THE ARABIC PHARYNGEAL APPROXIMANT

Ashraf Alkhairy
MIT Room 36-547, 50 Vasser Street, Cambridge, MA 02139, USA

ABSTRACT

The acoustic properties of the Arabic pharyngeal approximant in intervocallic sequences are studied. This sound has traditionally been classified as a voiced fricative, but our research indicates that it is an approximant, with no IPA symbol. The attenuation in the intensity of the consonant from the neighboring vowels is around 11 dB, and the fundamental frequency of the consonant is less than that of the vowels by 25 Hz. Study of the formant trajectories indicates that F1 increases and F3 decreases into the consonant for all contexts. The bandwidths remain unchanged from that of vowels, except in the context of /a/ for which both B2 and B3 increase. Examination of the estimated glottal volume velocity indicates an open quotient of 100%, and a spectral tilt that is highly dependent on the context with an average of 36 dB, and a value of 21 dB for the context of /a/.

1. INTRODUCTION

The Arabic language consists of a relatively large number of consonants produced with constriction in the uvular and pharyngeal regions, in addition to locations in the oral cavity between the velum and the lips. The articulation and characteristics of the pharyngeal consonants is still not clearly understood because of lack of research, and the difficulty of measurements for these regions. These consonants involve articulation of the tongue root, back wall of the pharynx, and possibly the epiglottis and other structures in the lower vocal tract [1,2].

Traditionally, Arabic has been considered to have two pharyngeal consonants, a voiced and a voiceless fricative [1,2,3,4]. The "voiced fricative" has also been classified as a voiceless stop, voiced glide, and an approximant [1,2,3,4,5,6,7]. While these discrepancies in categorization of the voiced fricative may be attributed to differences in dialect or context, we believe that this not to be the case. Existing studies have relied on X-ray images, and spectrographic plots to reach their conclusions. In all cases, the images are not clear enough to make distinctions as far as manners of articulation are concerned, and the plots are not of adequate resolution to make appropriate judgements. In this paper we make a detailed study of the nature of the consonant and conclude that it is indeed an approximant, and a new IPA symbol is required.

Classification of a sound, as far as manner of articulation is concerned is important in its acoustic analysis, as well as modeling, synthesis, and the determination of perceptual cues. An approximant is produced with two articulators close to one another, but without a narrow enough constriction to generate turbulence. In contrast, the area and shape of the cavity between the two constricting articulators in a fricative are such as to produce turbulent airflow. In the first case, the acoustic signal is deterministic, whereas in the second case the pressure waveform is stochastic, comprising of both deterministic and pseudo-random components.

In addition to the study of the nature of the sound, a comprehensive acoustic analysis is conducted. The analysis is directed towards investigation of acoustic variables that aid in articulatory and voice source modeling of the pharyngeal approximant, as well as in determining acoustic cues for their perception.

This paper is divided into six sections. Section 2 discusses the speech samples that were used for analysis. Section 3 investigates the nature of the sound. Section 4 studies the formants. Section 5 presents results related to the glottal characteristics. Section 6 concludes the paper and provides suggestions for future research.

2. EXPERIMENTS

The Modern Standard Arabic (MSA) was used to conduct our study. MSA is employed by the entire Arab population for purposes of official communication, teaching, and literature. The speakers consisted of three educated adult male Saudi natives, with no speech or hearing impairments, and without phonetic training. Speech samples were recorded in a sound-treated room into a digital computer with sampling rate of 10kHz.

The linguistic material for the experiments were of the form /z/VCV/z/, where C represents the approximant, and V denotes the vowels. The approximant was in geminate form to allow more accurate measurements, and the vowels consisted of /a/, /u/, and /i/. The VCV sequence were embedded within /z/ to provide the most natural context. In a number of cases the properties of the consonant are compared to that of the vowels. To accomplish this, reference utterances consisting of /z/V/z/ were also recorded, where V represents the vowels in long form. The long versions provide better circumstances for the study of the vowel steady state characteristics.

