations (described in chapter 3) it appears that the aryepiglottic sphincter that is created by the epiglottis and the aryepiglottic folds provides a narrow cylindrical wall that creates friction.

The Spectrogram: The vowel portions of the spectrogram in Figure 12, show clear vertical striations that are equally spaced and correlate with glottal pulse periods, whereas the consonant section shows irregular behaviour that is associated with friction. Cited in Lieberman and Blumstein is research that shows that as a major frequency peak is lowered from 5kHz to 2.5 kHz, perception goes from [s] to [ʃ]. Based on this comparison, pharyngeal behaviour should be lower still, since the constriction is farther back in the vocal tract. This would allow a larger cavity that accommodates lower frequencies to resonate. This is seen in the spectrogram in Figure 12. Having such a large cavity in front of the place of articulation available for resonance results in the first and second formants in the noise portion of the spectrogram. These formants and the resonance characteristics will be discussed in Section 2.5.

Conclusion: The pharyngeal fricative has the general acoustic feature of a fricative in that there is random noise observable in the waveform, in the spectrogram, and auditorily. The other marked observation in the spectrogram is the presence of formants. This implies that the source of noise is produced in a location that allows the entire vocal tract to resonate.

One question that Johnson's (1997) general discussion of fricatives inspires is, whether the pharyngeal fricative is produced by a wall, or an obstacle. Based on the acoustic observations here the pharyngeal fricative is likely a wall-articulated fricative since an obstacle would create more turbulence and therefore be higher in amplitude. The pharyngeal fricative tends to be quite low amplitude. This would imply that the epiglottis cannot be an obstacle that causes friction but more accurately the whole larynx tube is narrowed.

2.4 The Voiced Pharyngeal Fricative [ʕ] as an Approximant.

The production of a pharyngeal approximant involves the larynx rising to meet the epiglottis and obscuring the larynx completely, in the same manner as the stop.

General Acoustic Description: Figure 13 is a spectrogram that shows approximant characteristics: There is no stop portion between the vowels and no trilling effects or obvious fricative qualities. In a sense, the approximant is unique because it looks almost vowel-like spectrographically. For other approximant segments, such as [w], the formant transitions from a neighbouring sound provide the acoustic cue. The vowel [ɑ] is considered pharyngeal because it is the lowest and most back vowel; for a pharyngeal approximant then, it is expected that the formant structure would resemble [ɑ].

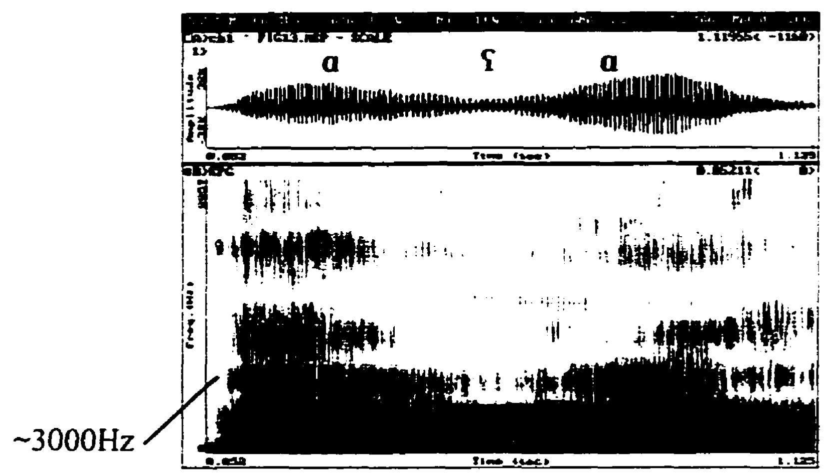


Figure 13: This spectrogram of the sequence [ɑʕɑ] shows no plosive, trilling or fricative-like qualities. The formants mimic those seen in the fricative in Figure 12, which suggests that these sounds are produced at the same place in the vocal tract.

The Spectrogram: The spectrogram of the pharyngeal consonant [ʕ] next to [ɑ] shows very little formant transition in the first three formants. In fact, if only the bottom half of the spectrogram were visible, one would likely confuse the spectrogram with a very long vowel. However, auditorily there is a definite change between the beginning and end of the utterance.

The low amplitude in the waveform during the consonant can be accounted for if the articulation is caused by the aryepiglottic sphinctering mechanism and is considered to be a wall articulation rather than an obstacle (Johnson 1997 p. 119). Looking at the formant structure in Figure 13 there is a notable change in formant behaviour above 3000Hz. Most notable is a weakening of energy above 3000Hz. However, this weakening is not expected to be a distinguishing feature since the fricative seen in Figure 12 has strong energy across the frequency spectrum. Because of the fricative's strong energy, the formant behaviour in the fricative is more valuable for examining formant changes between the [ɑ]-vowel and pharyngeal consonant place of articulation. The formants in the fricative are examined by looking at a Fast Fourier transform (FFT) display of the vowel [ɑ] compared with the pharyngeal fricative [ħ] and are discussed in Section 2.5.

Conclusion: The pharyngeal approximant shows a change in the upper formants that are not normally considered in phonetic acoustic analysis, except for singing. This suggests that formant structure may not be a salient feature of the [ʕ] compared with [ɑ]. However, the marked drop in amplitude during the consonant appears to be the most obvious feature that may be salient to the listener. It seems that the manner of articulation features account for the cardinal consonants discussed in this chapter, the most puzzling distinction is between [ʕ] and the vowel [ɑ]. Chapter four discusses the effect of the physiological production of the pharyngeal consonants and their effect on vocal fold behaviour, which seems promising for describing the salient acoustic features of the pharyngeal consonant with the vowel.

2.5 Acoustic Theory

The previous sections have illustrated qualities of the manners of pharyngeal articulation and only alluded to the features that support these consonants being produced in the same place in the vocal tract. This section will discuss in more detail acoustical theory and how it supports one place of articulation for the cardinal pharyngeal sounds.

In order to determine how sound is modified in the vocal tract, it is useful to draw an analogy. The vocal tract is essentially a tube, like a brass instrument, which has its source of noise at the vocal folds rather than the lips. Consequently, a brass or wind instrument player is adding another tube and uncoupling the noise source from the vocal folds so that air from the lungs is changed into air pressure changes at the lips. This is a long-winded way of saying a brass instrument player is playing with an external vocal tract. Using this analogy of the vocal tract with a musical instrument can help to explain acoustic behaviour of sound in the vocal tract and pharynx in particular.

The vocal tract isn't nearly as simple as a brass instrument in reality because of its soft walls and other fleshy bits, but modeling the vocal tract as tubes presents an approximation of the vocal tract with pharyngeal constriction. The human vocal tract is complex not just because the structures are fleshy, but because there are a few 'tubes' that are attached and are within the vocal tract. For example, the nasal cavity can be opened to and closed off from the vocal tract proper at the velopharyngeal port. The vocal tract can also be chopped up into smaller tubes by the tongue. For example, the vowel [i] has the tongue constricting near the front of the mouth creating a small tube at the front of a long tube at the back. Also in the back tube is another tube, namely the epilaryngeal lumen, that can pop up into the back tube. It is this little tube that is the focus of this study, since it 'pops up' in all the cardinal pharyngeal sounds.

Before considering the acoustic effects of the larynx tube that is inside the pharynx, cardinal pharyngeal sounds can be described according to the vowel [ɑ]. This vowel is the lowest and most backed of all the vowels and thus describes the most extreme sound of this kind that is commonly found in human speech.

A two-tube model implies that for the vowel [ɑ], the extreme backedness essentially chops the vocal tract in half. Johnson (1997) describes this model with the following formula for a tube open at one end: $f_{b\,or\,f} = (2n-1)c/4l_{b\,or\,f}$. The subscript *b* stands for back cavity and *f* stand for front cavity; *f* -represents (f)requency *c* -the speed of sound in air (about 330 meters per second), and *l* -(l)ength of the tube. This mathematical formula states that the frequency of a sound in a tube open at one end is increased with the speed of air in sound and decreases as the length of the tube gets longer.

When l*b*=l*f*, or the front and back cavities are of equal length, there should be a concentration of energy at about 1000 Hz, if the vocal tract is about 17 cm long and the speed of sound is about 340$^{m}/_{s}$. This would show up in a spectrogram as F1 and F2 merging, and illustrated in Figure 10.

Fant's (1960) data show formant frequencies for the vowel [ɑ] as follows:

Formant	Frequency	Bandwidth
1	650	94
2	1075	91
3	2463	107
4	3558	198
5	4631	90

These frequencies show that the vocal tract is not quite chopped in half, but the frequencies of the first and second formants are certainly closer to 1000Hz than for an [i]-vowel at 250Hz and 2500Hz (see Figure 10).

The effect of the larynx tube popping up into the back cavity has been said by Sundberg (1987) to have its own resonance frequency, which is quite a bit higher in frequency, at about 2500Hz since the tube has a shorter length (*l* in the formula above). This small tube has been made analogous with the mouthpiece on a brass instrument (Titze et al. 1996). That is, it has its own 'popping' frequency (Rossing, 1990), but perhaps more importantly to Titze et al. is that it matches the high impedance of the vocal folds with the low impedance of the vocal tract, thus having a marked effect on how the vocal folds vibrate. This topic is further covered in chapter 4, for now the 'popping' frequency of the larynx is something more pertinent because it creates a small tube, and since a smaller tube resonates at higher frequencies it is likely to show up as changes in the upper formants.

Figure 14 shows the FFT power spectrum taken during the [ɑ]-vowel portion (dark line). Overlaid is a spectrum during the pharyngeal fricative consonant portion which shows the shifts in formants: The energy in the first formant region remains at about 800Hz for both sounds. There is a shift downward in the second formant to about 2100Hz for the fricative. The fourth formant shows a pronounced peak and is shifted up to about 4400Hz. Also, the energy in the 5-6kHz region is more concentrated during the consonant portion.

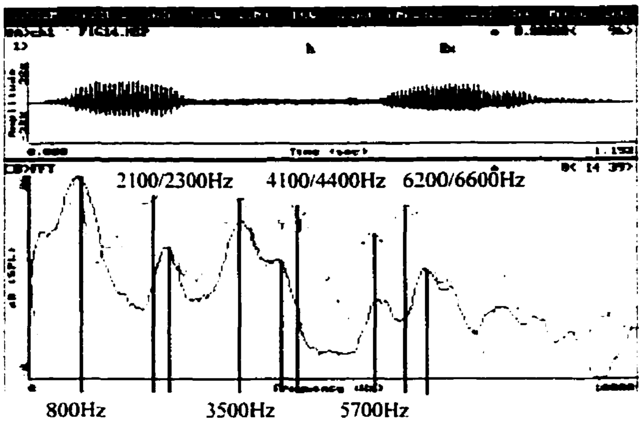


Figure 14: FFT power spectrum at a point during the vowel (dark line) compared with the consonant.

A model of upper formant changes is provided by Fant (1960). Fant's nomogram of a horn shaped three-parameter model (three sections of a tube) without the lip section (1.4-11 (c) p. 84) seems to best represent the observations seen here. The curve in his nomogram that matches the closest with the behaviour observed here implies a very small degree of opening at the constriction closest to the glottis. This is consistent with what is observed in the video data here, however, even Fant's predictions do not account for the shifts in energy above 3500Hz.

Though Fant claims that the upper formants bear no phonetic relation, it appears the difference, spectrographically, between the approximant and the vowel [ɑ] is seen in the upper formants. The effect of this pharyngeal constriction on vocal fold vibration is examined in chapter 4 and suggests that pitch, and other more fricative-like qualities are more likely the distinguishing feature between the pharyngeal consonant [ʕ] and the vowel [ɑ].

2.6 Conclusions

The acoustical facts presented here lend support to the claim that there are structures within the pharynx that can produce varying manners of articulation. Moreover, there does not appear to be much evidence to suggest the need for more than one place of articulation in the pharynx. This would modify the 'epiglottal' category in the present IPA treatment of post-velar sounds. Thus, the model that posits epiglottal and pharyngeal places of articulation for post-velar differences in the auditory category, is not supported by the evidence presented throughout this chapter.

The acoustic description of these cardinal sounds shows that they have qualities that define manner of articulation for other regions of the vocal tract. Namely, there is stoppage in the waveform for stops, noise in the spectrogram for fricatives, pulsing that is not associated with the vocal folds for trilling, and vowel-like qualities for the approximant. Comparison with general acoustical theory suggests that the formant behaviour of these sounds is consistent, implying that they differ only in manner of articulation.

Low amplitude that is noted in the waveforms of the pharyngeal sounds suggests the constriction is formed by a wall-like structure rather than an obstacle, as is discussed in Johnson (1997). Johnson's discussion suggests that if the constriction were caused by an obstacle, such as the epiglottis, the waveform amplitude would not be affected to the degree observed here. This fact gives support for the aryepiglottic sphincter, rather than the epiglottis or tongue, being the active articulator, as proposed by Esling (1996).

Of importance to the acoustical theory of speech production, the results of this acoustic description illustrate the need to modify the present source-filter theory. Modifications need to account for the change in upper formants that are observed in the post-velar segments examined here. Also, the model must account for the effect of pharyngeal constriction on vocal fold vibration. It is expected that pharyngeal effects are more pertinent to voice quality settings, however, the description developed in this chapter, with respect to sound segments, offers a more thorough account of movement in this region of the vocal tract.

One issue that remains is the justification of a voiceless pharyngeal trill when there are no other voiceless trills in the IPA taxonomy. There is another notable imbalance with respect to the fricative manner of articulation. Arguments in Laufer (1996) and the acoustic evidence presented in this thesis motivate re-classifying the voiced pharyngeal fricative as and approximant; however, doing so would leave a gap in the voice fricative box. One argument to support having only a voiceless pharyngeal fricative is that a voiced fricative and the approximant would be easily confused because of the increase in friction noise noted in the glottal source for an approximant. These issues beg further research; however, noting the characteristics of all pharyngeal sounds in this systematic manner provide characteristics that can be used for accurate speech recognition of the range of pharyngeal sounds that are humanly possible.

Chapter 3 - Video Description of Cardinal Pharyngeal Postures

3.0 Introduction

The purpose of this chapter is to gain a better understanding of pharyngeal behaviour by examining the cardinal pharyngeal consonants through a video-nasendoscope. It is possible to position the nasendoscope so that there is a clear view of the pharynx and larynx and so that is does not affect speech production. The positioning of the nasendoscope, and the view that it shows are illustrated in Figures 1 and 2. Quantifiable measurements from the two dimensional images are difficult to obtain; however, the measurement technique described by Painter (1986) is modified in this thesis and used to compare the most constricted point of each cardinal consonant. This was done in order to determine if the manner of articulation could be distinguished by the degree of constriction visible in the pharynx.

Also, a sequence of video pictures is examined along with the acoustic output in order to see if there can be an acoustic to articulatory mapping of pharyngeal behaviour from the video. Sundberg (1987) describes that when the ratio of laryngeal opening is 16% of the size of the pharyngeal opening, and the larynx is low, then there should be a concentration of energy in the 3-5kHz range. This is commonly referred to as the singer's formant. This ratio is easily measurable using two dimensional video pictures and is used in the sequence of pictures in order to determine if the pharyngeal sphinctering observed in these sounds contributes to changes in upper formant frequencies.

Using data from a videonasendoscope has one major drawback which is that it is difficult to compare pictures because the camera depth can change during a session, between sessions or between subjects. To minimize this, Painter's (1986) technique is tested to see if a glottal length measurement can be used to normalize all video photos so that the measurements essentially come from the same camera depth. In this study, an adaptation of Painter's normalization technique is used to determine if the range of constriction for the proposed pharyngeal consonants can be seen and measured in the video pictures. Normalization is also used when measuring the ratio of laryngeal opening to pharyngeal opening in order to confirm that the camera has not moved drastically.

The complexity of movements in the larynx is difficult to capture because there is movement in three dimensions. Also, some of the laryngeal structures in this area are small, agile and are sometimes obscured by fleshy folds; not to mention that, in addtion to from antero-posterior and lateral movements, the larynx itself can tilt up toward the epiglottis. Measurement of activity in the vertical dimension is difficult, requiring specialized modifications to the nasendoscope and development of a measurement technique. In spite of all of these drawbacks, the goal of this part of the research is to devise a method of quantitative video measurement without modification to the laryngoscope. Painter's technique is intriguing since it also provides a method of normalization that allows for comparisons across subjects and recording sessions.

Another technique of measuring video pictures is Peppard and Bless' (1990) method of tracing two images that are similar, using landmark structures such as the arytenoid cartilages, ventricular folds and epiglottis. The traced images are taken from two separate sessions under similar conditions, in order to produce images that are nearly identical. This method of measurement is useful for observing a change in nodule size after treatment, for example. However, an assessment of general relative posture is not possible from this technique, nor is it possible to compare postures.

Painter's normalization of measurements involves acquiring a video picture of a deep inhalation from the subject, for each recording session. In each of these inhalation pictures, the length of the glottis is measured and normalized to 23mm; which is considered a reasonable length for an adult male. All pictures that are taken from the particular camera lens position, or session, are then measured and normalized to the glottal length of 23mm. While this allows for valuable comparisons, it is unclear, in Painter's papers, if a number of deep inhalations were measured and tested to determine if measurements for the deep inhalations are consistent. If the measurements for deep inhalations are not consistent, then the normalization technique is not valid. In this thesis, therefore, normalization of measurements is tested by taking a series of inhalations, making measurements, and testing statistically in order to determine if the values are consistent.

In Painter's papers, a deep inhalation is used at the beginning of a recording session for normalization. Since the camera can move during a recording session, inhalation pictures for this thesis were taken as often as possible. Also, the inhalations for this thesis were considered to be the normal breaths a subject takes during a session, rather than a requested exaggerated inhalation. The rationale for taking normal breaths as the benchmark photo is that there should be less variation in a normal speaking breath than for a 'requested' deep inhalation. Also, if a normal breath can be used for normalization, any laryngoscopic video ever made can be normalized, measured, and used for comparison, since every subject takes a breath at some point.

It is easiest, in terms of describing vocal tract behaviour mathematically, to assume the vocal tract does not change over time and has uniform shape. Consequently, the point of greatest constriction of the pharyngeal sphincter during the production of the cardinal pharyngeal consonants is used as the measurement picture for this study. Since the reality is that speech changes over time, it is useful to analyze a series of video frames, which is also done in this thesis.

