Place of articulation cues for voiced and voiceless plosives and fricatives in syllable-initial position

Willa S. Chen and Abeer Alwan
Department of Electrical Engineering, UCLA, Los Angeles, CA 90077

ABSTRACT

In this paper, the acoustic correlates of the labial and alveolar place of articulation for both plosive and fricative consonants are investigated, and the results are analyzed in terms of vowel context, voicing and manner of articulation. Several measurements, including formant and noise measurements, are reported for CVs spoken by two male and two female talkers. It was found that the spectral amplitude of frication noise relative to F1 at vowel onset results in 84% or better correct classification for the fricatives in 3 vowel contexts. For plosives, a measure which quantifies the amplitude of noise at high frequencies relative to F1 at vowel onset (Av-Ahi [8]) resulted in 81 % or better correct classification in the three vowel contexts. Formant frequency cues, on the other hand, were not reliable measures for all vowel contexts.

1. INTRODUCTION

Various studies have attempted to find invariant acoustic cues for the place of articulation feature for plosives and fricatives. For plosives, several studies have focused on the role of the formant transitions (especially the second formant, F2) in acoustic classification and perception of place of articulation (e.g., [2], [5], and [7].) In [9], the authors were able to classify alveolars and labials across 9 vowel contexts when applying locus equation analysis of F2 onset versus F2 steady state frequency for voiced plosives from ten male and ten female talkers. Other studies analyzed the temporal and spectral characteristics of the burst and aspiration segments (e.g., [1] and [10].) In [4], the authors reported on perceptual experiments using synthetic voiceless plosives and showed that the burst amplitude relative to vowel onset in the F4/F5 region is a cue that distinguishes between voiceless alveolars and labials. Stevens et al. [8] measured F1 and F2 frequencies and three relative spectral amplitudes of the burst (Av-Ahi, Ahi-A23 and Av-maxA23) for word-initial plosives as spoken by two male and two female talkers. The study showed the importance of noise measures as correlates of place of articulation.

Fricative studies have shown differences in the amplitude of noise as a function of place. For example, [6] showed that the difference between the minimum amplitude value between 0-2 kHz and the maximum amplitude between .5-17 kHz is higher for /*s*/ than /*ʃ*/. In addition, perceptual studies have demonstrated the perceptual importance of the amplitude of noise relative to vowel onset in the F3 and F5 regions [3].

This paper attempts to find invariant acoustic cues for place of articulation for both fricatives and plosives. The results are analyzed in terms of vowel context, voicing and manner of articulation.

2. METHODS

2.1 Stimuli

The stimuli analyzed consisted of consonant-vowel syllables (CVs), where C was one of /b,d,p,t,f,s,v,z/ and V was one of /a,i,u/ as spoken by 2 male and 2 female talkers, 4 repetitions each, of American English. The data were sampled at 16 kHz.

2.2 Analysis

Two types of acoustic measurements were made: formant frequency and noise measurements. The first set of measurements included the duration, frequency and amplitude of the formants. The second set included measures which quantify the duration and spectral characteristics of the noisy segments (frication, burst, and aspiration).

Formant Measurements

Temporal and spectral measurements of F1, F2 and F3 were obtained manually from the time waveform, wide-band spectrogram, short-time DFT and LPC spectra using Matlab. Spectral analysis was done by analyzing 20 ms (for male talkers) or 15 ms (for female talkers) frames of speech which were windowed with a Hamming window and overlapped by 2.5 ms. Pre-emphasis was used. The order of the LPC spectrum was between 8-12 (depending on the talker). The onset of the vowel was defined as the center point of the frame which shows an abrupt change in the waveform and in spectral features. The end of the formant transitions, chosen automatically, was defined as the frame during which the rate of change of the formant frequency

fell to less than 5 Hz per 2.5 ms, and the average rate of change for the next 5 frames was also less than 5 Hz per 2.5 ms. Since locating the transition offset is prone to error, a third point, called the steady-state, was measured at 95 ms after vowel onset. At each of these three points (vowel onset, offset and steady-state), the corresponding time, frequency and amplitude of F1, F2, and F3 were recorded. The difference between vowel onset and steady-state formant frequencies are referred to as F1Δ, F2Δ and F3Δ. The locus equations [9] were also obtained for all CV syllables using F2 onset and steady-state frequencies.