3. THE NATURE OF THE SOUND

A number of analysis methods were used to determine the classification of the consonant under consideration. Both time-domain and frequency-domain analysis results indicate that the sound is produced without the generation of any turbulence noise and without nasalization; and hence should be classified as an approximant - with a new IPA symbol. The analysis methods used are pressure waveform examination, intensity analysis, spectrogram, and spectrum.

3.1. The Pressure Waveform

A first indicator of the nature of the sound, although inconclusive, is obtained from visual inspection of the waveforms in the time-domain. An example is shown in Figure 1 (Top). The plot is free from random noise and is representative of all the waveforms.

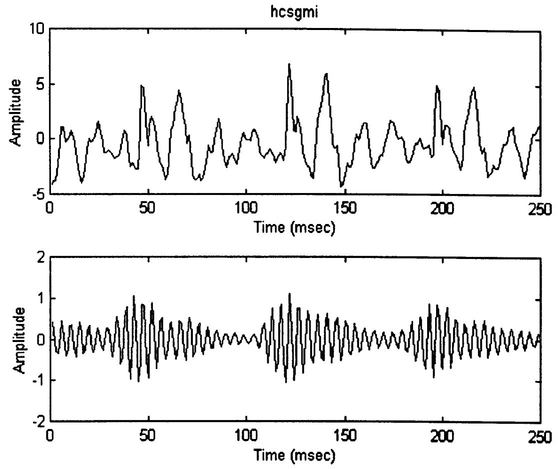

Figure 1: Top: The pressure waveform of the consonant in the context of /i/. Bottom: The bandpass filtered waveform.

Absence of turbulence noise in the consonant can be confirmed by inspection of the bandpass filtered waveforms centered on the third formant, with a bandwidth of 600 Hz [8]. This methodology was applied to the waveforms for all speakers and contexts. The resulting filtered signals consist of damped sinusoids synchronized with the original waveforms; a typical example of which is shown in Figure 1 (Bottom). We thus infer that the sound is an approximant rather than a voiced fricative.

3.2. Intensity Analysis

An approximant is likely to have higher energy in comparison to voiced fricatives because the constriction in the vocal tract for an approximant is less obstructive. Although voiced fricatives are generally attenuated by 20 dB, no exact ranges are known for approximants and voiced fricatives. To examine intensity characteristics of our approximant, differences in intensity between the steady state values of the approximant and that of the following vowels were computed. The mean values are as follows : /a/ context (12 dB), /u/ context (14 dB), /i/ context (11 dB), and overall average (12 dB). These results support the observation that the consonant is an approximant.

3.3. The Spectrogram

To study the global characteristics of the approximant, digital spectrograms of the utterances were computed using a Hanning window of length 7.97 ms, an overlap of 97%, length of 512 points, bandwidth of 180 Hz, and no pre-emphasis. A representative example is displayed in Figure 2. The spectrogram has regular striations and is not diffused, indicating continuous voicing and lack of turbulence noise.

3.4. The Spectrum

Wideband spectra at the steady state positions of the approximant were computed using a Hanning window of size 6.4 ms with no pre-emphasis to determine the general spectral characteristics of the approximant. As the representative example plotted in Figure 3 indicates, the energy is clustered in the mid-frequency range, and no signs of anti-resonance are present.

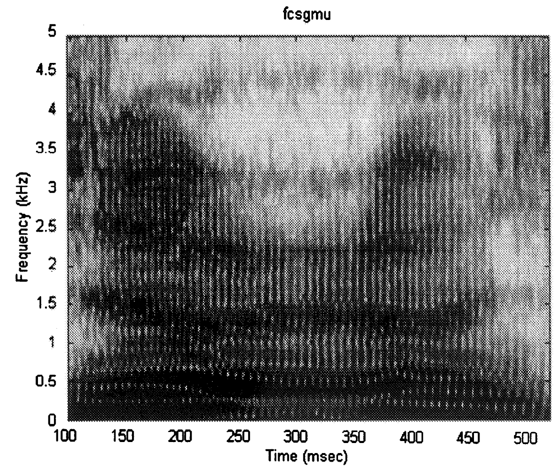

Figure 2: Spectrogram of the consonant in the context of /u/.