The analysis of a video sequence attempts to correlate pharyngeal posture with the posture described for the singer's formant. The singer's formant (which is a concentration of energy at around 3kHz) is said to be caused when the opening of the larynx tube is 1/6 the cross-sectional area of the expanded cross-sectional area of the pharynx. One way to correlate the video with the acoustic data is with the Singing Power Ratio (SPR) measurement described by Omori, Kacker, Caroll, Riley and Blaugrund (1996). This measurement takes the greatest harmonic peak between 2 and 4 kHz and the greatest harmonic peak between 0 and 2 kHz to produce a ratio used to describe singing voice quality. This ratio however, is only valuable for fine tuning a singing style by generally increasing energy within a frequency range and doesn't serve as a linguistic distinctive feature.

Past research has provided two methods of measuring the video data for describing cardinal pharyngeal sounds and pharyngeal sphinctering in general. One method is the quantifiable video measurement technique described by Painter and the other is the ratio of pharyngeal openings that are seen in the video data and is inspired by Sundberg's singer's formant investigations.

3.1 Normalization of Inhalation Pictures

In order to determine if normalizing video pictures to a glottal length measurement is a valid method for comparing video pictures, a collection of 19 pictures of inhalations were captured from two recording sessions. The pictures were saved as Targa files and imported into CorelDraw to be measured. Eight points on the inhalation picture that were visible in all pictures were identified and marked. These points were chosen since the author felt certain that in each picture these points were the same physical structures and that these points could be accurately measured. The points are shown in Figure 15.

Figures 15 and 16

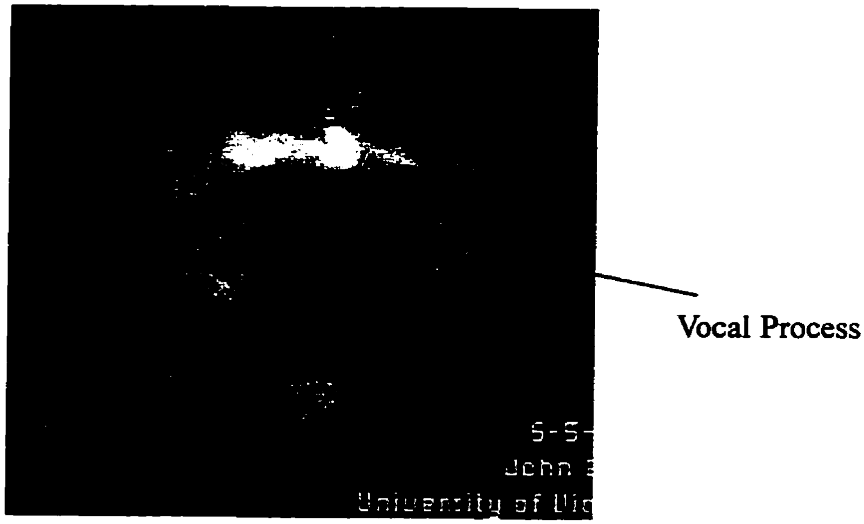


Figure 15: Points used for inhalation measurements.

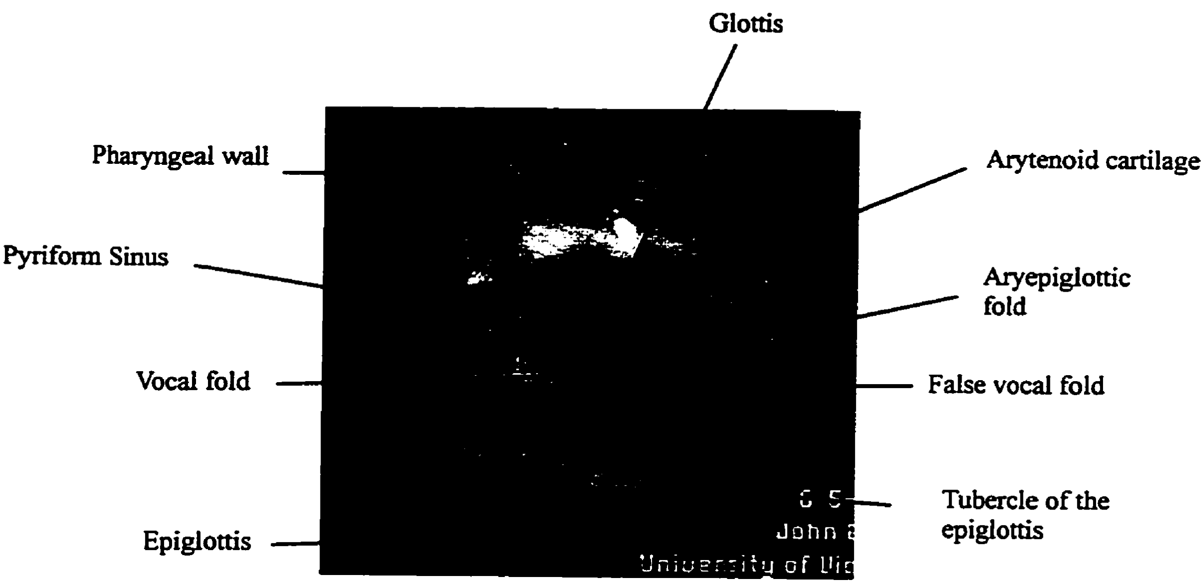


Figure 16: Landmark structures in the larynx and pharynx..

The landmark structures used for measurement appear in two vertical planes. One plane is the 'glottal' plane and the other, is the 'aryepiglottal' plane. Figure 16, shows landmark structures found in these two planes. The glottis, vocal folds and false vocal folds all lie in the glottal plane. The apex of the arytenoid cartilages (formally, but not so often called corniculate cartilages) and the cuneiform cartilages which are embedded in the aryepiglottic folds, along with the aryepiglottic folds themselves, constitute the aryepiglottal plane. Movement for the pharyngeal sounds appears to be in each plane, that is, the epiglottis has the ability to move back toward the pharyngeal wall, making contact with the arytenoid and corniculate cartilages and obscuring the glottal plane below.

The concept of two planes and the value of having points in both planes is important since, most of the pharyngeal articulations measured in this chapter do not have any visible features in the glottal plane. If all measurements that are taken and are normalized to a glottal measurement, then it is necessary to determine if the structures in both planes can be accurately measured; particularly if the glottal opening is rarely seen in the pictures that are measured. The value of using the glottal measurement as the benchmark measurement for normalization and not something in the aryepiglottic plane, is that it can be measured most accurately. All other structures used to take measurements in the video pictures are identified by the brightness of the structure relative to the surrounding tissue. The glottal length, however, is obvious because the glottis is a 'hole', with clearly defined edges.

Referring back to Figure 15, the eight points allow five relatively accurate linear measurements to be taken and one angle measurement. The total number of permutations for the eight points would involve more than five measurements, however, points A, C and D are lines that lie along a fleshy fold, and it would be difficult to find a consistent central point. Points A, C and D, however, do allow for accurate measurement across planes with point B. The one angle measurement is taken from point B to both vocal processes and is consistently 90° in all inhalation examples.

The measurements taken are: AB (glottal length), AD, CB, DB. The raw measurements were taken and then normalized to the equivalent of the glottal length (from A to B). Rather than using Painter's 23mm measurement, the glottal length was considered to be one (1) since, only relative measurements are of interest. Table 1, below shows the normalized measurements between these pertinent points. Also shown are the mean and standard deviations for each measurement.

Table 1. Normalized measurements from inhalation photos. Parameters are on the horizontal axis and photo identification on the vertical axis. Note: Token 'inh5' was removed along with the picture that would have been normalized to that picture since there was not enough lighting to make points measurable.

token	AB	AD	CB	DB	EF
inh1	1	.48	1.22	1.48	2.15
inh2	1	.48	1.16	1.48	2.28
inh3	1	.41	1.18	1.41	1.91
inh4	1	.44	1.13	1.44	2.00
inh6	1	.43	2.05	2.43	2.24
inh7	1	.52	1.10	1.52	2.24
inh8	1	.40	1.15	1.40	2.10
inh9	1	.55	1.09	1.55	2.27
deepa	1	.57	1.11	1.57	2.70
deep1	1	.54	1.27	1.54	2.27
deep2	1	.52	1.31	1.52	2.62
deep3	1	.54	1.32	1.55	2.61
deep4	1	.52	1.32	1.52	2.58
deep5	1	.56	1.15	1.56	2.29
deep6	1	.56	1.41	1.56	2.63
deep7	1	.55	1.37	1.57	2.80
deep8	1	.46	1.23	1.46	2.31
deep9	1	.48	1.31	1.48	2.52
	Mn	0.56	1.22	1.50	2.36
	SD	.09	.11	.06	.25

All values from Table 1 for each parameter fall within two standard deviations of the mean. This suggests that in both the glottal and aryepiglottal planes the measurements do not vary significantly from the mean. This also suggests that normalizing to a standard glottal length is acceptable even for measurements taken from photos that do not show any features in the glottal plane. The results from examining the inhalation pictures show that it is possible to measure structures in a video picture and normalize them so that the values in all pictures effectively come from the same camera depth.

Measurements are then taken from a series of pharyngeal articulations in order to determine if a pharyngeal stop, trill, fricative or approximant can be identified based on a video picture. The picture that is chosen to represent the pharyngeal articulation is the point of maximum constriction during production. This was chosen with the intent that it would give insight into the shape or area function of the resonating cavity for each consonant. The value of gaining information on area function is so that it can be checked against acoustic theory. In this case, the information potentially can be used to create an acoustic theory of pharyngeal behaviour. As seen in chapter two and suggested by Fant (1960) changes in the pharynx result in upper formant changes; the present acoustical theories of speech production focus on the lower frequency formant information, and do not provide a model of upper formant behaviour. Before discussion of the analysis of the still pictures, there will be a subjective auditory and visual description of the sequence that is examined for each pharyngeal sound.

3.2 Observational Description of Cardinal Pharyngeals

The Pharyngeal Stop: The pharyngeal stop sounds like a more forceful version of a glottal stop. The glottal stop is the sound occuring in the middle of the paralinguistic speech feature "uh-oh!". A more forceful version of this is the middle stage of swallowing. In swallowing, the aryepiglottic sphincter closes off the larynx. This posture is protective and allows channeling of food into the esophagus.

The cardinal pharyngeal stop was produced in an [iʡi] environment. This offers the best contrast of movement since the tongue is high in the oral cavity for the production of [i], thus maximally exposing the pharynx.

When viewing a pharyngeal stop with a nasendoscope, the larynx (the structure that contains the vocal folds), rises up to meet the epiglottis (Figure 17a), thus obscuring the larynx. The glottis is obscured by the aryepiglottic folds and the arytenoid cartilages which are pressed against the epiglottis. The arytenoid cartilages tilt forward and meet the epiglottis above the epiglottal tubercle. Thus, the seal for complete constriction is between the larynx tube and the epiglottis rather than the epiglottis and the pharyngeal wall. The pyriform sinuses remain visible, but take on a smaller, more rectangular shape.

The Pharyngeal Trill: A pharyngeal trill sounds like a growl, and is associated with the singing style of the jazz artist Louis Armstrong. This sound can be voiced (involving the vocal folds) or unvoiced. Observation of this behaviour with the nasendoscope shows the pharyngeal structures in the same posture as the other pharyngeal sounds, with the inclusion of trilling that is caused by the aryepiglottic folds vibrating. The ability for these folds to vibrate requires the larynx to approximate the epiglottis, usually in raised mode (Figure 17b). This is similar to the stop, but the aryepiglottic folds are slackened enough to vibrate rapidly against the epiglottis. Vibration is identified in the still photo by the blurring at the aryepiglottic folds.

The Pharyngeal Fricative: A fricative is caused by producing friction noise as in an English /f/ or /s/. In order to cause this type of noise in the pharynx, the arytenoid cartilages remain against the pharyngeal wall with a gap between them which allows the aryepiglottic folds to remain away from the epiglottis (Figure 17c). It appears that the entire larynx tube is elevated in the case of a fricative compared with the tilting action for a plosive or trill.

The Pharyngeal Approximant: The approximant is the Arabic 'ain', which is classed as a fricative but argued to be an approximant by Laufer (1996). The most constricted point of an approximant looks similar to the plosive and trill except that the epiglottis and pharyngeal wall form a less rounded view of the arytenoid cartilages and aryepiglottic folds (Figure 17d).

3.3 Measurements and Discussion

The landmark structures in the larynx that are of interest and used as anchors for the measurements taken are shown in Figure 16. Painter uses 15 measurements that are visible in the data he examined; in this study, there are only four of these measurements that are visible in the cardinal pharyngeal still picture. Of the four measurements, only two are visible in all of the pharyngeal data. The other two measurements are only visible in the pharyngeal fricative and, consequently, serve to distinguish this consonant. A verbal description of each measurement along with the acronym used throughout the chapter are as follows:

1. Glottal Length (GL): This is the measurement of the opening between the vocal folds along the antero-posterior dimension, from the anterior commissure of the vocal folds at the epiglottal tubercle to the vocal processes of the arytenoid cartilages. In the case of the fricative, the glottis is obscured somewhat by the arytenoid cartilages, the aryepiglottic folds and the epiglottis. Since the fricative is the only posture examined here that has a visible glottis, the measurement for this and GW are recorded as one (1).

2. Glottal Width (GW): The glottal width is the widest part of the opening between the vocal folds, which is between the vocal processes. The fricative is the only example that has a visible glottal opening; in the fricative, the vocal processes are obscured so, as with GL, it is given a value of one (1).

3. Epiglottic Blade Position (EBP): The epiglottal blade position is measured from the posterior wall of the pharynx and runs between the arytenoid cartilages, which form a small triangle against the epiglottis in the analyzed pictures, to the blade of the epiglottis. The blade measurement is taken where the tubercle intersects with the blade itself.

4. Pyriform Sinus Width (PSW): This measurement is the maximum distance from the aryepiglottic fold to the pharyngeal wall, perpendicular to the aryepiglottic fold. This measurement is taken from the pharyngeal wall to where the arytenoid cartilage joins the

aryepiglottic fold since, in the pharyngeal articulations, it consistently shows up as a relatively clear right angle.

Method: A fibreoptic nasendoscope was inserted to show a suitable view of the larynx and pharynx as described with respect to Figures 1 and 2. It was possible to maintain a stable camera position while recording by monitoring markings on the nasendoscope that line up with the outside edge of the nose (external nares). Variability in camera position is also accounted for by normalization of values as described above. The video data are obtained from the *vowel-consonant-vowel* , sequence [i_i]. The vowel [i], as in 'beet', is used since the tongue is high in the oral cavity leaving the pharynx wide open and offering the greatest contrast with pharyngeal behaviour. Sequences are spoken in isolation or taken from the carrier phrase "he hit VCV quickly". The data are taken from two different recording sessions using one subject, who is a trained phonetician.

The point of maximum constriction during the consonant production is considered to be the 'pharyngeal articulation'. The images are obtained using an Olympus ENF-P3 flexible fibreoptic laryngoscope attached to the Kay Elemetrics Rhino-Laryngeal Stroboscope 9100, a Panasonic KS152, camera and a Mitsubishi S-VHS video cassette recorder BV-2000 (running at 30 frames/sec). The images are saved as a Targa file and exported to a PC computer for processing in CorelDraw.

Data captured for this study consist of at least eight examples of each pharyngeal consonant articulation (stop, trill, fricative and approximant) and, at least two inhalation samples for each set of four or five consonant examples. Some pictures were discarded; in one case the picture transferred from the Targa file was unreadable and in a second case, the inhalation picture to be used for normalization was too dark to be measured and was discarded along with the articulation picture that followed.

In all cases except for pictures 'stp2t7-stp7t7', the inhalation picture used for normalization was taken before the articulations were uttered. Pictures 'stp2t7-stp7t7', were noticeably closer to the camera than the picture preceding them. However, there was no inhalation prior to these frames. Consequently, the inhalation immediately following these pictures was used for normalization.

The raw data measurements taken from the inhalation pictures, along with the articulation pictures that follow are shown in Table 2. Since the inhalation pictures are only used for normalizing to the glottal length, only the glottal length value is recorded. In the cases where no measurement was possible, the value was recorded as zero.

Table 2. Raw measurement values from video pictures. The parameters are on the horizontal axis, photo identification is on the vertical axis. Zero indicates measurement was not possible. One (1) indicates the glottis is visible but an accurate numerical measurement is not possible.

	GL	GW	EBP	PSW		GL	GW	EBP	PSW
inh1	27				deep7	30			
stp1	0	0	22	11	fric1t7	1	1	38	13
stp2	0	0	20	8	deep8	26			
inh2	25				fric2t7	1	1	37	15
stp3	0	0	19	10	deep9	29			
stp4	0	0	19	9	fric3t7	1	1	42	15
deep1	26				fric4t7	1	1	45	15
stp1t7	0	0	32	14	fric5t7	1	1	47	25
stp2t7	0	0	29	11	fric6t7	1	1	41	20
stp3t7	0	0	34	11	inh3	22			
deep2	29				fric1	1	1	16	12
stp4t7	0	0	36	14	inh4	16			
deepa	27				fric2	1	1	10	0
stp5t7	0	0	43	13	fric3	1	1	10	0
stp6t7	0	0	52	12	inh6	21			
stp7t7	0	0	50	15	aprx1	0	0	21	12
inh8	20				aprx2	0	0	23	13
tril1	0	0	25	16	aprx3	0	0	22	14
tril2	0	0	22	13	aprx4	0	0	22	12
inh9	22				inh7	29			
tril3	0	0	25	15	aprx5	0	0	27	14
tril4	0	0	25	14	deep3	28			
tril5	0	0	29	15	apx1t7	0	0	41	17
deep5	34				apx2t7	0	0	43	20
tril1t7	0	0	40	17	apx3t7	0	0	44	22
tril2t7	0	0	39	19	deep4	31			
deep6	32				apx4t7	0	0	41	21
tril3t7	0	0	42	20	apx5t7	0	0	41	20
tril4t7	0	0	42	20	apx6t7	0	0	43	22
tril5t7	0	0	42	12	apx7t7	0	0	40	18
tril6t7	0	0	39	9					

The values in Table 2 are then normalized to the inhalation pictures and shown in Table 3. The inhalation tokens have been removed from the table.

Table 3. Normalized values from video pictures. The parameters are on the horizontal axis, photo identification is on the vertical axis and normalized values take from measurements are in the cells.