Noise Measurements

• **Burst, VOT, and frication noise duration**: The durations of the burst, VOT, frication noise and aspiration noise were obtained manually from the time waveform and wide-band spectrogram of the signals.

• **Relative Spectral Measurements** Burst and noise spectra were estimated using the Welch periodogram method using a 6 ms window with a 3 ms overlap. If the duration of the burst is shorter than 9 ms, then a 3 ms window with a 1.5 ms of overlap was used. The first three of the following relative spectral amplitude measurements are similar to those suggested by Stevens et al. for plosive consonants [8]. **(1)Ahi-A23**: **Ahi** represents the peak spectrum amplitude of the combined burst and aspiration region (hereafter referred to as BUAS) in plosives, and frication noise in fricatives in the frequency range above 3500 Hz for females and 3000 Hz for males. **A23** is the average spectral amplitude in the F2-F3 range. This measurement characterizes the spectral tilt of the burst/aspiration region [8]. **(2)Av-Ahi**: **Av** represents the spectral amplitude peak around F1 at vowel onset. **Av-Ahi** is a measure of the amplitude of the noise at high frequencies relative to the vowel[8]. **(3)Av-maxA23**: **maxA23** is the maximum of the BUAS spectral amplitude around the F2-F3 region. **Av-maxA23** was calculated only for plosives to determine mid-frequency spectral prominence [8]. **(4)Av4-A45** Similar to the definition for **Av**& **A23**, **Av4** represents the peak amplitude of the spectrum around F4 at vowel onset. **A45** is the average amplitude of the BUAS or noise in the F4-F5 region. This measurement was inspired by studies [3] and [4]. **(5)Av4-maxA45**: This quantity is similar to **Av4-A45** except that we calculate the maximum amplitude of the BUAS/noise in the F4-F5 region. **(6)Amid-Avmid**: **Amid** is the average of the BUAS spectral amplitude, and **Avmid** is the average of the vowel onset spectral amplitude between 3200 and 4800 Hz. This quantity is measured only for plosives to characterize the difference in amplitude between the burst/aspiration and the vowel at mid-frequencies. **(7)Av-Anoise**: This measurement was measured for fricatives only. **Anoise** represents the average amplitude of the entire noise spectrum.

	F1 onset	F2 onset	F3 onset
/fa,sa/	**100%**	84%	81%
/va,za/	**100%**	84%	78%
/fi,si/	59%	69%	62%
/vi,zi/	59%	69%	66%
/fu,su/	63%	84%	75%
/vu,zu/	69%	81%	75%
/ba,da/	94%	97%	75%
/pa,ta/	66%	69%	63%
/bi,di/	69%	69%	69%
/pi,ti/	69%	56%	63%
/bu,du/	81%	88%	88%
/pu,tu/	59%	88%	66%
/b,d/	65%	78%	68%
/p,t/	56%	68%	55%
/f,s/	67%	68%	70%
/v,z/	68%	72%	67%

Table 1: Percent correct classification using F1, F2 or F3 onset values for individual CVs and for each consonant pair. Numbers in boldface represent the highest percent correct classification for place of articulation using a particular feature.