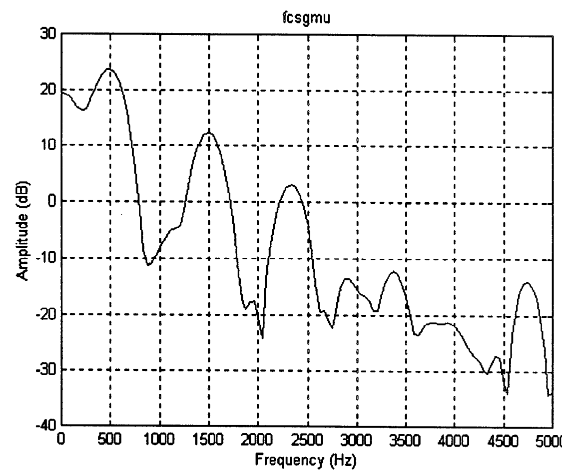

Figure 3: Wideband spectrum of the consonant within /u/.

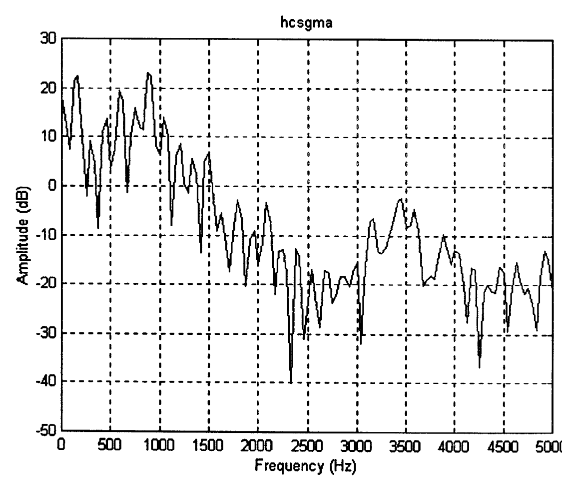

Figure 4: Narrowband spectrum of the consonant within /a/.

To examine aspects related to the duty-cycle of the glottal source from the spectrum, narrowband spectra were computed with a Hanning window of size 25.6 ms. As the illustrative example in Figure 4 indicates, the first harmonic has a large amplitude, implying a high open quotient. This is examined in greater detail in a latter section.

4. THE FORMANTS

The primary parametric identifier of a phoneme are formant frequencies; the first three of which can be predicted based on articulatory models in many cases [9,10]. In this section we present results related to the first five formant frequencies and bandwidths. Although issues related to bandwidth and higher formants have not been investigated in-depth in the literature - as far as modeling, synthesis, and recognition are concerned, we have included them in this study to provide more comprehensive results that can be used in the future.

The formant frequencies and bandwidths were computed for all word sequences using the synchronous covariance LPC method, with the pre-emphasis of 0.95 and filter order of 12 (sometimes 10 or 14). Although the LPC method has been criticized for underestimating the bandwidth, it is the most practical accurate method to date [11]. To ensure accuracy in bandwidth estimation, we validated the resulting LPC values with the measurements of the spectra.

The statistical analysis results can be presented as either absolute values or as deviations from the references. We can either compute the statistics of the formants at the steady state times of the approximant within particular vowel contexts, or compute the statistics of the difference between the formants of the approximant within a particular vowel context and the formants of the corresponding long vowel in a reference utterance.

Tables 1 and 2 present the average deviations in formant frequencies and bandwidths respectively. Statistics of differences, rather than absolute values, are tabulated because they are more suitable for modeling purposes, and because significant differences can be distinguished from insignificant ones more readily. Cells with "*" indicate statistically insignificant changes.