	GL	GW	EBP	PSW		GL	GW	EBP	PSW
stp1	0	0	.82	.41	fric1t7	1	1	1.27	.43
stp2	0	0	.74	.30	fric2t7	1	1	1.42	.58
stp3	0	0	.76	.4	fric3t7	1	1	1.45	.52
stp4	0	0	.76	.36	fric4t7	1	1	1.55	.52
stp1t7	0	0	1.23	.54	fric5t7	1	1	1.62	.86
stp2t7	0	0	1.11	.42	fric6t7	1	1	1.41	.69
stp3t7	0	0	1.31	.42	fric1	1	1	.73	.55
stp4t7	0	0	1.24	.48	fric2	1	1	.63	0
stp5t7	0	0	1.59	.48	fric3	1	1	.63	0
stp6t7	0	0	1.93	.44	aprx1	0	0	1	.57
stp7t7	0	0	1.85	.56	aprx2	0	0	1.09	.62
tril1	0	0	1.25	.8	aprx3	0	0	1.05	.67
tril2	0	0	1.1	.65	aprx4	0	0	1.05	.57
tril3	0	0	1.14	.68	aprx5	0	0	.93	.48
tril4	0	0	1.14	.64	aprx1t	0	0	1.46	.61
tril5	0	0	1.32	.68	aprx2t	0	0	1.54	.71
tril1t7	0	0	1.18	.5	aprx3t	0	0	1.57	.79
tril2t7	0	0	1.15	.56	aprx4t	0	0	1.32	.68
tril3t7	0	0	1.31	.63	aprx5t	0	0	1.32	.65
tril4t7	0	0	1.31	.63	aprx6t	0	0	1.39	.71
tril5t7	0	0	1.31	.38	aprx7t	0	0	1.29	.58
tril6t7	0	0	1.22	.28					

The means, standard deviations and percent variation for each of the pharyngeal articulations are calculated and shown in Table 4. The GL and GW measurements are binary, that is, for all pictures except the fricatives there is no measurement possible. Consequently, these measurements distinguish the fricative and are not shown in Table 4.

Table 4. Statistics from normalized values.

stats	EBP	PSW	
stop	mean	1.21	.44
	SD	.43	.08
	variation	36%	18%
aprx	mean	1.25	.64
	SD	.22	.08
	variation	18%	13%
trill	mean	1.22	.59
	SD	.08	.15
	variation	7%	25%
fric	mean	1.19	.59
	SD	.41	.14
	variation	35%	24%

The mean values for both EBP and PSW measurements, between consonants, are similar, suggesting that these measurements alone do not distinguish the four manner-of-articulation consonants. The only marked values appear to be the percent of variation. For the epiglottal blade position (EBP), the stop and fricative articulations vary considerably more than the approximant or trill. This is reasonable since the required posture to maintain trilling, and avoid friction (in the approximant articulation), would have to remain stable. The pyriform sinus width (PSW) should be more variable for trilling since the aryepiglottic folds trill into the pyriform sinus space and in fact the values for the trill vary more than for the stop and approximant.

Discussion: These video measurements are consistent and suggest that normalizing of all measurements to the glottal measurement seen in one of the depths is appropriate. The measurement used to normalize test data in Painter (1986, 1991), Painter et al. (1991), and in this chapter is the glottal length. Although the glottis is not seen in any of the test data for this experiment, the glottal length is used from inhalations because it is the most consistent and accurate measurement. The first part of this chapter tests the validity of normalizing a test photo, to a glottal length. This is a valid method which allows for comparison between different photos, sessions and even speakers. Measurements that are used to test normalization are taken from points which appear at two different depths in the pharynx.

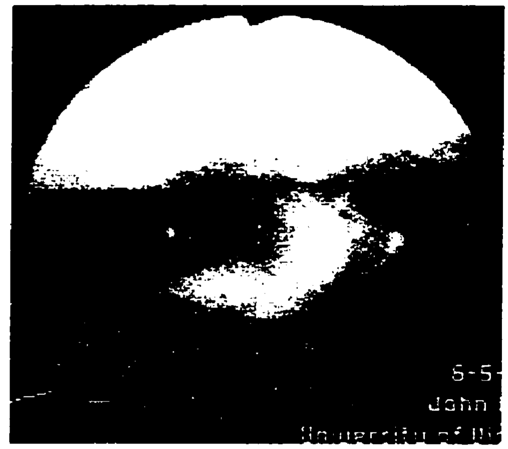
The normal breath a subject takes during speech is used for the normalization photo in this thesis because this act occurs often during speech and shows a maximally open glottis. Taking glottal length measurements from many inhalation photos during a session, and monitoring the camera depth externally, converts each test picture's measurements into the same virtual depth.

Since the pharyngeal sounds in this experiment are proposed to be, according to the IPA taxonomy, varying degrees of constriction, one might expect this to be visible in the pictures. Glancing at the photos in Figure 17, the stop, trill and approximant 'look' very similar. Differences that are visible between these still photos include: The stop appears to have a more rounded closure between the epiglottis and pharyngeal wall compared with the trill and approximant. Also, the trill has a 'blurred' aryepiglottic fold that is caused by its vibration. The fricative is the most markedly different since the glottis is visible through an interarytenoidal gap at the aryepiglottal level. Statistical testing of distances in the antero-posterior plane (from the pharyngeal wall to the epiglottis), do not suggest that the stop, trill and approximant pharyngeal postures, are different.

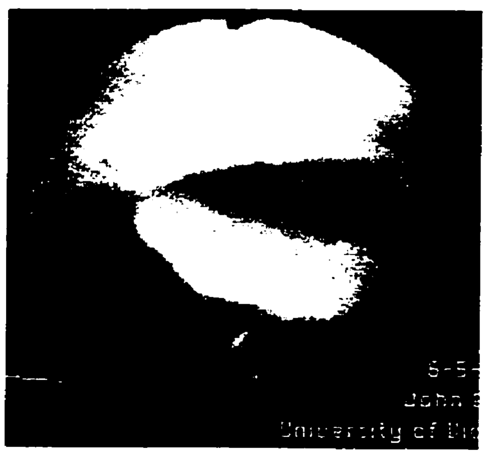
Instead of varying degrees of constriction in the antero-posterior dimension, the pharyngeal stop, trill, fricative and approximant appear to make use of the larynx position below the aryepiglottal constriction because in the case of the approximant the view of the glottis is obscured, yet quite normal voicing is heard. In the case of the pharyngeal stop, there may be complete closure between the larynx and the epiglottis along with full glottal closure. For the trill and approximant, the percent of variation in the EBP suggests that these articulations rely on precise positioning of structures in the pharynx. In order to maintain trilling, the aryepiglottic folds must be slack enough to vibrate and there must be adequate pressure from the lungs to set these folds vibrating. The approximant, requires a posture that allows air to pass through the narrow pharyngeal constriction and avoid causing friction.

A task for future research is to observe the larynx below the pharyngeal constriction described in this thesis. With the tools available to the author, it is impossible to see what the larynx is physically doing below the pharyngeal constriction: X-ray, MRI or perhaps ultra-sound measurements along with video and acoustic measurements are necessary in order to observe behaviour below the pharyngeal sphincter.

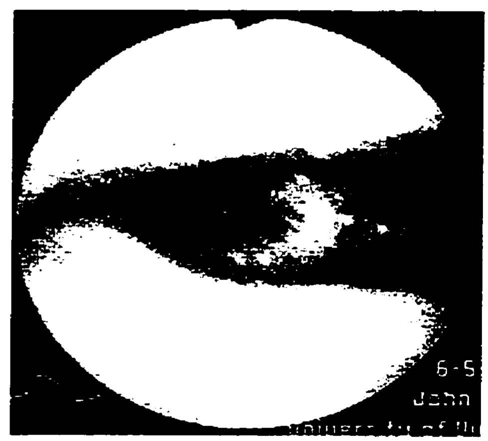
Figure 3a. Stop



b. Trill



c. Fricative



d. Approximant

Figure 17: Sample picture of each pharyngeal articulation.

The phonetic segments examined here are better defined by movement in the time dimension and by positioning of structures that are out of view. Measurement of one still photo is more valuable for describing general, pharyngeal and laryngeal postures that contribute to voice quality settings, rather than phonetic segments. The use of this video measurement in describing voice quality settings is the focus of Painter's papers.

The investigation of pharyngeal constriction in this thesis also provides support for the pharyngealized voice quality setting. That is, a long term vocal tract posture that maintains constriction in the pharynx is possible since, as we have seen with the pharyngeal approximant and trill, the epiglottis and arytenoid cartilages can be approximated while the vocal folds below vibrate. Also, we will see evidence that the Hebrew speaker is section 3.4 possesses this voice quality setting and consequently, his video pictures are difficult to analyze.

Since speech varies over time, one still photo does not describe how long it takes to produce a phonetic segment, nor does it describe the way structures move to get to the most constricted point. A brief examination of duration, based on video frames showed that on average, the stop takes the longest to produce, being anywhere from 9-19 frames from the vowel through to the following vowel. The approximant and trill, averaged about 6 frames and the fricative 4 frames. However, all of these consonants save the pharyngeal plosive can be sustained, therefore, the number of frames is not very useful. More information about the production of phonetic segments may be gained from examining measurements from a sequence of frames rather than a picture of the most constricted point (see section 3.4). However, observing activity below the constriction is really necessary to complete the theory.

The parameter that is missing from the test pictures examined here is depth of the larynx before, during and after the pharyngeal constriction. This information would provide evidence for raised and lowered larynx positions that are presumed to occur beneath the pharyngeal sphincter. Larynx height is valuable for the investigation of voice quality since Laver's voice quality taxonomy has terms for raised and lowered larynx settings. Also, there is a correlation between raised larynx and constriction in the pharynx as discussed earlier with respect to Esling et al. (1994). Namely, Laver's raised larynx voice quality setting involves the same posture as a pharyngealized setting, and the difference between the two settings is pitch related. That is, a raised larynx setting would look similar to the pharyngeal postures examined here, except perceived pitch would be relatively high. Pharyngeal constriction with lower pitch is perceived as pharyngealized voice quality.

Titze and Story (1996) suggest that the space created between the larynx and pharyngeal constriction matches impedance between the glottis and vocal tract, thus causing a drop in fundamental frequency of the vocal folds. This is discussed in chapter 4 and this effect would also contribute, along with resonant frequency of the vocal tract, to the pitch relationship between raised larynx and pharyngealized voice quality settings. Of relevance to the phonetic sounds that have been the focus of this thesis, we see in chapter 5 that the terms raised and lowered larynx, that can be long-term voice quality postures, may function as distinguishing segmental features in languages such as Bruu and Mpi and even Native North American languages.

The results of measuring still photos imply that an analysis of a sequence of video pictures is necessary in order to describe the way the pharyngeal consonants are produced. A sequence of video pictures for one utterance are examined in the following section.

3.4 Measurement and Movement

The purpose of this section is to map a sequence of video pictures with acoustic data in order to give more insight as to the acoustic effect of pharyngeal movement. This is accomplished by capturing a series of video pictures and the corresponding acoustic output. The video captures movement at 30 frames per second. With this fact the researcher can gain a better estimate of where the video picture corresponds with the acoustic data. The data used here are the sequence [iʕi]. The sequence of pictures was captured using the Kay Elemetrics Rhino-Laryngeal Stroboscope 9100 described earlier, and the acoustic output corresponding to the video sequence was captured direct from the video recorder into a Sony model DTC-750 DAT recorder. The video pictures were then transferred to CorelDraw and the audio was captured into the MultiSpeech 3700 software package. Along with the sequence of video pictures an inhalation picture was taken before and after the sequence just to ensure that the scope has not moved markedly.

Analysis of the video pictures was designed to determine if the laryngeal opening was less than (<), equal to (=), or greater than (>) 1/6 the opening of the pharynx. This value of 1/6 has been described in Sundberg (1987) and Titze (1993), but in general, when the laryngeal opening is less than one sixth the opening of the pharynx opening the result is a tube that possesses its own resonant frequency as explained in section 2.5. This ratio is easily measurable from the video pictures and thus, should correlate with changes in high frequency energy.

In order to determine the ratio of laryngeal opening to pharyngeal opening, a clear plastic grid was placed over the picture. The laryngeal opening was measured along the inside edge of the epiglottis, aryepiglottic folds and arytenoid cartilages. The pharyngeal measurement included the inside edge of the epiglottis and coursed out to the pyriform sinus, crossing the attachment of the aryepiglottic fold to the epiglottis. Figure 17 shows the first four pictures that were measured.

There were 19 frames captured that correlated with the second example in a series of productions of [iʕi] spoken in the sentence "he hit [iʕi] quickly". The [iʕi] sequence was isolated from the captured audio data and measured to be about 1 second. The 19 video frames captured correspond to 0.63 seconds of time since only a few frames of the vowel surrounding the consonant were included and not the entire [iʕi] sequence. In the sequence of pictures, there is a marked change in posture from frames 3 to 4 where the ratio becomes

<1/6. This change is sustained for about 9 frames, or 0.3 sec which corresponds with the marked change in formant structure in the spectrogram of the audio signal in Figure 20.

It is possible then, to correlate the formant behaviour with the video pictures within a few milliseconds. This is because the sounds in the sequence are extremes and, consequently, the formant behaviour can cue the researcher as to where the frame correlate in the waveform. Specifically, the energy from 0-700Hz is expected to be low for the vowel [i] and energy between 1500 and 3000Hz is expected to be high. This energy, which is composed of the first three formants, merges together as the sound goes to an [ɑ] like articulation. Thus, the changes in oral behaviour are accounted for by the first three formants leaving the upper formants to correlate with the pharyngeal and laryngeal ratio, just as Fant (1960) and Sundberg (1987) suggest.

Comparing Figure 17 with the spectrogram in Figure 19; the first 3 videos correlate with the beginning of the marked portion. Frame 4 has a ratio of < 1/6 and correlates with a marked drop in energy starting at about 2500Hz compared with the preceding frames. All the frames that follow frame 4 have a ratio of < 1/6 and their behaviour is the same. Unfortunately, the audio data from the laryngoscope only captures at 10kHz; consequently only formant information up to 5kHz is available. From discussions in section 2.5, it is necessary to have a good view of the data above 4000Hz in order to try and describe formant behaviour for pharyngeal articulations. This closer examination of the video and acoustic data further suggests that the laryngeal/pharyngeal ratio has a marked effect on the upper formants.

Interestingly, data from a native Hebrew speaker was matched to the acoustic output in the same way as for the phonetician. However, this speaker's ratio of <1/6 occurred for far more frames than the entire acoustic sequence. Two different video and audio sequences were captured for this speaker and both did not have a direct correlation between laryngeal/pharyngeal opening and acoustic output. Examining the video of the Hebrew speaker, it appears that his voice quality is in a pharyngeally constricted position

Figure 18: The first four frames in the video sequence containing [i] (the first 3 frames and the pharyngeal approximant (last frame).

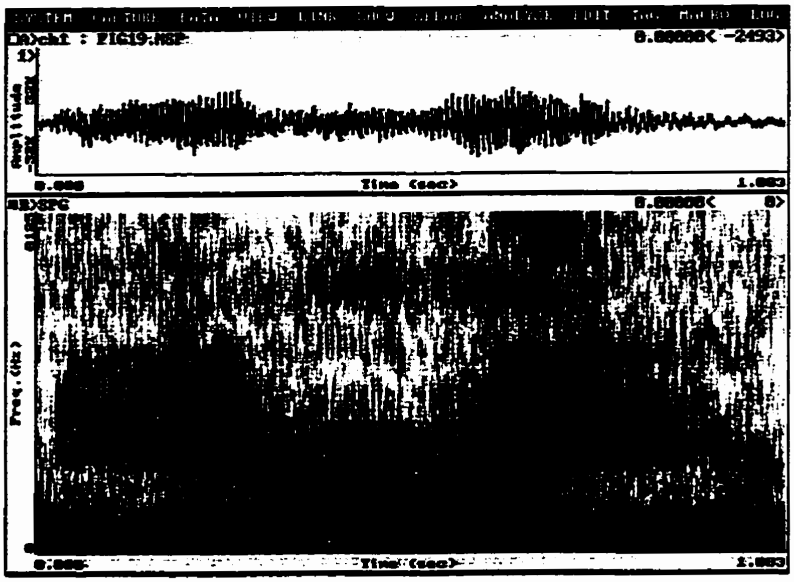


Figure 19: Formant behaviour from the utterance that corresponds with the video pictures in Figure 18. The sharp change in formant structure is about 0.3sec.

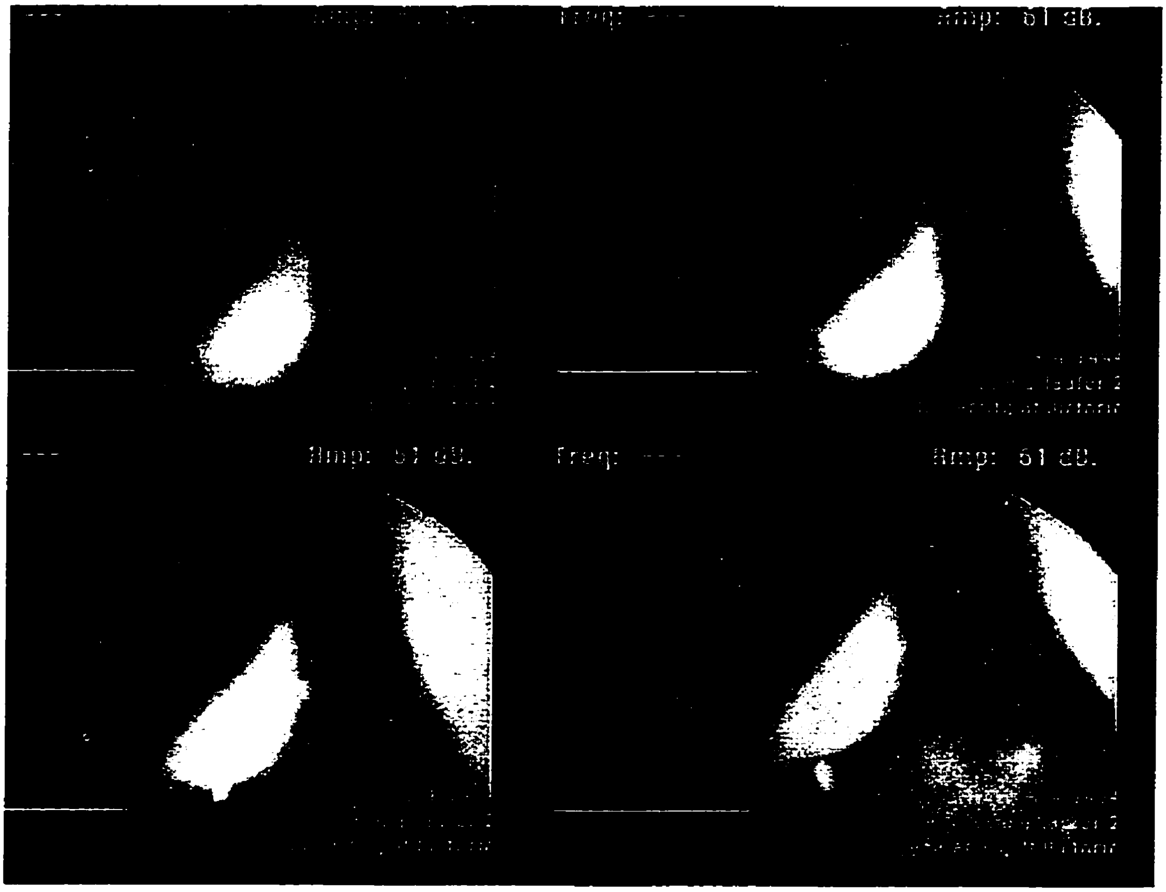


Figure 20: This Hebrew speaker's larynx is lower as the epiglottis moves back compared with Figure 18.

long before and after the utterance. Therefore, the 1/6 ratio can be explained as the product of a long-term setting rather than a brief segmental articulation.

Figure 20 shows the first four frames of the Hebrew speaker's data. These frames have the least amount of constriction and appear to have occurred before the utterance was spoken. The pictures show this speaker's larynx is low as the epiglottis moves back toward the pharyngeal wall. Since the pictures had less than 1/6 as a ratio, and Hebrew has a natural class of post-velar sounds, it is assumed from the video and acoustical data, that this speaker has a 'pharyngeal' voice quality setting that is present more or less all the time. In order to produce the pharyngeal consonant, this speaker must raise his larynx up to the already backed epiglottis. The phonetician, on the other hand, has his larynx and epiglottis meet in the middle.

Using the 1/6 ratio as a measurement does not necessarily mean that there will be a strong peak in the upper frequencies as Sundberg suggests; the posture that creates this peak is likely more complex. As we have seen, it is possible to raise and lower the larynx below the pharyngeal sphincter and it is hypothesized that a singer's formant requires a balance between the laryngeal to pharyngeal opening and larynx height below.

3.5 Conclusion

This chapter shows that it is possible to normalize measurements taken from video pictures to the glottal length of an inhalation picture taken during the recording session. This procedure converts all measurements so that all the test data pictures come from the same virtual camera depth, and thus, the pictures can be compared. A further examination of the normalization technique could test to see if a requested deep inhalation, like the one Painter used, is significantly different from the normal breaths used here.

Measurement of a picture that represents the most constricted portion of a pharyngeal consonant only shows distinguishing features in the fricative manner of articulation. Consequently, it is necessary to consider other things, such as what is happening at the larynx while the pharyngeal sphincter engages.

It appears that the manner of articulation for the trill, approximant and stop may be created below the pharyngeal constriction as is evidenced by the view of the glottis during the fricative and the approximant which has audible voicing with no visible glottis in the video data. Also, the Hebrew speaker has a backed epiglottis most of the time during his speech and must use the structures below. It is hypothesized and suggested by this data that the larynx can rise and lower while there is pharyngeal constriction, and it is this ability that identifies the varying manners of pharyngeal articulation produced here. The only way to test this hypothesis is to observe the larynx with some other tool along with nasendoscopic video and acoustic output.

The attempt to correlate the video data with acoustic output appears to be possible with the phonetician's data, particularly because the sequence that is examined includes two extreme postures. This allows the number of video frames to be mapped with the formant behaviour. The mapping showed that the first three formants account for the change in oral cavity size. Unfortunately, the mapping made in this experiment does not help with describing the effect of pharyngeal sphinctering on upper formant behaviour because the audio that is taken from the video tape is captured at 10 kHz. This sampling rate only allows energy up to 5000Hz to be visible. It would be necessary in a future examination of this nature to capture the audio data that corresponds to the video onto DAT for later analysis.

Chapter 4 - Effects of Aryepiglottic Sphinctering on Vocal Fold Vibration

4.0 Introduction

The purpose of this chapter is to determine the effect that constricting the aryepiglottic sphincter has on vocal fold vibration. Motivation for this examination comes from Titze and Story (1996) who suggest that pharyngeal constriction, produced by constriction of the aryepiglottic folds, alters vocal fold behaviour. That is, narrowing in the pharynx provides an area that matches the high impedance of the vocal folds with the low impedance of the vocal tract. Matching impedance should allow the vocal folds to vibrate more slowly by reducing the oscillation pressure threshold, and this will be apparent in measurements of vocal fold behaviour.

The most direct vocal fold measurement technique is via the electroglottograph (EGG). Various measurements can be derived from the EGG signal. The most common measurements are quotients of openness and closedness. These quotients measure the relative amout of time the vocal folds are open or closed. The quotient measurements were examined preliminarily on the pharyngeal sounds used in this thesis and did not show significant differences between the pharyngeal sounds. The voiced pharyngeal consonant values for open and closed quotients remained stable and from this, it wasn't expected that other permutations of quotients would be valuable. However, pitch, jitter (pitch fluctuation), shimmer (amplitude fluctuation), and signal harmonic-to-noise ratio (HNR) showed promise for describing vocal fold behaviour. It so happens that these measurements can also be derived from the acoustic waveform, making these measurements more appealing to examine than quotients.

Jitter and shimmer are interperiodic variations in pitch and amplitude respectively and harmonic-signal-to-noise (HNR) ratio reports additive noise that is in the acoustic signal (Bough, Heuer, Sataloff, Hills and Carter, 1996). Consequently, these measurements are generally used in identifying breathy or harsh voice qualities; however, the use of jitter, shimmer and HNR measurements with respect to cardinal consonant sounds are able to describe fricative-like behaviour with respect to a segmental distinction rather than a long-term voice quality effect.

As mentioned earlier, aspects of pharyngeal constriction have been observed in the literature with respect to singing qualities. Titze and Story (1996), and Yanagisawa, Estill, Kmucha and Leder (1989) have observed constriction in the region of the pharynx that they believe is a desirable posture for good singing quality. Titze and Story (1996) in particular believe aryepiglottic constriction, as described in Yanagisawa et al. (1989), enhances vocal fold oscillation by reducing the amount of air required from the lungs. Less air from the lungs for driving the vocal folds implies a more efficient system.

Using mathematical modelling to predict the effect of constriction in the pharynx on the vocal folds, Titze and Story (1996) concluded that "a narrow epilarynx acts a bit like the mouthpiece of a brass instrument, matching the high internal impedance of the glottis to the lower impedance of the vocal tract and free space" (p. 33). The observations of pitch changes between the pharyngeal approximant and [ɑ] vowel that are made in chapter two motivate examining vocal fold vibration more closely. Recent findings from other researchers also suggest that changes in the vocal tract area function in the pharynx will have an effect on vocal fold vibration. The measurements used for this part of the study are chosen because they can be applied to the greatest amount of data and show the most potential for accurately describing the effect of constriction of vocal fold vibration.

4.1 Method and Measurement

The data for this chapter are examples the author collected on DAT of Esling's cardinal pharyngeal consonants and included examples of the pharyngeal sounds where raised and lowered voice quality settings were superimposed onto them. The consonants were produced in either an [ɑ] or [i] environment. Some examples were also produced with a target pitch that was played before the speaker produced the sound sequences.

The language samples used were obtained from two databases, the University of Victoria Phonetic Database (PDB) and the UCLA Sounds of the World's Languages (SOWL). The language data include speech segments from Ahousaht, Arabic, Agul and Hebrew and involves a comparison of voiced pharyngeal sounds and preceding or following vowels. All examples are voiced, since the algorithms used to measure the electroglottographic and acoustic signals depend on voicing. There are statistical measures that may be useful in examining voiceless consonants in particular, called Spectral Moments (Forrest et al., 1988), however, this will be left for another investigation.

The audio and electroglottographic (EGG) data were recorded onto a DAT using a Sony model DTC-750 DAT recorder. All of the data were captured at 20 KHz and saved using the Computerized Speech Lab (CSL). The EGG data were initially processed using a laryngographic signal analysis program (EGG) from Kay Elemetrics Corp, however, the results are the same as the voicing analysis routine in CSL, consequently, the voicing analysis was used for these data.

CSL's voicing analysis routine measures jitter, shimmer, average fundamental frequency and HNR. Jitter is the term used for frequency perturbation and "is the ratio of the short-term average pitch period durations to the momentary pitch period duration." (Visi-PitchII Instruction Manual p. 50). Jitter is measured in % and is the average period difference between consecutive cycles divided by the average period, times 100 (Horii, 1982). The normal range for jitter is considered to be around 0.1-0.5% according to Horii (1982). Shimmer is the average decibel difference between peak amplitudes of consecutive cycles, and the normal range is considered to be from 0.17dB to 0.39 dB (Horii, 1982). Harmonics-to-noise ratio is also a decibel measure that "is the ratio of harmonically related energy to noise in the speech signal" (Visi-PitchII Instruction Manual p. 50). An ideal HNR should have high acoustic energy in the harmonic components and low in noise in between; the average is 11.5 dB to 15.12 dB (Horii, 1982).

The settings for the voicing analysis were adjusted to account for a large variation in fundamental frequency that is seen within the data, and required the tolerance parameter threshold in the voicing analysis to be set to 20 msec. The offset parameter was adjusted to the 20K sampling rate and was set to the lowest default value of 2 (CSL Instruction Manual).

The voicing analysis results were analyzed statistically using NPSTAT, a software package distributed at the University of Victoria. Since the collection of language data is relatively small, and almost half of the examples are from one subject, the non-parametric Wilcoxon signed-rank test is used. This test does not assume that the data are from a normal distribution. Also, the numbers are ranked and then compared rather than using the absolute value of the measurement. The statistical analysis compares audio and laryngographic data on various vowels with the neighbouring pharyngeal consonant.

Included in this comparison are isolated examples of vowels and pharyngeal consonants provided by the phonetician. The vowels [i] and [ɑ], produced by the phonetician are compared and analyzed using the same procedures as with the pharyngeal sounds and their neighbouring vowels. This is done since the vowel [ɑ] is considered pharyngeal. Petersen and Barney (1952) observe that for high vowels such as [i] the pitch is generally higher compared with low vowels such as [ɑ]. Consequently, it can be expected based on past research on vowels that there would be variation in pitch and, because [ɑ] is considered a pharyngeally articulated vowel, the pharyngeal consonant is likely to show similar characteristics to [ɑ]. This analysis of vowels gives a baseline of what to expect when comparing vowels and neigbouring consonants for cardinal pharyngeal sounds and natural language data.

4.2 Results

Although the EGG software was not used for statistical analysis, it did provide a useful visual comparison of the vowel and consonant waveforms. The Electroglottographic waveform for both the phonetician's pharyngeal trill and the approximant are somewhat disturbed compared with the vowel example; Figure 21 illustrates the waveform differences between the vowel (top waveform) and the approximant consonant (bottom waveform):

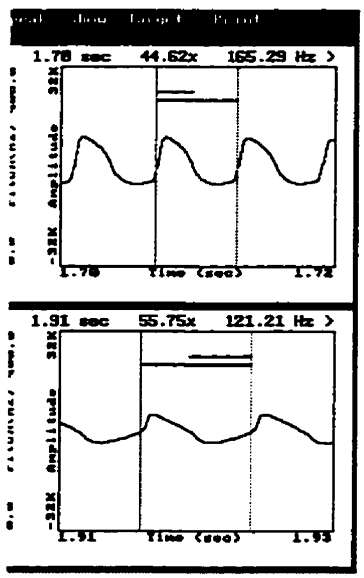


Figure 21. Laryngographic illustration of a pharyngeal approximant. The top waveform shows vowel production [ɑ] and the bottom waveform shows approximant consonant production [ʕ].

The waveform in the bottom of the figure is more 'disturbed' than the smooth looking vowel waveform at the top. This less than smooth nature of the waveform will be described by the CSL voicing analysis performed on the data in this chapter. The lowered amplitude could be from the narrow-walled pharyngeal constriction that is described in the previous chapters, however, it could also be caused by the rising larynx which would move the EGG up consequently, losing the signal.

In order to establish a baseline of what to expect with the pharyngeal sounds, the vowels [i] and [ɑ] are compared. Table 5 presents an analysis summary of the phonetician's production of [i] and [ɑ]. The target pitch for the isolated vowels is 110Hz; this target pitch was reached by playing a reference tone for the speaker. The values obtained for the vowels are as follows:

Table 5. *Values for pitch, jitter, shimmer and harmonic signal-to-noise ratio for* [i] *and* [ɑ].

	mean pitch	jitter	shimmer	HNR
[i]-112 (frames)	112.193	0.692	0.378	11.455
[a]-92	109.943	0.981	0.561	5.926
[i]-110	112.068	0.747	0.272	10.340
[a]-80	110.793	0.907	0.515	3.251
[i]-Lx,113	112.187	0.707	0.083	6.334
[a]-Lx, 106	110.044	0.721	0.120	5.693
[i]-Lx, 113	111.931	0.731	0.086	8.337
[a]-Lx, 79	110.397	0.674	0.111	5.519
[i]-Rlx, 160	113.151	0.781	0.392	9.536
[a]-Rlx, 112	112.793	6.920	0.686	0.388
[i]-Rlx, Lx, 160	113.107	0.281	0.134	6.263
[a]-Rlx, Lx, 111	111.815	0.307	0.133	8.512

Note: Normal range for jitter is: .5-.1%, shimmer: .17-.39 dB, signal-to-noise: 11.5 dB to 15.12dB. Rlx refers to raised larynx setting, Lx refers to the EGG signal itself is used for analysis.

The reason that values for the vowel data in Table 5 vary outside the normal range for jitter is unknown. The author was careful to make sure the impulse marks on the waveform were accurately placed, therefore, the results should be reliable. Only the raised larynx examples fall within the normal range suggested by Horii (1982). The values for HNR are quite low also. Only three examples fall outside the norm for shimmer, all are the low vowel examples. The numbers for jitter and HNR imply that there was poor recording quality, however, the shimmer and pitch values appear normal; particularly the pitch values since a reference tone was played for the speaker. The variation in pitch for the vowels is minimal and can be accounted for by Titze and Story's (1996) discussion.

Statistical analysis to compare pitch between the two vowels reveals that, for this sample, the probability that it is by chance that pitch is higher for the vowel [i] is 0.00098. This means, for these isolated vowels, that pitch for [i] is consistently and significantly higher than [ɑ]. The same test for jitter and shimmer values suggests that [ɑ] has a consistently higher value than [i]. And, for the HNR, [i] has a consistently higher value.

The vowel analysis is consistent with the literature with respect to the pitch being higher in high vowels (Petersen and Barney 1952), and suggests that findings should be the same in the comparison between a vowel and its neighbouring pharyngeal consonant. Values for pitch, jitter, shimmer and HNR obtained from the vowel versus pharyngeal consonant comparison are summarized in Table 6:

Table 6. *Values for pitch, jitter, shimmer and harmonic signal-to-noise ratio for vowel/consonant comparison.*

N=29	*Mean Fo*	*Jitter %*	*Shimmer dB*	*Sig/noise dB*	*N=29*	*Mean Fo*	*Jitter %*	*Shimmer dB*	*Sig/noise dB*
aho123CCC1	157.846	5.718	0.512	4.212	**jheV**	166.337	1.138	0.241	8.827
aho123VV17	235.131	6.427	1.523	4.155	**jheC**	147.007	1.883	1.305	-5.487
aho123C	177.966	1.592	0.462	-1.348	**jheV**	155.096	1.986	0.324	7.133
aho123V	217.722	1.823	0.297	11.13	**jheC**	140.888	2.438	0.835	-4.97
are023V	159.635	1.33	0.691	1.386	**jheV**	169.731	0.9	0.255	5.33
are023C	104.428	7.28	1.781	-6.247	**jheC**	141.176	0.674	0.403	-6.301
are024V	151.829	9.666	0.743	-3.316	**jheV**	155.642	0.666	0.291	3.505
are024C	136.876	9.277	2.572	-5.884	**jheC**	132.653	1.114	0.475	-7.341
are029C	126.404	1.527	0.657	-4.731	**jheTRILL**				
are029V	152.284	1.65	0.604	-0.127	**jheV Lx**	169.247	4.123	0.379	-5.594
are030C	131.944	7.787	1.347	-2.773	**jhe-CLx**	123.851	11.832	1.681	-4.676
are030V	202.109	1.508	1.047	7.427	**jhe-VLx**	184.12	0.078	0.395	3.391
are033V	189.394	0.728	0.717	2.812	**jhe-CLx**	111.554	5.575	1.565	-5.551
are033C	142.315	7.919	1.962	-5.034	**jheTRILL**				
are062V	195.865	2.367	0.741	4.261	**jhe-V**	165.041	0.55	0.318	7.858
are062C	108.696	5.707	3.249	-2.682	**jhe-C**	105.152	20.442	2.948	-5.843
muʕagulV31	269.331	1.95	0.173	0.32	**jhe-V**	169.95	1.048	0.235	6.111
muʕagulC35	153.106	3.433	1.112	-3.588	**jhe-C**	120.267	11.538	2.966	-5.321
muʕarV	206.022	0.63	0.401	5.317	**jhe-V**	163.569	1.227	0.242	7.978
muʕarC	130.344	6.404	1.655	-0.765	**jhe-C**	117.904	25.059	4.512	-5.303
muʕarV	195.369	4.246	0.624	-0.38	**Lx /l/**	168.7	1.37	0.182	5.526
ʕor-C	95.256	3.591	1.795	1.139	**Lx aprx**	137.104	3.853	0.487	-8.753
ʕor-V	102.214	1.59	0.432	2.658	**Lx /l/**	167.214	3.966	0.316	9.411
naʕar-V	83.128	3.128	0.793	-0.952	**Lx trill**	121.112	16.683	2.394	-4.99
naʕar-C	79.424	5.938	1.028	-1.701	**Audio /l/**	165.505	1.152	0.242	8.610
naʕar-V	104.05	0.576	0.86	3.934	**Audio aprx**	152.815	1.731	1.128	-5.223
jaʕimV	107.116	3.749	1.209	3.689	**Audio /l/**	162.602	1.174	0.245	7.868
jaʕimC	75.758	12.463	3.073	-4.409	**Audio trill**	126.08	18.301	4.842	-4.949
jaʕimV	218.216	2.944	1.148	7.982					

Table Legend: jhe=phonetician, Lx refers to the EGG waveform, Rlx=raised larynx setting, aho=Ahousaht language, are=Arabic language, the remaining are words from Agul (Caucasian).