	/a/ context	/u/ context	/i/ context	Average
Δ F1	119	41	124	95
Δ F2	78*	277	-434	-26*
Δ F3	-231	-351	-268	-283
Δ F4	-23*	-11*	-95*	-43*
Δ F5	-243*	334*	-117*	-9*

Table 1: Deviations in formant frequencies (Hz).

	/a/ context	/u/ context	/i/ context	Average
Δ B1	-8*	-10	2*	-5*
Δ B2	66	22*	-73*	15*
Δ B3	109	-27*	-49*	11*
Δ B4	69*	86	38*	67*
Δ B5	-5*	-56*	-64*	-42*

Table 2: Deviations in bandwidths (Hz).

The data in Table 1 clearly indicates that the first formant of the approximant is larger than that of all the vowels, and that the third formant is smaller. This is in agreement with modeling results [6]. In the context of /i/, however, the second formant is smaller. Further research in acoustic phonetics is required to explain the result.

Bandwidth values for the approximant in the context of /a/ indicate a widening of the second and third formants when compared to that of the adjacent vowels. This feature may aid in distinguishing the approximant from the closely sounding back vowel, /a/; perceptual studies can validate this hypothesis. The bandwidth data is in disagreement with predicted model-based results, and further research is needed to develop more accurate models.

5. THE GLOTTAL CHARACTERISTICS

Conventional acoustic analysis of a sound does not usually include investigation of its glottal attributes. This is because they are considered secondary in importance to formant characteristics, and until lately, no reliable and accurate algorithms have existed for determining the glottal features [12].

A number of recent publications, however, have demonstrated the importance of glottal characteristics in producing high-quality synthesized sounds [8]. In addition, aspects related to articulatory-acoustic models for the glottal waveform have received considerable attention, and are likely to encourage future investigations into analysis of the glottal characteristics.

In this paper, we examine the glottal attributes of the pharyngeal approximant in considerable detail. First, we examine the fundamental frequency of the approximant. Next, we investigate temporal and spectral characteristics of the glottal waveform.

5.1. The Fundamental Frequency

To study fundamental frequency trajectories statistically, time duration between impulse markers of the words were computed and reciprocated to compute the fundamental frequency in Hz at each pitch period. Values at the steady states of the consonant, preceding and following vowels were selected, while making sure that jitters are avoided. Means of the differences between these values and fundamental frequencies of the corresponding long vowels spoken by the same speakers were computed and are listed in Table 3.

	/a/ context	/u/ context	/i/ context	Average
Pre-vowel	-7	0	-1	-3
Consonant	-27	-24	-37	-29
Post-vowel	-12	-13	-12	-12

Table 3: Deviations in fundamental frequencies (Hz).

Examination of the results indicate that the fundamental frequencies of the approximant are noticeably lower than that of the vowels by around 30 Hz. The values at the consonant, when compared to that of the adjacent vowels, indicates that this is a phoneme characteristic rather than a suprasegmental artifact. The deviations for the context of front vowels is larger than that of back vowels because the fundamental frequencies for front vowels are higher [13]. The fundamental frequencies for /a/, /u/ and /i/ are 168, 173, and 180 respectively. Such changes in fundamental frequency can be expected because of the proximity of the constriction for a pharyngeal approximant to the vocal folds.

5.2. The Glottal Waveform

Temporal and spectral characteristics of the glottal waveform were directly computed from the calculated glottal waveform at the steady-state positions of the approximant. A decomposition method was used to estimate the glottal signal [14]. An example, of the obtained waveform is shown in Figure 5.

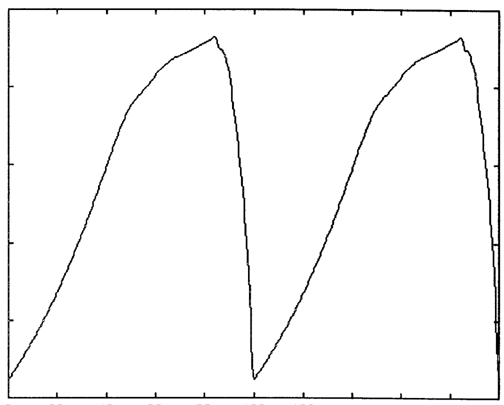

Figure 5: The glottal waveform for the approximant in the (two pitch periods shown).