The language data voicing analysis values that occur outside the established norms outlined by Horii (1982) can be accounted for because the measurements are performed on less than ideal recordings, particularly with the SOWL data which will be discussed later.

However, the vowel analysis above shows a varied distribution, and the data came from an excellent recording. As suggested by Gelfer and Fendel (1995), the measurement of jitter, shimmer and HNR are unstable and require direct input into the analysis software for the most accurate analysis. Also, Karnell et al. (1995) show that there is variation among voicing analysis between different electroglottographic analysis systems. However, in most of the data in this chapter the measurements are compared with a neighbouring segment, and consequently the measurements are relative and do not depend on the absolute value. Nor does the statistical analysis depend on the absolute value of the measurements.

An initial analysis shows that for all 29 examples pitch is lower for the pharyngeal consonant compared with the vowel preceding or following. For jitter, 24 out of 29 pharyngeal consonants have a higher value than for vowels. Shimmer shows 28 out of 29 consonant examples having a higher value, and for the harmonic signal-to-noise ratio 27 out of 29 vowel examples had a higher value. In order to investigate the statistical significance of these values, the Wilcoxon signed-rank test was applied and confirmed the results are statistically significant. Similar to the results for the analysis of [i] and [ɑ],the pitch of vowels compared with pharyngeal consonants is consistently higher. Jitter and shimmer have consistently higher values for the consonants. And vowels have a higher value for harmonic signal-to-noise ratio than neighbouring pharyngeal consonants.

4.3 Discussion

A preliminary examination of the EGG waveform data, in order to determine pertinent measures to be examined on a larger corpus of data revealed that the EGG waveform values for quotients were not significant. Since the quotient values are not significant in the EGG signal and, it is possible to acquire pitch and jitter from CSL, both the EGG signal and the audio signal are analyzed in CSL. This allows the inclusion of shimmer and HNR measures. Since only a limited amount of electroglottographic data is available, it is valuable to have the capability of relying mainly on acoustic data from the various languages in the analysis, thus providing more speakers and examples for analysis and comparison.

One visible quality of the EGG examination seen in Figure 21 is that the waveform for the pharyngeal consonant is less smooth and lower in amplitude. This implies traditionally that harshness would be a perceived quality, however, based on Omori, Kojima, Kakani, Slavit and Blaugrund (1997), this may not be the case. Their study shows how these two parameters can be eliminated, yet roughness is perceived based on sub-harmonic frequency. These results are promising if aryepiglottic constriction is considered desirable for good singing, since the increase in jitter and shimmer observed here would not necessarily imply a harsh or undesirable quality. It is possible that the extreme pharyngeal articulations examined here introduce more disruption to vocal fold vibration than good singing and, thus causing more extreme jitter and shimmer values; however, this does not necessarily mean that these articulations involve harshness based on Omori et al. Also, the great excursion seen from the normal jitter and HNR values established by Horii make the use of these measurements questionable as descriptors of salient features.

The most consistent result from the analysis presented here is that pitch decreases during the production of postures that involve constriction in the pharynx. The values for pitch, with respect to vowels, is predicted and supports the findings of Peterson and Barney (1952) whose work suggests that high vowels are generally higher in pitch. Considering Titze and Story's (1996) examination and other simulations run by Story (personal communication), introducing constriction in the pharynx lowers the fundamental frequency. Story has also observed a small decrease in fundamental frequency for the [ɑ] vowel compared with [i], similar to the observations in this study. The observations here fit well with Titze and Story's (1996) observations. That is, more extreme constriction in the pharynx would impose a greater impedance load on the vocal folds (as described in Titze, 1988), thus lowering the fundamental frequency.

This thesis also examines some of the vowels produced in the Esling et al. (1994) study, and found a slight variation in fundamental frequency between the [i] and [ɑ], similar to Story's observations. Interestingly, and consistent with Esling et al. (1994), are the readings for the vowels that have a superimposed raised larynx voice quality, and that are produced with a target pitch of 110Hz. In these vowels, pitch is higher than the modal-[i] vowel with a target of 110Hz (Table 5). Consequently, this supports Esling et al.'s finding that raised larynx setting is associated with a rise in pitch.

Of significance to this study is that the comparison between the [i] and [ɑ] vowels for raised larynx voice are consistent with 'modal' vowels and the vowel/consonant comparison. That is, [i] has higher pitch relative to [ɑ] no matter what the voice quality setting. Of interest for further study, is the reason for pitch to increase for raised larynx voice. The auditory quality of raised larynx voice requires higher pitch in order to be perceived as raised larynx (Esling et al., 1994). For a pharyngealized setting, the impedance load effects should lower fundamental frequency. Raised larynx voice then likely increases pitch by pushing the larynx up to the level of the pharyngeal sphincter and eliminating the impedence matching space.

Shimmer and jitter measurements for articulations that involve aryepiglottic constriction show results that are consistently and significantly higher. This implies that when constriction is introduced into the pharynx for the vowel [ɑ] compared to [i], and for pharyngeal consonants compared to vowels, there is more pitch variation and more amplitude variation.

Harmonics signal-to-noise ratio is higher for non-pharyngeal constricted articulations. This implies that articulations that involve aryepiglottic constriction have less harmonically related energy than non-constricted articulations. Combining all the voicing analysis results shows that aryepiglottic constriction lowers pitch but raises the modulation of both pitch and amplitude, and increases the amount of non-harmonically related noise in the acoustic and EGG signals.

These results can be generalized to a larger population because the language data includes male and female speakers, and each language example is spoken by a native speaker of the language. It seems that the values for pitch, jitter, shimmer and HNR are useful in determining if one voiced speech articulation contains more aryepiglottic constriction than another. It seems also that these measurements can be used either within a certain speaker's production of vowels or consonants, or between different speakers with respect to consonant-vowel combinations since the measures depend on relative values.

Also, even with a larynx height voice quality setting superimposed on vowels, an increase in aryepiglottic constriction would be detectable based on these measurements. That is, if raised larynx voice is superimposed onto an utterance, the results should still show a decrease in pitch and HNR, and an increase in jitter and shimmer for articulations with increased pharyngeal constriction.

4.4 Conclusion

First of all, it should be noted that the voicing analysis measurements, with the exception of pitch, are somewhat inaccurate measurements; not only by the observations noted here but by the results of the Karnell et al. (1995) and Gelfer ad Fendel (1995) studies, which illustrate the variability of these measurements between analysis systems and mediums. However, in this thesis, voicing analysis measurements are compared between a vowel and a neighbouring consonant; therefore, the measurements are relative and even if the absolute measurement is indeterminate, the relative measurements are consistent, and thus somewhat useful. The statistical analysis method used to compare the sound segments also minimizes the inaccurate absolute measurements.

The voicing analysis measurements of pitch, jitter, shimmer and HNR used in the CSL package are valuable since they can be used on either an audio signal or an EGG signal and produce consistent results. Also, these measurements from sounds that involve aryepiglottic constriction compared with other types of articulations are consistent and statistically significant. That is, pitch, jitter, shimmer and harmonic signal-to-noise ratio are measurements that can be used predictively to determine if one sound contains more aryepiglottic constriction than another. These measurements, based on their statistical assessment, can be used on a single speaker, between different speakers, and on voices that have a quasi-permanent voice quality, such as raised larynx, since the measurements are relative. Also, the assumption that increased values for jitter and shimmer imply harshness is not supported since jitter and shimmer values increase for the vowel [ɑ] compared with [i], and it is not common that [ɑ] is perceived as more harsh than [i].

It appears then, that a shift downward in pitch is the most salient feature that identifies changes in pharyngeal posture. The other voicing measurements are consistent with respect to each other and with what is predicted, but the absolute values are unreliable. This results study, similar to Karnell et al. (1995) and Gelfer and Fendel (1995), question the use of jitter, shimmer and HNR as measurements that describe speech features.

Chapter 5 - Acoustic Language Data

5.0 Introduction

This chapter examines real language data from an acoustic phonetic perspective in order to determine to what extent natural languages use the different manners of pharyngeal articulation that are investigated throughout this thesis. In order to do this, languages that contain post-velar segments were isolated from two databases, the University of Victoria Phonetic Database (PDB) and UCLA's Sounds of the World's Languages (SOWL).

The theory of four manners of articulation presented in this thesis will be tested acoustically using natural language data to offer the phonologist some facts to consider when assessing how a language uses these sounds. It is important to consider how a language uses the segment in the phonology in order to determine the status of manner of articulation. Al-Ani (1970), Catford (1983), Traill (1986), Jacobsen (1969), Bessell (1992, 1993) and Krauss (1979) all discuss phonological evidence that motivates the investigation of varying manners of pharyngeal articulation in Arabic, Caucasian, African and North American languages. Consequently, there is evidence in the literature to support this acoustic phonetic examination.

Both databases that were used to obtain language data are excellent resources for language data; the PDB is geared toward more experienced linguists since complete word lists, sentences, stories and in some cases, songs are stored. The PDB data is captured at either 10 or 20 kHz and 16-bit quantization, thus the quality of digitization is considered excellent even if the original data were recorded in a noisy environment. The SOWL database consists of small amounts of data presented to illustrate interesting aspects of languages. The data in this database are captured at 8kHz sampling rate and 8-bit quantization; consequently there is noise associated with the quantization rate that contaminates the data, making it difficult to analyse in some cases.

Not all pharyngeal segments in the databases were analyzed acoustically since some pharyngeal segments were identifiable not by their acoustic quality but by their effect on neighbouring vowels. This phonological quality of pharyngeal segments is one of the features that motivates the pharyngeals to be in a natural class along with uvular and laryngeal sounds.

In section 5.1, various acoustic and auditory descriptions produced by other researchers will be discussed, and it will be shown that their descriptions comply with what is observed in the cardinal consonant productions covered throughout this thesis. For example, Bessell (1993, p. 45) associates pharyngeals with creak, which is a very slow vibration of the vocal folds. The lowered pitch that is associated with pharyngeal sphinctering (see chapter 4) accounts for this auditory assessment.

As we have seen throughout this thesis, the pharyngeal sphinctering mechanism may or may not involve the larynx rising up toward the epiglottis. Bessell (1993, p. 46) shows a table that has an articulatory description of /ʕ/ and /ħ/ involving larynx height, or at least a laryngeal component. This ability to move the larynx within the space of the pharynx provides another potential distinctive feature, namely raised or lowered larynx. A phonological discussion of larynx height as a distinctive feature can be found in Trigo (1991). Phonetic evidence of this is seen in the Mpi and Bruu language data discussed in section 5.2.

Also in section 5.2 is a description of Agul, a language in the SOWL database that provides a good illustration of post-velar sounds. These sounds are used to test the features observed in the previous chapters for the cardinal examples of pharyngeal manners of articulation. The Agul data could alternatively be interpreted with larynx height as the distinguishing feature rather than a change in manner of articulation.

The acoustical analysis of language data shown in this chapter provides phonetic evidence to support the hypothesis that there can be four manners of articulation in the pharynx. Acoustic evidence of larynx height adjustments suggests this may also be a distinctive feature available for phonologists to consider in the assessment of a language's use of pharyngeal sounds.

5.1 Acoustic Phonetic Support for Four Manners of Articulation

The Pharyngeal Stop:

Pharyngeal plosive-type behaviour is described by Al-Ani (1970) as the more common allophone of the voiced pharyngeal fricative in Arabic. Butcher and Ahmad (1987) in Iraqi Arabic identify alternatives to the pharyngeal fricative as being either a pharyngeal approximant or stop articulation. These observations illustrate not only phonetic evidence of the existence of a pharyngeal stop, but also the great variation in production of pharyngeal speech segments in general (Thelwall 1990).

In the SOWL database, there is Egyptian Arabic data that shows the emphatic alveolar plosive and its contrasting non-emphatic version. The emphatic series is said to have a pharyngeal component to its articulation (McCarthy 1994), and consequently is examined in Figure 22 for pharyngeal effects. The first utterance in the spectrogram in Figure 22 is the non-emphatic **[t]** in the word **[ti:n]** meaning 'mud'; the second utterance is the emphatic **[tˤ]** in the word **[tˤi:n]** meaning 'figs'. Note that the formants in the emphatic version appear to be coming from an **[ɑ]**-like vowel (raised first formant (F1) and lowered F2). This formant transition is much less marked than in the non-emphatic [ti:n]. This spectrographic comparison provides evidence that the emphatic is produced further back in the vocal tract, however, it does not necessarily show if the production is uvular or pharyngeal. This question of whether the emphatic series is uvular or pharyngeal in place of articulation arises from the literature (McCarthy 1994) and from the discussion of the Hebrew post-velar series in Figure 28.

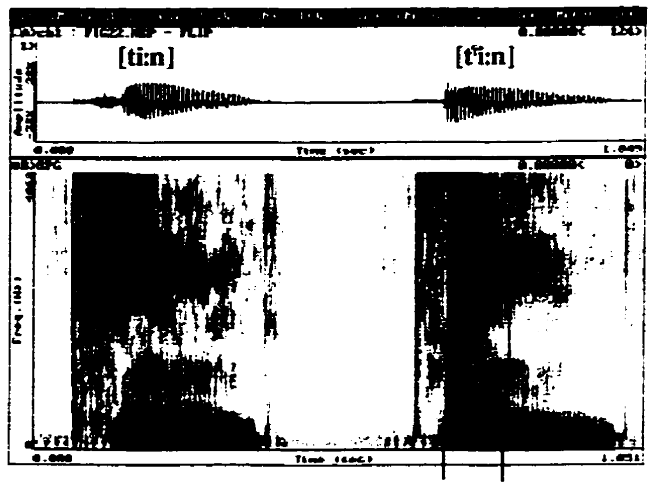


Figure 22: Arabic (Syrian dialect) [ti:n] compared with the emphatic [tˤi:n]. The formant structure in the emphatic suggests the release involves post-velar constriction.

Figure 22 shows how pharyngealization can be applied to another consonant and the acoustic effect of such behaviour. Figure 23 is a Hebrew example from the SOWL database that shows similar formant effects in the contrast between a glottal stop and its pharyngeal correlate. In chapter two, the glottal stop was compared with the pharyngeal stop in order to illustrate the difference between these two similar sounding segments. In this case the two are compared because McCarthy (1994) suggests that the glottal stop is in the same natural class as other post-velar segments.

The Hebrew data in Figure 23 show acoustic effects similar to Figure 22. In the first utterance of the word /ʕor/ meaning 'skin', the formants in the pharyngeal sequence suggest the vocal tract was in an [ɑ]-like posture, whereas the glottal consonant in the word /ʔor/ meaning 'light', shows no formant transitions. This acoustic comparison identifies the differences between these two segments but does not help the phonologist with acoustic features that would suggest the glottal stop is in the post-velar natural class.

Also in Figure 23, note the amplitude burst of the two Hebrew sounds compared with the cardinal examples in Figure 3. In Figure 3, the slope amplitude in the cardinal pharyngeal stop is steeper than the glottal. In Figure 23, the slope amplitude of the word-initial pharyngeal is less compared with the glottal. This suggests that burst amplitude does not define the pharyngeal from the glottal, for Hebrew speakers at least, and it questions whether this acoustic feature is salient.

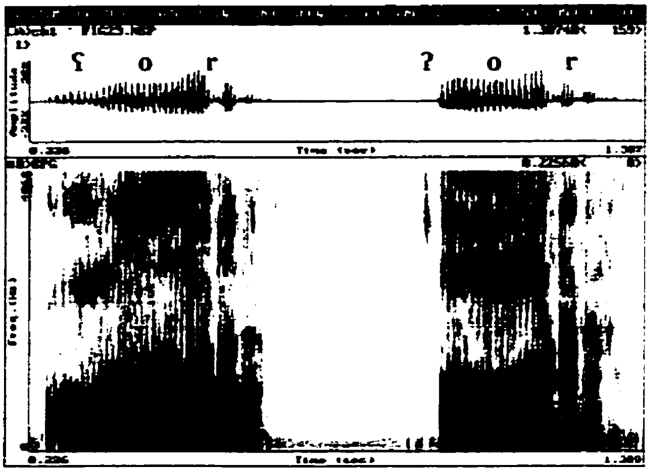


Figure 23: A comparison of the Hebrew (Oriental dialect) pharyngeal stop with the glottal stop. The words mean 'skin' and 'light' respectively.

Rose (1976) describes a plosive in the Ahousaht dialect of Nuuchahnulth (Nootka), a Native North American language. Rose illustrates phonologically that the /ʕ/ alternates with /q/, the uvular stop, providing evidence for /ʕ/ as a plosive. Also, Jacobsen (1969) argues that proto-Nootkan uvular stops merged to /ʕ/. Massett Haida also contains a pharyngeal that apparently patterns similarly to Ahousaht and exhibits, a stop-like articulation acoustically (Bessell, 1993).

Figure 24 shows formant transition evidence from Ahousaht (PDB) that supports a pharyngeal plosive interpretation for [ʕ]. The formant transition at the beginning of the utterance shows F1 and F2 diverging from about 1000Hz, similar to the cardinal example. There is evidence of pharyngeal articulation in the middle of the utterance aiso. The fricative portion has pharyngeal formant characteristics. The segment then changes to a region of less friction so that there is almost no amplitude in the waveform but there is a hint of energy in the spectrogram. Auditorily, this sequence sounds like the speaker almost comes to a stop. Acoustically, this sequence appears to be two segments because of the sharp change in acoustic behaviour and the region of acoustic stability of about 10ms following the [ħ]. These two sounds must differ in manner of articulation since evidence seen in the stable formants in the frequency domain shows that place of articulation remains the same. This observation of acoustic stability along with the consistent formant behaviour through both regions of acoustic stability, lends support to Steven's (1972) Quantal Theory. Acoustically, the lowest amplitude portion must be a stop since there is no perceptual category for a 'really weak' fricative.

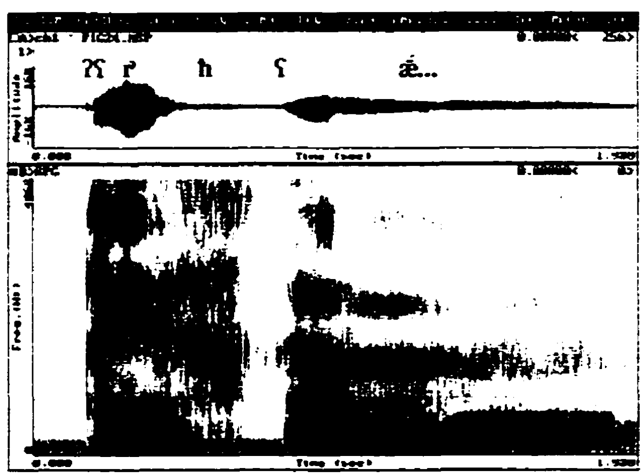


Figure 24: Evidence from Ahousaht of a pharyngeal fricative and plosive. The utterance is an onomatopoetic expression describing the sound of weeping.

The literature provides auditory reports of languages using pharyngeal stops and the languages analyzed in this section show acoustic evidence to support these observations. The formant evidence surrounding the stop portion in the utterance suggests the vocal tract is constricted in the pharynx.

The Pharyngeal Trill

Trilling-like qualities in the pharynx have been described auditorily, in Hydaburg Haida (Krauss, 1979; in Bessell 1992), in !Xóõ (Traill, 1986) and in this investigation with respect to Agul[1] (see section 5.2 for the Agul discussion). The Hydaburg Haida description is an auditory description and not supported by an acoustic description to date. Bessell (1992) has suggested that the Haida segment behaves more like an affricate phonologically; as mentioned in section 2.1 the amplitude release in stop-like sequences has been suggested as an acoustic cue for identifying affricates. However, with respect to the discussion of Figure 23, this feature is not likely salient.

In !Xóõ, Traill (1986) provides a video and acoustic description similar to what is seen in this investigation. Traill measured trilling-like qualities that occur about 50 times per second, which is consistent with the cardinal trill in this study. Traill's account also describes the aperiodic, or irregular nature of trilling that is observed here. In Figure 25, it is possible to see trilling during the strident vowel portion particularly in the second speaker's example of the word. The trilling behaviour is visible in the waveform by the more broadly spaced pulses and in the spectrogram by broadly spaced vertical striations. Also noteworthy is the low amplitude of the waveform during the vowel in question; and the formant structure which is consistent with pharyngeal constriction.

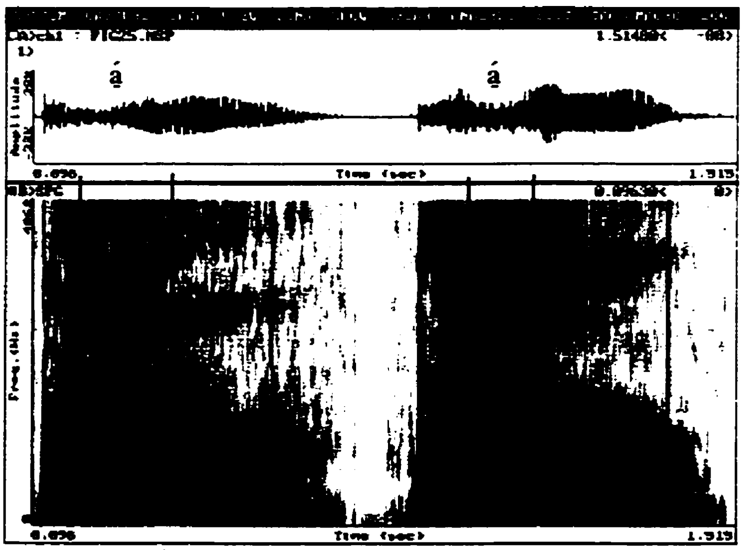


Figure 25: Pharyngeal trilling in the !Xóõ. The word is k!ạo meaning 'base', and is spoken by two different speakers.

The trill produced by the phonetician in Figure 9 does not show as much irregularity or frication as in Figure 25, likely because the trill is produced in a controlled manner, and it is produced in isolation. A speaker of !Xóõ may show more regular trilling if asked to produce the trilled vowel in isolation. In addition, there is the inherent problem of quantization noise in the SOWL database which produces noisy spectrograms.

The articulatory description of this trilling behaviour in !Xóõ is described in the SOWL database is as follows:

> "The whole body of the tongue is much lower for the strident vowels, the back wall of the pharynx is drawn forward, and the epiglottis vibrates rapidly. There is also a constriction between the part of the tongue below the epiglottis and the tips of the arytenoid cartilages" (SOWL, !Xóõ card in HyperCard stack).

Evidence presented throughout this thesis suggests that in this type of sound, the pharyngeal wall does not draw forward but rather the arytenoid cartilages and the larynx as a whole tilts anteriorly. The epiglottis itself does not vibrate but the aryepiglottic folds vibrate, which may cause the epiglottis to vibrate.

The use of trilling in !Xóõ is not as a consonant segment but is said by Traill to be a phonation type superimposed onto a vowel. Thus, a phonation type is used as a distinctive feature in this language. This poses a problem for phonological theory since the feature [voice] is no longer adequate for such phenomena. That is, there are voiced and voiceless segments in this language along with 'voicing due to aryepiglottic fold trilling' for which there is no distinctive feature that is used in phonological theory. In this thesis, the evidence of trilling capabilities in the pharynx is what is pertinent, leaving the phonological problem open for discussion.

Evidence that Ahousaht uses this trilling activity is found in the PDB. Figure 26 shows the Ahousaht rendition of the "sound of a barking dog." The dark striations in between the dark bars in the spectrogram are argued to be aryepiglottic trilling because of the spacing that occurs about 50 times per second and the formant structure which is consistent with constriction in the pharynx.

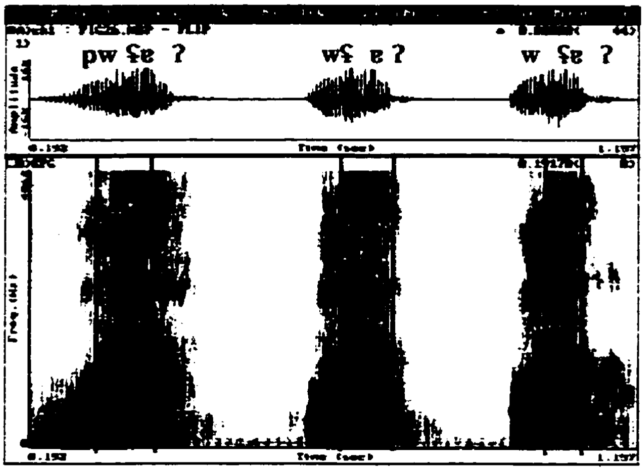


Figure 26: Ahousaht data of the onomatopoetic use of pharyngeal trilling: the sound mimics a barking dog.

It appears that at least two languages make use of trilling behaviour in the pharynx. One language uses a trilled segment as a sound in onomatopoetic expressions and the other language uses the segment as a vowel that differs from other vowels by some type of voicing feature. The new voicing feature is both interesting and problematic for theoretical linguistics, but phonetically, it provides evidence for the manner of articulation of trilling in the pharynx.

The Pharyngeal Fricative:

The pharyngeal fricative is common in Semitic languages. However, as mentioned earlier, Butcher and Ahmad (1987) and Laufer (1996) suggest that the voiced pharyngeal segment can also be realized phonetically as an approximant or a stop. This evidence illustrates the phonetic variability of these phonological units. This variation is hypothesized to be a result of the sound being produced in a part of the vocal tract that is not visible, so that articulatory cues that identify the segment are not available to the listener.

An Egyptian Arabic example from the PDB is illustrated in Figure 27. Figure 27 shows clearly the fricative portion of the consonant. The formant structure throughout the noise portion shows enhanced energy at about 1000Hz where F1 and F2 merge, and a strong F3 along with strong upper frequency energy in general.

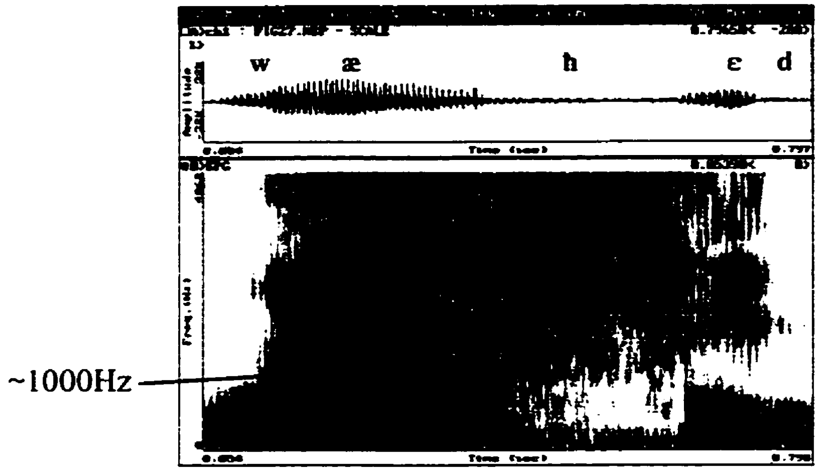


Figure 27: The pharyngeal fricative in the Egyptian Arabic word meaning 'one'.

Figure 28 shows a series of post-velar fricatives in Hebrew (SOWL). The center example is the pharyngeal fricative that shows aperiodic 'noise' with a merged first and second formant along with a strong third formant, suggesting pharyngeal constriction. The lower formants then split into the [o]-vowel following.

The first utterance in Figure 28 is the word, 'χimiɑ', meaning 'chemistry', which contains the uvular fricative in the word intial postion. This sound contains a second formant transition that starts at about 1000Hz. This [ɑ]-like transition seen in the uvular presents a bit of a problem when compared with Figure 27 since it could be argued acoustically that the consonant in Figure 27 is produced at the uvular place of articulation.

The challenge that these two Hebrew words present is to identify the acoustic difference between the uvular and pharyngeal consonants. Ideally, the first and second formants should be essentially one formant for the pharyngeal; however, the cardinal example in Figure 10 shows that the pharyngeal consonant in the environment of the [i] vowel does not reach its merged F1 and F2 state, yet it is perceived as maximally pharyngeal. The most promising difference between the uvular and pharyngeal consonants in Figure 28 is the third and fourth formants, which are stronger in the pharyngeal. This formant comparison points toward the need for an upper frequency model of vocal tract behaviour. A valuable task would involve a perceptual and acoustical study of the effects of changing from uvular to pharyngeal articulation.

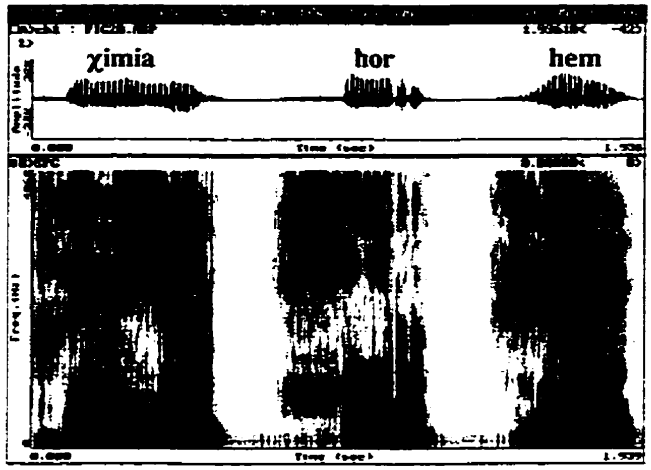


Figure 28: The Hebrew uvular, pharyngeal and glottal fricatives respectively, in word initial position.

Considering for a moment McCarthy's proposal that glottals are considered to be in the same natural class as the post-velars, the third utterance in Figure 28 contains a glottal fricative which shows that the glottal fricative mirrors the formants in the following vowel. The lower frequency energy is quite weak and there is a formant transition in the glottal at the 4000Hz mark. The upper formant movement suggests that the key acoustic feature for the post-velar class may well be formant information in the upper frequencies. Another acoustic feature that may identify the post-velar class is the presence of energy across the frequency spectrum, although the formant information in the glottal [h] is quite weak.

Figure 29 shows data from Ahousaht (PDB) and fricative behaviour with clear formants that correlate with pharyngeal-like behaviour; that is, F1 and F2 have merged. The most notable pharyngeal effect in Figure 29 is the presence of energy at the low frequencies. The previous language examples in Arabic and Hebrew also show energy across the frequency spectrum; consequently, these data suggest that the entire vocal tract is resonating.

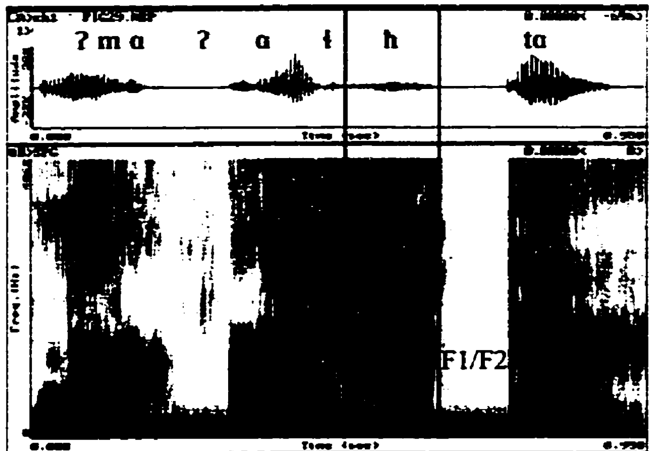


Figure 29: Ahousaht data showing energy across the entire frequency spectrum for the fricative consonant. F1 and F2 are quite close suggesting pharyngeal constriction.

Figure 30 shows Avar (SOWL) language data that compares the pharyngeal fricative with the glottal fricative. This is an ideal acoustic comparison because the words are identical except for the final consonant. The pharyngeal fricative in the second example of the word **maн** meaning 'odour', has stronger energy across the frequency spectrum compared with the glottal word **mah** meaning 'bundle'. The formant information in the 2000-4000kHz range is stronger and more defined, showing an upward shift compared with the glottal. The one quality that is similar between these segments, aside from the manner of articulation, is the fact that both sounds have energy across the frequency spectrum, thus in both cases the entire vocal tract resonates. This similarity may be useful in providing phonological evidence that would motivate including glottals such as [**h**] in the post-velar class.

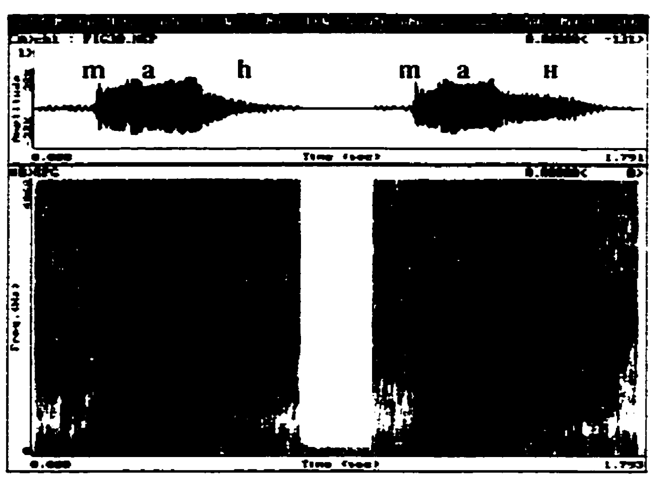


Figure 30: The Avar glottal fricative in word-final position compared with the pharyngeal fricative.

There appears to be plenty of acoustic evidence of the pharyngeal fricative in the language data, supporting its status as a symbol on the IPA chart. The questions that arise are: What distinguishes the pharyngeal from the uvular articulation? And, if the glottal is considered to be in the same natural class as the pharyngeals, then what is the distinctive feature that they appeal to? Evidence presented here suggests the upper frequencies may distinguish between pharyngeal and uvular sounds; and that the acoustic feature that includes glottals in the class of post-velars may be formant changes in the upper frequency range or, the presence of energy across the frequency spectrum.

The Pharyngeal Approximant:

The pharyngeal approximant appears phonetically in Arabic (along with its many other phonetic variants). Bessell (1992) provides both phonetic and phonological evidence for the pharyngeal approximant as a segment in Interior Salish. Laufer (1996) suggests that in Arabic the common voiced pharyngeal is actually an approximant and not a fricative. It appears that there is no controversy among researchers that a pharyngeal approximant is possible. This section will illustrate the approximant and the pitch feature that is observed in chapter four in the cardinal pharyngeal approximant.

Hebrew (SOWL) and Arabic (PDB) have clear acoustic phonetic examples of a pharyngeal approximant interpretation for the [ʕ]. Figure 31 shows the Hebrew example where there are clear striations that correlate with vocal fold behaviour. The striations are more broadly spaced than during the vowel portion suggesting a lowered vocal fold vibration rate and thus, lower pitch. This lowering of pitch is predicted by Titze and Story based on their model of pharyngeal constriction and the lowering of vocal fold vibration rate observed throughout this thesis.

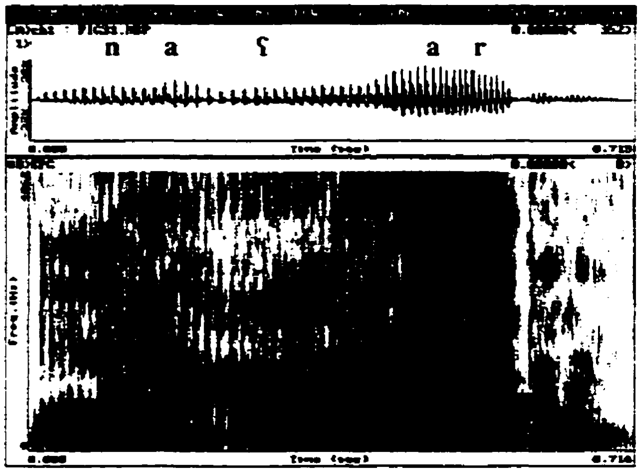


Figure 31: The spectrogram of the Hebrew /ʕ/ in the word **naʕar** The clear striations during the consonant suggest it is an approximant. Slower vibration and merging first and second formants suggest the sound is pharyngeal.

The formant behaviour indicates change in vocal tract behaviour compared with the vowels, and corresponds with the description in section 2.5, that is, merging of the first and second formants at about 1000Hz, and changes in formant behaviour in the upper frequency region.

Figure 32 shows Egyptian Arabic (PDB) data with similar characteristics to Hebrew, in that there is distinct voicing throughout the consonant portion. In Figure 32, the vocal fold behaviour is slower and more fricative-like, however, the fricative behaviour is not as extreme as the examples in the pharyngeal fricatives (Figs 27-30) since there are visible voicing striations throughout the consonant portion. Also noteworthy in both Figure 31 and 32 is the marked lowering in amplitude in the waveform of the approximant during the consonant portion.

There appears to be little debate as to the existence of a pharyngeal approximant as suggested in the literature and shown in Figures 31 and 32. The evidence of vocal fold vibration during the consonant offers support for the segment being classed as an approximant manner of articulation rather than a fricative. The slowing of vocal fold vibration is consistent with the evidence presented in chapter 4.