The graph shows a low-bandwidth signal with an open quotient (OQ) of 100% and a large speed quotient (SQ). These results are representative of all occurrences of the approximant, as indicated by the statistical analysis shown in Table 4. The last row in the table lists the spectral tilt (TL), which measures the attenuation (in dB) between the amplitudes at 300 and 3000 Hz. The average spectral tilt is around 40 Hz, which is the same as that of vowels, and is in agreement with the attribute of ideal models. The spectral tilt for the approximant in the context of /a/, is however low at around 20 dB. This may be a feature that is used to enhance the distinction of the approximant from the closely sounding vowel /a/.

	/a/ context	/u/ context	/i/ context	Average
OQ (%)	100	100	100	100
SQ (%)	417	356	305	359
TL (dB)	21	49	39	36

Table 4: Glottal characteristics.

6. CONCLUSION

This paper has investigated the acoustic properties of the Arabic pharyngeal approximant. While some of our findings are in agreement with existing predictions, others need to be examined in light of further research. Improved articulatory models need to be developed to explain the formant trajectory for the consonant in the context of /i/. Also, loses in articulatory models need to be better understood to explain the bandwidth changes that occur when the approximant is spoken in the context of /a/.

Significant modeling research can be conducted in view of the results provided concerning the glottal characteristics. Variations in the open quotient and spectral tilt can be explained based on future modeling approaches.

A number of perceptual studies can be conducted to examine the acoustic features that effect either the intelligibility or the quality of the sound. The contribution of bandwidth, open quotient, speed quotient and spectral tilt form major factors that may significantly effect the quality of the sound.

ACKNOWLEDGMENTS

The author was sponsored by KACST during his sabbatical leave at MIT, at which time this work was conducted. He wishes to thank Kenneth Stevens for numerous discussions.

REFERENCES

[1] Al-Ani, S. (1970). Arabic Phonology. Mouton, The Hague.

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

[3] Delattre, P. (1971). Pharyngeal Features in the Consonants of Arabic, German, Spanish, and American English. Phonetics 23: 129-155.

[4] Butcher, A. and Ahmad, K. (1987). Some Acoustic and Aerodynamic Characteristics of Pharyngeal Consonants in Iraqi Arabic. Phonetica 44: 156-172.

[5] Al-Ani, S. (1967). An Acoustical and Physiological Investigation of the Arabic /*/. Acts of the Xth International Congress on Linguistics. Bucharest.

[6] Klatt D.H. and Stevens K.N. (1969). " Pharyngeal Consonants", QPR no. 93, RLE, MIT, 207-215

[7] Alwan, A. (1986). Acoustic and Perceptual Correlates of Pharyngeal and Uvular Consonants, unpublished Master's thesis, MIT.

[8] Klatt, D. and Klatt, L. (1990). Analysis, Synthesis and Perception of Voice Quality Variations Among Female and Male Talkers. Journal of the Acoustical Society of America, 87:820-857.

[9] Fant, G. (1970). Acoustic Theory of Speech Production. Mounton.

[10] Stevens, K. (1998). Acoustic Phonetics. To be published.

[11] Wakita, H. (1980). New Methods of Analysis in Speech Acousics. Phonetics 37:87-108.

[12] Fant, G. (1993). Some Problems in Voice Source Analysis. Speech Communication, 13:7-22.

[13] Lehiste, I. and Peterson, G. (1961). Transitions, Glides and Diphthongs. Journal of the Acoustical Society of America, 33:268-277

[14] Alkhairy, A. (1999). An Algorithm for Glottal Volume Velocity Estimation. Proceedings of ICASSP, March 1999.