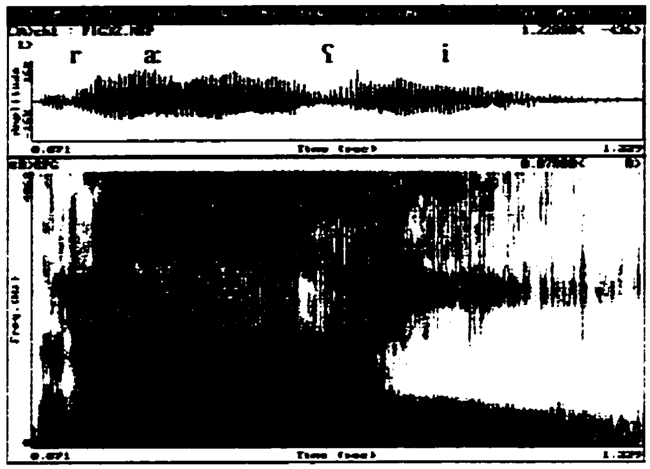


Figure 32: The Arabic word raːʕi . The formant behaviour and slow vibration suggest it is an approximant. Note too the lower amplitude in the waveform during the consonant.

This chapter has presented language evidence to suggest that languages use, at least phonetically, four manners of articulation in the pharynx. The following section discusses the use of four manners of articulation with respect to one language that contains pharyngeals as well as epiglottals. This language is also discussed with respect to the larynx height feature discussed throughout this thesis.

5.2 Agul and Raised Larynx as a Distinctive Feature

This section presents the Agul language data which supports the use of four manners of articulation as well as larynx height as a distinctive feature. The categorization of the Agul language data in the SOWL database offers a challenge to the pharyngeal behaviour theory proposed here, because the SOWL database makes a distinction between two places of articulation in the pharynx, namely, pharyngeal and epiglottal. This implies that there are two regions in the pharynx that can act as places of articulation. The assumption made in this thesis is that the pharynx acts as one place of articulation with varying manners.

In Agul, an 'epiglottal plosive' is noted by Ladefoged and Maddieson (1996). Figure 33 shows this sound with the first and second formant transitions going toward and coming from about 1000Hz, and a strong third formant. The broadly spaced striations in the spectrogram preceding the complete closure imply either a slowing of the vocal folds or 'flapping' of some other structures such as the ventricular or aryepiglottic folds, depending on the frequency.

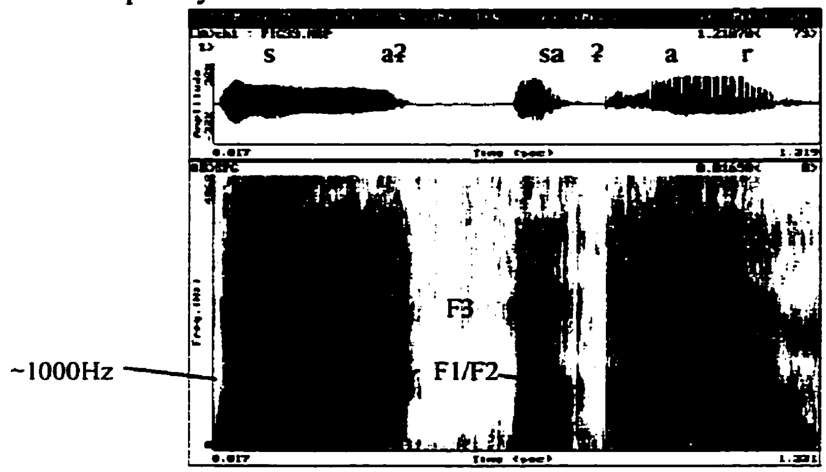


Figure 33: The Agul plosive showing the first and second formant merging at around 1000Hz.

Unfortunately, the glottal stop is not represented in the Agul data to compare the vowel and stop lengths. Attempts were made to synthesize the Agul pharyngeal stop into a more glottal sounding stop in order to determine if the pitch, amplitude and vowel shortening effects are the salient features that distinguish glottal and pharyngeal stops. However, it was not successful. This failure to synthesize is due to the lower sampling and quantization rates of the SOWL data which are difficult for the synthesis algorithm to accommodate.

The Agul epiglottal fricative in Figure 34 and the phonetician's example of a voiceless pharyngeal trill are similar auditorily. The Agul example can be argued to have trilling-like qualities in the spectrogram and waveform as seen in Figure 34. The measurements taken from this example reveal pulsing at a rate of about 40Hz, which is consistent with the cardinal pharyngeal trill, and far too low to be due to glottal voicing produced by the vocal folds.

The trilling interpretation of this Figure could alternatively be argued that the larynx rises. This would cause the formants to rise during the consonant, which could be argued to be the case in both Figure 33 and 34. In fact, both trilling and raised larynx could be argued to be occurring in this example of Agul.

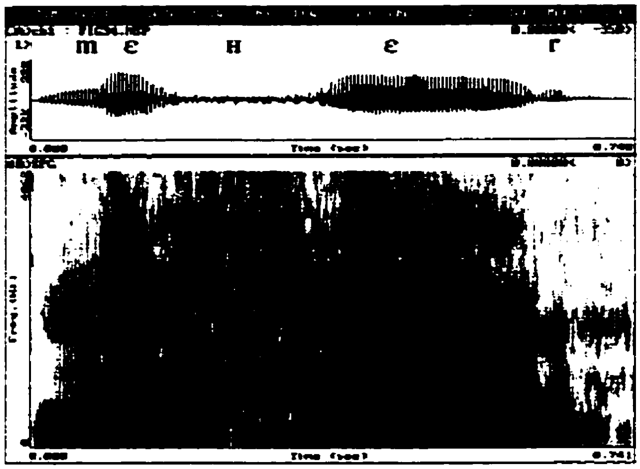


Figure 34: The Agul epiglottal fricative with trilling-like qualities. The interval for the pulsing is about 40Hz.

The Agul example of a pharyngeal fricative is seen in Figure 35. In word-final position the fricative shows energy across the frequency spectrum but no strong formant energy. In the second word, the [ħ] is between the vowels /u_a/ and there is a formant transition that appears to be the second formant rising toward about 1700Hz; this may be a transition from a more open vocal tract in /u/ to where the epiglottis cuts the cavity in two for /a/. In Figure 34, there is a concentration of energy at about 1700Hz which is associated with the third formant and is somewhat lower than F3. This lowered concentration of energy in the pharyngeal fricative could be due to a lowered larynx setting when compared with the epiglottal trill in Figure 34.

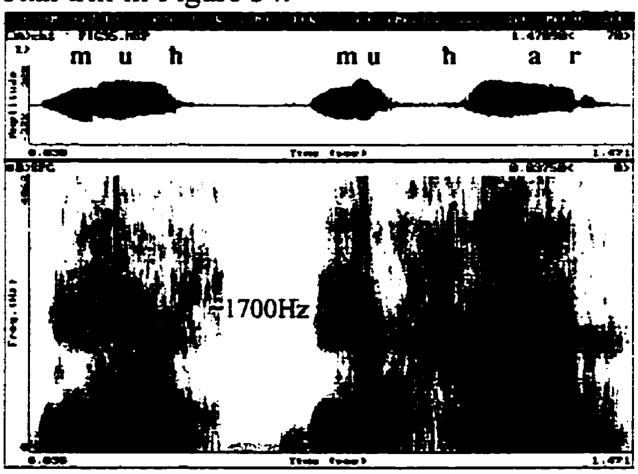


Figure 35: The voiceless fricative in the Agul words **muħ** and **muħar.**

Figure 36 shows the pharyngeal approximant. There is [ɑ]-like formant behaviour but there are clear striations indicating the vocal folds are vibrating. The spacing of striations is broader than during the vowel portion, suggesting the vocal folds are vibrating more slowly because of the impedence matching quality of the aryepiglottic sphincter.

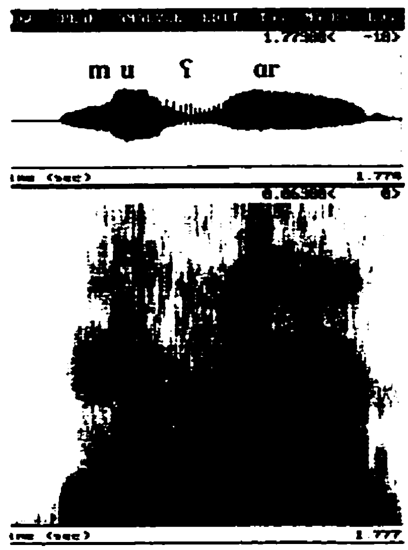


Figure 36: The pharyngeal approximant in the Agul word **muʕɑr**

The preceding interpretation of the Agul language data suggests that four manner of articulation in the pharynx can account for this range of post-velar sounds. Language evidence from Mpi and Bruu of larynx height adjustments also suggests that larynx raising may be a second distinguishing acoustic feature.

The theory of post-velar articulation presented in this thesis can also be considered with respect to the Daghestanian languages which have a full pharyngeal series including uvulars which contrast in "voicing, aspiration, ejective, strength, labialization and palatalization" (Kodzasov, 1987). Historically Kodzasov reports that, [h] goes to [ħ] and [ʔ] goes to [ʕ]. Along this continuum, laryngeals evolve to pharyngeals and pharyngeals evolve to epiglottals. Again, this continuum would be explained, in the current theory, as a change in degree of constriction, or manner of articulation or the inclusion of raised larynx along with all the other unusual contrasts mentioned in Daghestanian. The phonological use of post-velar segments in a language would likely reveal if the language is using a manner of articulation continuum or larynx height change.

Other languages appear to use larynx height as a distinctive feature. Mpi and Bruu show acoustically that a raised larynx setting distinguishes between sounds. The term used to describe this effect in the SOWL database is 'laryngealized', but the acoustic effect of raised formants is clearly consistent with raised larynx as seen in Figure 37.

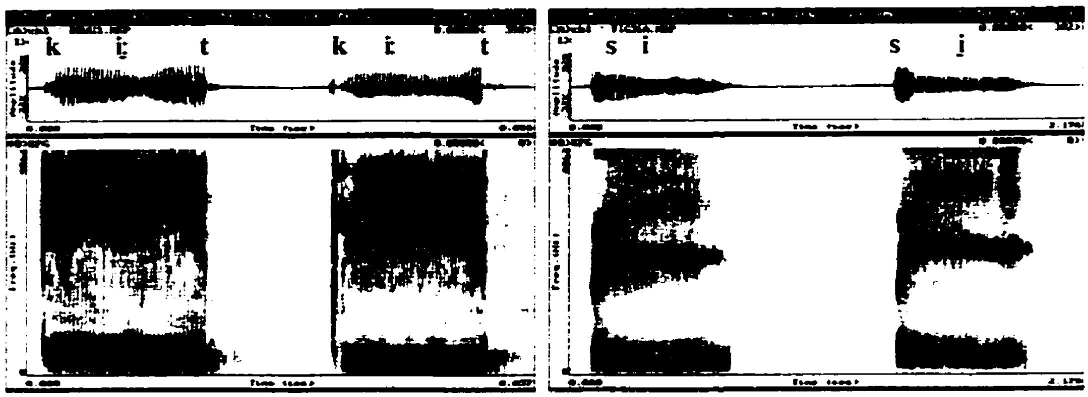


Figure 37: Bruu and Mpi words that show the effects of larynx height adjustments, and the use of this as a distinctive feature.

It appears from the acoustical evidence from Agul that the series of post-velar segments may be a full set of manners of articulation of pharyngeal consonants. It is now the phonologist's challenge to test if the phonology corresponds to these interpretations.

5.4 Conclusion

The language evidence collected from the two databases (PDB and SOWL) suggest that, phonetically, languages of the world use varying manners of articulation in the pharynx, and that larynx height could be argued to be a distinctive feature.

The Agul data in particular provide a description of a language that uses a full range of post-velar consonants and can be interpreted as having varying manners of pharyngeal articulation. Or, the epiglottal consonants could be interpreted as pharyngeal with raised larynx as a distinctive feature. The Mpi and Bruu data provide evidence that larynx height is a plausible interpretation, since these languages use this feature.

Literature reports such as Laufer (1996) and acoustic evidence from Semitic languages provide motivation for a pharyngeal fricative and approximant. Evidence of a stop is presented in Hebrew, Ahousaht and as a feature superimposed onto a set of stops in Arabic. The Arabic data illustrates the pharyngeal feature of F1 rising and F2 lowering and a strong F3 when it is combined with another consonant. The Hebrew data provide a comparison between the glottal which shows no formant changes on neighbouring segments.

Since the Ahousaht example is an onomatopoetic expression, it cannot be said that the pharyngeal stop is used as a phoneme in the language, however, the noted features of the pharyngeal fricative and its neighbouring segment illustrate the acoustic stability of segments and the types of acoustic features that are considered when determining manner of articulation.

The trill is perhaps the most unusual segment auditorily and acoustically. The use of it in !Xóõ as a feature that is applied to vowels provides a unique voicing feature. This trilling quality is used paralinguistically in Ahousaht and thus provides weak evidence for its use as a sound segment in the language, but more strongly suggests trilling as a voice quality feature. Agul offers the best evidence to suggest this sound is a segment used in languages. It is necessary to examine how all sounds in Agul pattern and then compare the post-velar series to determine if they pattern as four manner of articulation. Another confounding factor is larynx height which may act as a distinctive feature or as a quality that is inherently a part of trilling.

The change in upper formant information appears to be the distinctive feature of post-velars. The acoustic data for the uvular in Figure 28 shows that F1 and F2 could be confused acoustically with the pharyngeal sound and poses a problem for distinguishing between uvular and pharyngeal manners of articulation. It is hypothesized that the upper formant frequencies help to describe the difference between uvular and pharyngeal places of articulation. In the case of an approximant it could be argued that a drop in pitch is the salient distinctive feature.

What is left for further research, along with creating a model of upper formant behaviour, is how these different phonetic pharyngeal segments are used phonologically in the languages that use these sounds. Of specific interest is: Are the different phonetic pharyngeal consonants separate phonemes, or alternates of one phoneme? The challenge will be to find a language that uses three or four manners of articulation in separate pharyngeal phonemes. The closest example of this is Agul, which uses multiple pharyngeal consonants. The task for further research is to determine how Agul makes phonological use of the set of pharyngeals as four manners of articulation, or if the contrasts can be described more accurately with the feature raised larynx.

[1]Hydaburg Haida is a language isolate belonging to the Na-Dene phylum. !Xóõ is a Khoisan language of the San subgroup, and Agul is a Nako-Dagestanian language of the northeast Caucasus.

Chapter 6 - Conclusions

The five preceding chapters have presented a theory that describes pharyngeal behaviour using IPA auditory classifications and Laver's Voice Quality taxonomies. The tools that are used to provide evidence for this theory include a videonasendoscope, electroglottograph, and a collection of acoustical analysis algorithms such as: Fast Fourier Transform (FFT), Spectrograms, Voicing analysis, Pitch extraction, and Analysis-by-Synthesis. These tools are used to examine one phonetician's production of so-called cardinal pharyngeal sounds.

One may ask how reasonable it is to use one phonetician's productions of such sounds? The primary value of using a trained phonetician to produce cardinal speech sound examples is that the subject can actually produce the sounds. This phonetician's examples of pharyngeal sounds provide a standard range of pharyngeal sounds that a human can produce. It is expected that the language data will vary because of inherent speaker voice quality differences and difference that are a result of the speaker's native language or dialect, but the differences should occur in a predictable manner and vary around the parameters that describe the phonetician's cardinal pharyngeal sounds.

Chapter 2 provides acoustic analyses that support one place of articulation for pharyngeal sounds and four manners of articulation. Acoustic evidence suggests that there is no need for an epiglottal place of articulation and the video data in chapter 3 illustrates that only one physical area in the vocal tract that has speech articulation capabilities. The cardinal pharyngeal manners of articulation show acoustic characteristics that are consistent with manners of articulation in general. The language data show that phonetically, languages use all four manners of articulation. The question that remains for the phonologist is whether these manners of articulation are phonemic units.

One problem that is solved by identifying four manners of pharyngeal articulation is the recognition of pharyngeal sounds in speech technology applications. A speech recognizer would have better success identifying the Arabic 'ain', for example, if the recognizer is aware that this phonemic segment can be produced as either an approximant, stop or fricative.

Chapter 3 offers a normalization technique that allows a quantitative video description of the pharyngeal consonants. This technique allows measurement in a two-dimensional image, however, the video data also show that larynx height is a characteristic of the production of cardinal pharyngeal consonants. The acoustic evidence and unmeasurable visual observations, show that the larynx can move below the pharyngeal sphincter. The range of laryngeal movement that is possible could translate into a distinctive feature that can be applied to the cardinal pharyngeal segments and result in the perception of another place of articulation such as 'epiglottal'. It could be that the larynx height feature has confused linguists and resulted in the inclusion of the epiglottal place of articulation on the IPA chart.

Having a sphincter above the vocal folds is likely to affect vocal fold vibration; chapter 4 provides indirect evidence of the effect of pharyngeal sphinctering by monitoring vocal fold vibration so that pharyngeal constriction can be identified on the bases of glottal source changes. The effect of pharyngeal constriction on vocal fold vibration is examined, using the voicing analysis measurements of pitch, jitter, shimmer and harmonic-to-noise ratio. Although the absolute values of these measurements (with the exception of pitch) appear to be variable, the results of the analysis are predictable. These measurements represent pharyngeal constriction consistently in that pitch and HNR lower and jitter, and shimmer increase. The pitch measurements in particular are consistently lowered when pharyngeal constriction is present, as predicted by Titze and Story (1996). Also, the laryngographic waveform of the pharyngeal consonant in Figure 21 shows more variation that is represented by higher jitter and shimmer values and lower HNR.

The language data in chapter 5 are examined with the investigative tools from chapters 2 through 4 to show acoustic qualities of pharyngeal segments that are consistent with the characterization of manner of articulation. The specific features include: for the stops, a break in the waveform along with the first and second formants in the burst diverging from around 1000Hz; for trills, pulsing that is around 50Hz. For the fricative, noise is visible in the spectrogram across the frequency spectrum and, consequently, there is formant behaviour that shows F1 and F2 merging at about 1000Hz. For the approximant, formant characteristics are consistent with those seen in the fricative, along with a lowered vocal fold vibration rate. In the case of the fricative, trill and approximant, there is marked amplitude lowering in the waveform.

This thesis has covered a new area of research using a broad range of analysis techniques and there are questions that remain and issues that are vague. One signal analysis that was not explored is the amplitude and pitch variation over time. The jitter and shimmer measurements examined here are cycle-to-cycle variations and since they do not appear to be salient acoustic features, it is hypothesized that some other type of pitch and amplitude variation may be more valuable. That is, perhaps pitch and amplitude variations are more salient when they create an amplitude cycle that repeats 10 or 20 times per second, rather than occurring between vocal fold pulses. This type of measurement is motivated by the observations made with respect to the uvular trill in section 2.2 which shows the waveform amplitude repeating a pattern at 25Hz.

The cycle-to-cycle voicing measurements, in spite of their unreliable absolute values, describe the effects of pharyngeal constriction on the vocal folds consistently. That is, pitch, jitter, shimmer and HNR are consistent when used to measure sounds that are in the same utterance. Also, these measurements are derivable from the acoustic waveform and provide consistent results with the EGG waveform. The value of being able to examine the acoustic waveform is that a variety of language examples can be examined in order to determine if pharyngeal constriction is present.

Analysis-by-synthesis was used minimally here but could easily constitute a new study of pharyngeal place and manner of articulation. It was difficult to determine how the pharyngeal stop differed acoustically from a neigbouring [ɑ]-vowel thus, it was necessary to use the synthesis technique to test what acoustic features are salient to listeners. Each of the cardinal pharyngeal segments could be tested by synthesis in order to confirm that the acoustic features that were measured correspond to meaningful changes auditorily. The results of this thesis provides a systematic description of source and filter parameters that would help synthesizing pharyngeal sounds as well as pharyngeal voice quality settings.

The observation of changes in the glottal source and formant characteristics with pharyngeal sounds has shown that the present source-filter model of speech production is inadequate. The present source-filter model has the source as the sound which is produced at the glottis by the vocal folds, this source is then filtered by the rest of the vocal tract. The evidence presented here forces a modification to this theory since it shows that constriction in the vocal tract can affect vocal fold vibration. Also, the acoustic and visual observation of trilling aryepiglottic folds shows that there can be another source of sound above the glottis. The following block diagram illustrates how the pharyngeal component can be included into the source-filter model:

This diagram shows that the glottal source sound passes through the pharyngeal component which may or may not affect the glottal source, and may act as a source itself, before being filtered by the rest of the vocal tract.

The observed changes in upper formant frequencies for all of the post-velar sounds examined here suggest that changes in the back part of the vocal tract are acoustically realized above the third formant. Current models of speech production have focused on the oral and nasal cavities and their affect on the resonance characteristics of the vocal tract; even Fant's (1960) model that includes up to 5 formants does not characterize the upper formant changes observed in the data in this thesis. Sundberg's (1987) description of the one-sixth larynx to pharynx ratio that contributes to the singer's formant does not imply that there will be a formant around 3000Hz. Titze (1993) develops Sundberg's ratio measurement further to include that the length of the larynx tube must also be one-sixth the length of the vocal tract in order to create the singer's formant. Based on the observations in this thesis Titze's modification of Sundberg's ratio measurement is likely to be a more precise description of the posture necessary for good singing.

Also, the language data presented in chapter 5 reveal that changes in larynx height can be a distinctive acoustic feature. It is shown from past research on voice quality settings that a raised larynx setting is on a continuum with pharyngeal constriction. This correlation between voice quality settings implies that there is an auditory and acoustic connection between larynx height and constriction in the pharynx that is likely to manifest itself in speech segments. It is not clear from this research how larynx height is used phonologically, however, it is plausible that all four manners of articulation can be produced with a raised or lowered larynx. If raised larynx has been associated with the term laryngealized, as in the Mpi and Bruu language data, then it may be the case that the glottalized pharyngeal sounds in the Native North American Interior Salish languages are raised larynx pharyngeal manners of articulation. Since no Interior Salish data is examined in this thesis this question remains open for future investigation.

Investigating pharyngeal sounds with the video nasendoscope shows the larynx rising and lowering, however, the is a need for a quantifiable depth measurement in order to confirm this scientifically. The most practical method would be to include MRI, x-ray or ultrasound along with the acoustic and video analysis used in this thesis. Other methods for depth measurement were considered for this thesis but would have involved adding extra fibers or lasers to the fibrescope and then testing and developing a measurement technique. The value of the measurement technique that is developed and applied in this thesis provides is that a quantitative two-dimensional analysis that can be used to compare movements in the pharynx between subjects and recording sessions.

The normalization of video pictures that was applied to video frames that represented the most constricted portion of a sound did not show significant differences between specific measurements for different sounds; however, the variation in measurements proved to be significant. The variation in particular measurements showed that the pharyngeal trill and approximant require a much more stable vocal tract position than a fricative or stop. These results are predictable since it is likely that the pharyngeal structures would have to remain stable for the trill to maintain trilling and for the approximant to avoid friction. The video measurements in general show that a one-frame analysis is more useful for determining voice quality settings, as Painter has done; however, the depth measurement along with a video frame may provide enough evidence to describe vocal tract configuration for a speech sound segment.

It is necessary to observe a sequence of video frames in order to gain information about changes in vocal tract shape. An attempt to analyze a sequence of video data was made in this thesis, and it was found that the picture sequence could be matched-up with the aid of the acoustical measurements relatively accurately. Formant transitions between the vowel [i] and the pharyngeal approximant helped to correlate the video picture sequence with the corresponding audio output. It was shown that the visible formant information is accounted for by the oral movements and that the pharyngeal changes are likely represented by the upper formants. A detailed analysis of upper formant behaviour was not possible on the sequence examined here since the audio signal from the video tape had no acoustic information above 5000Hz. It is necessary to capture the audio separately onto DAT during the video recording rather than taking it from the video.

The same correlation of a sequence of video frames with audio output was applied to data of a Hebrew speaker. Since Hebrew has pharyngeal sounds it was expected that the video data would show clearly the changes in the vocal tract for pharyngeal sounds. Interesting information is obtained from this analysis. The video data of the Hebrew speaker did not correlate with audio output since there were far more frames of video pictures than audio data for this speaker. The results of this attempt to correlate a video sequence with audio data provides evidence that pharyngeal constriction is associated with a voice quality setting, since the Hebrew speaker had a relatively constricted pharynx most of the time.

This observation of pharyngeal voice quality in a native speaker of a language that uses post-velar sounds implies that if a Hebrew phonetician were used to produced the 'cardinal' examples there would be a pharyngeal quality on all of the other speech segments uttered. The phonetician used for the data in this study is not a native speaker of a language that uses post-velar sounds, however, he is trained with the IPA and Voice Quality taxonomies and provides a valid standard for developing a theory of pharyngeal behaviour. The phonetician also produced the segments with the intent of viewing behind the epiglottis to see what is happening at the larynx, which is not always practical with the average speaker.

The discussions thus far and the results of this thesis, characterize the acoustic distinctive features of cardinal pharyngeal sounds. This combined with the articulatory and perceptual distinctive features can be summarized as follows: Articulatorily, pharyngeal sounds are close to the glottis which suggests that there will be some affect on the vocal folds. This is shown by the change in pitch and other features observed in chapter 4. Also, the pharyngeal sounds are a combination of sphinctering in the horizontal plane and vertical displacement of the larynx. The larynx height feature was not specifically examined in this thesis with respect to the cardinal sounds; however, the language data evidence and the relationship of pharyngeal and raised larynx voice quality settings suggest that examining larynx height adjustments on cardinal pharyngeal sounds would be a worthwhile pursuit.

The perception of voicing in relation to pharyngeal constriction is likely to be confused because of the interaction of pharyngeal constriction and vocal fold vibration. This implies that there could be more distinctions among voiceless pharyngeal pairs in languages since the voicing distinction would be difficult to hear in pharyngeal sounds. The glottal stop is suggested to be a partner for the pharyngeal stop if a voicing contrast is used phonologically in a language that has both of these sounds. It is necessary to further investigate how voicing distinctions are determined in the phonologies of languages that use post-velar sounds in order to clarify the role of voicing in pharyngeal consonants.

Ultimately, classifying and understanding the acoustic results of pharyngeal behaviour and how humans perceive sounds made in this region of the vocal tract is useful to researchers who study languages that use these sounds and it can help in identifying voice quality. This combined with the proposed changes to the source-filter model can be applied to such products as speech-to-text processors, speaker identification and verification applications, and voice alteration applications.

References

AL-ANI, S. (1970). *Arabic Phonology: An Acoustical and Physiological Investigation. (Janua Linguarum 61).* Mouton: The Hague.

BESSELL, NICOLA J. (1993). The case for a pharyngeal glide: Evidence from Interior Salish. Proceedings of Western Conference on Linguistics. Fresno: California State University.

BESSELL, NICOLA J. (1992). *Towards a Phonetic and Phonological Typology of Post-Velar Articulation.* Ph.D. Dissertation, UBC. Single spaced version.

BICKLEY, C.A. AND STEVENS, K.N. (1991). Effects of vocal tract constriction on the glottal source: Data from voiced consonants. In *Laryngeal Function and Respiration* , eds. Thomas Baer, Clarence Sasaki and Katherine Harris, 239-253.

BOUGH, D., HEUER, R., SATALOGG, R., HILLS, J., AND CATER, J. (1996). Intrasubject variability of objective voice measures. *Journal of Voice* **10(2)** 166-174.

BUTCHER, A. AND AHMAD, K. (1987). Some acoustic and aerodynamic characteristics of pharyngeal consonants in Iraqi Arabic. *Phonetica* **44**, 156-172.

CATFORD, J.C. (1983). Pharyngeal and laryngeal sounds in Caucasian languages. In Bless, D.M. and Abbs, J.H. (editors), *Vocal Fold Physiology: Contemporary Research and Clinical Issues,* 344-350. San Diego: College Hill Press.

COOK, PERRY R. (1989). Synthesis of the singing voice using a physically parameterized model of the human vocal tract. *International Computer Music Conference*, Nov. 2-5. San Francisco, CA., 69-72.

DELATTRE,. D.C., LIBERMAN, A.M. AND COOPER, F.S. (1955). Acoustic loci and transistional cues for consonants. *JASA*, 27. 769-73.

ESLING, J.H. (1996). Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Aassociation,* 26, 65-88.

ESLING, J.H. (1996). Pharyngeal voice qualities: Auditory change with pitch and fibreoptic filming of aryepiglottic articulations. *Canadian Acoustics,* **24**(3), 22-25.

ESLING, J.H., HEAP, L.M., SNELL, R.C., AND DICKSON, B.C. (1994). Analysis of pitch dependence of pharyngeal, faucal, and larynx-height voice quality settings. *ICSLP* **94**, 1475-1478. Yokohama:Acoustic Society of Japan.

FORREST, K.,WEISMER, G., MILENKOVIC, P. AND DOUGALL, R.N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *JASA* **84**(1), 115-123.

FOURCIN, A.J. (1974). Laryngographic examination of vocal fold vibration. In Barry Wyke (ed) *Ventilatory and Phonatory Control Systems: An International Symposium*, 315-333. Oxford University Press.

GELFER, M.P. AND FENDEL, D.M. (1995). Compraisons of jitter, shimmer and signal-to-noise ratio from directly digitized versus taped voice samples. *Journal of Voice* **9**(4), 378-382.

HALLE, M., HUGHES, G.W. AND RADLEY, J.P.A., (1991). Acoustic properties of stop consonants. Readings in Clinical Spectrography of Speech (eds) Ronald J. Baken, Raymond G. Daniloff. 217-225.

HEAP, L.M. (1996). Acoustic interpretation of pharyngeal articulations. *Canadian Acoustics,* **24**(3), 26.

HEAP, L.M. (1997). Acoustic analysis of pharygeal approximant, fricatives, trills and stops. *The Working Papers of the Linguistics Circle of the Univeristy of Victoria.*

HONDA, K., HIROYUKI, H., ESTILL, J. AND TOHKURA, Y. (1991). Contribution of vocal tract shape to voice quality: MRI data and articulatory modeling. In Fujimura, O. and Hirano, M. (editors), *Vocal Fold Physiology: Voice Quality Control,* 23-38. Singular Publishing Group, Inc. San Diego, California.

HORII, Y. (1982). Jitter and shimmer differences among sustained vowel phonations. *Journal of Speech and Hearing Research* 25. 12-14.

JACOBSEN, W.H. JR (1969). Origin of Nootka Pharyngeals. *Int'l Journal of American Linguistics,* **35**(2), 125-153.

JOHNSON, K. (1997). *Acoustic and Auditory Phonetics.* Blackwell, Oxford.

KENT, R.D. AND READ, C. (1992). *The Acoustic Analysis of Speech.* San Diego, California: Singular Publishing Group, Inc.

KARNELL, M.P., HALL, K.D. AND LANDAHL, K.L. (1995). Comparison of fundamental frequency perturbation measurements among three analysis systems. *Journal of Voice* **9**(4), 383-385.

KODZASOV, S. V. (1987). Pharyngeal features in the Daghestan languages. *Proceedings of the XIth International Congress of Phonetic Sciences,* vol. 2. Tallinn: Academy of Sciences of the Estonian SSR. 142-144.

KRAUSS, M. (1979). "Na-Dene and Eskimo-Aleut," In Campbell and Mithun (editors), *The Languages of Native America.*

LADEFOGED, P. AND MADDIESON, I. (1996). *Sounds of the Worlds Languages.* Oxford: Blackwell.

LAUFER, A. (1979). The epiglottis as an articulator. *JIPA* **9**, 50-56.

LAUFER, A. (1996). Modifications of the IPA chart. *JIPA* **25**(2), 81-82.

LAVER, J. (1980). *The Phonetic Description of Voice Quality.* Cambridge University Press, Cambridge.

LEIBERMAN, P. AND BLUMSTIEN, S. (1991). *Speech Physiology, Speech Perception and Acoustic Phonetic.* Cambridge University Press: Cambridge.

LISKER, (1978). *Rapid vs. rabid:* A catalogue of acoustic features that may cue the distinction. Haskins Laboratories (New Haven, CT). *Status Report on Speech Research.* (No. SR-54), 127-132.

McCarthy, J.J. (1994). The phonetics and phonology of Semitic pharyngeals. In Patricia Keating (ed.), *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*, 191-233. Cambridge: Cambridge University Press.

Minifie, F., Hixon, T., and Williams, F. (1973). *Normal Aspects of Speech, Hearning and Language*. Prentice-Hall, New Jersey.

Omori, K., Kojima, H., Kakani, R., Slavit, D.H., and Blaugrund, S.M. (1997). Acoustic characteristics of rough voice: Subharmonics. Journal of Voice 11(1). 40-47.

Omori, K., Kacker, A., Carroll, L.M. Riley, W.D. and Blaugrund, S.M. (1996). Singing power ratio: Quantitative evaluation of singing voice quality. *Journal of Voice*. **10**(3), 228-235.

Painter, C., Watanabe, H. and Tsazuki, T. (1991). Laryngeal configuration of 99 disphonic patients. *Arch Otorhinolaryngol* **248**, 381-385.

Painter, C. (1991). The laryngeal vestibule, voice quality and paralinguistic markers. *Arch Otorhinolaryngol* **248**, 452-458.

Painter, C. (1986). The laryngeal vestibule and voice quality. *Arch Otorhinolaryngol*. **243**, 329-337.

Peppard, R.C. and Bless, D.M. (1990). A method for improving measurement reliability in laryngeal videostroboscopy. *Journal of Voice* **4**(3), 280-285.

Petersen, G.E. and Barney, H.L. (1952). Control methods used in the study of vowels. *J. Acoust. Soc. Amer.* 24: 175-84.

Rose, S. (1976). Lenition and Glottalization in Nootka. MA Thesis, University of Victoria.

Rossing, (1990). The Science of Sound. Addison-Wesley Publishing Co. Don Mills, Ontario.

Rossiter, David and Howard, David M. (1996). ALBERT: A real-time visual feedback computer tool for professional vocal development. *Journal of Voice* **10**(4), 321-336.

Scherer, Ronald (1996). Formants in Singing. *NCVS Status and Progress Report* **9**, 27-30.

Solomon, N.P., Luschei, E. and Liu, K. (1994). Fundamental frequency and tracheal pressure during three types of vocalizations elicited from anesthetized dogs. *NCVS Status and Progress Report* **7**, 139-137.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In, E.E. David and P.B. Denes (eds.), *Human Communication:A Unified View*, pp 51-56. New York: McGraw-Hill.

Stevens, K.N. (1991). Vocal-fold vibration for obstruent consonants. In Jan Gauffin and Britta Hammerberg (eds) *Vocal Fold Physiology*. Singular Publishing Group Inc.

SUNDBERG, J. (1987). *The Science of the Singing Voice*. Northern Illinois University Press, Dekalb, Illinois.

SUNDBERG, J. (1974). Articulatory interpretation of the singing formants. *JASA* **55**, 838-844.

THELWALL, ROBIN, AND SA'ADEDDIN, M. ADRAM. (1990). Arabic. *Journal of the International Phonetic Association*, 20(2), 37-40.

TITZE, I.R. (1993). *Principles of Voice Production.* Prentice-Hall. Englewood Cliffs: Jersey.

TITZE, INGO R. AND STORY, BRAD H. (1996). Acoustic interactions of the voice source with the lower vocal tract. *NCVS Status and Progress Report* **10**, 25-34.

TITZE, I.R. (1988). The physics of small-amplitude oscillation of the vocal folds. *JASA* **83**(4), 1536-1552.

TITZE, I.R., HORII, Y., SCHERER, R.C. (1987). Some technical considerations in voice perturbation measurements. Journal of Speech and Hearing Research, 30. 252-260.

TRAILL, A. (1986). The laryngeal sphincter as a phonatory mechanism in !Xoo Bushman. In Singer, R. and Lundy, J.K. (editors). *Variation, Culture and Evolution in African Populations*, 123-131.

TRASK, R.L. (1996). A Dictionary of Phonetics and Phonology. Routledge: New York.

TRIGO, L. (1991). On pharynx-larynx interactions. *Phonology*, 8. 113-136.

Visi-Pitch II Instruction Manual. Kay Elemetrics Corp. Pine Brook, New Jersey.

VERDOLINI, K., DRUKER, D., PALMER, P.M. AND SAMAWI, H. (1994). Physiological study of "Resonant Voice". *NCVS Status and Progress Report* **6**, 147-153.

YANAGISAWA, E., ESTILL, J., KMUCHA, S.T. AND LEDER, S.B. (1989). The contribution of aryepiglottic constriction to 'ringing' voice quality--a videolaryngoscopic study with acoustic analysis. *Journal of Voice* 3(4), 342-350.

IMAGE EVALUATION TEST TARGET (QA-3)

150mm

6"

APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989