3. RESULTS

Tables 1 - 5 summarize the results in terms of percent correct classification for all syllables. Numbers in bold face represent the highest percent correct classification for place of articulation using a particular feature. For example, in the /fa,sa/ case, F1 onset values separate well (100% correct classification) between the labial and alveolar place of articulation, but not in the /fi,si/ case (59%). Formant frequency measures (especially F1 and F2 onset values, F2Δ and F3Δ) cue place for several /Ca/ and /Cu/ syllables but not for /Ci/ syllables. Relative noise spectral measurements seem to provide better discrimination cues for place of articulation than formants. For example, the Av-Anoise (for fricatives) and Av-Ahi (for plosives) measures resulted in more than 81% correct place classification for both voiced and voiceless consonants and for both vowel-dependent and vowel-independent measures. VOT (not shown in the tables) does not cue place for plosives. Burst duration signals place of articulation for plosives in /Ca/ and /Ci/ syllables (above 81% correct classification) but not for /Cu/ syllables; noise duration,

	F1Δ	F2Δ	F3Δ
/fa,sa/	78%	93%	96%
/va,za/	72%	94%	84%
/fi,si/	59%	75%	69%
/vi,zi/	69%	72%	88%
/fu,su/	63%	72%	78%
/vu,zu/	59%	59%	84%
/ba,da/	78%	**100%**	97%
/pa,ta/	75%	78%	63%
/bi,di/	66%	59%	75%
/pi,ti/	66%	66%	72%
/bu,du/	78%	75%	**91%**
/pu,tu/	66%	78%	75%
/b,d/	61%	74%	78%
/p,t/	61%	67%	56%
/f,s/	60%	68%	73%
/v,z/	57%	68%	70%

Table 2: Percent correct classification using F1Δ, F2Δ or F3Δ.

on the other hand, does signal some place information for fricatives in all vowel contexts (above 75% correct classification). Measurements that showed 100% classification for place include F1 onset frequency for /fa,sa/, /va,za/, F2Δ for /ba,da/, Av-Ahi for /pu,tu/ and Av-Anoise for /fu,su/.

To examine locus equations [9] in different vowel contexts, the slope and Y-intercept values are calculated for each CV. Bar graphs of the slope and Y-intercept values are shown in Figures 1- 2. For the plosives, F2 slope measures cue place for voiced plosives but not for the voiceless ones (with the exception of /pa,ta/.) Y-intercept values are higher for the voiceless alveolars in the /Ci/ and /Cu/ contexts, but these values are higher for the voiced labials in the /Ca/ and /Ci/ contexts. Fricatives do not show a consistent pattern in the slope values when considering individual vowel contexts and the Y-intercept values are higher for the alveolars except for /fa,sa/ and /fi,si/. However, if the slope and Y-intercept values are obtained using CVs across *all* vowel contexts, alveolars consistently show smaller (or flatter) slope values and larger Y-intercepts. Slopes and Y-intercept values are summarized in Table 6.

4. SUMMARY

This study attempts to find invariant acoustic cues for place of articulation and investigate the effect of vowel context, voicing, and manner of articulation. Several acoustic measurements (formant and noise) were made for both voiced and voiceless plosives and fricatives in three vowel contexts. There was no single measure that can cue place for both plosives and fricatives and in all vowel contexts.

Results showed that noise measurements are reliable cues for place of articulation. For example, the Av-Anoise measure for fricatives and Av-Ahi

	Ahi-A23	Av-Ahi	Av4-A45	Av4-maxA45
/fa,sa/	66%	75%	75%	78%
/va,za/	66%	66%	72%	78%
/fi,si/	59%	59%	78%	**94%**
/vi,zi/	63%	63%	78%	81%
/fu,su/	72%	72%	84%	88%
/vu,zu/	66%	69%	69%	88%
/ba,da/	59%	91%	84%	84%
/pa,ta/	91%	84%	72%	75%
/bi,di/	75%	94%	69%	75%
/pi,ti/	**97%**	81%	78%	81%
/bu,du/	78%	**91%**	66%	69%
/pu,tu/	94%	**100%**	84%	91%
/b,d/	68%	**90%**	64%	72%
/p,t/	**89%**	86%	79%	79%
/f,s/	59%	67%	76%	85%
/v,z/	65%	64%	71%	81%

Table 3: Percent correct classification using Ahi-A23, Av-Ahi, Av4-A45 or Av4-maxA45.

	Av-Anoise	Noise Duration
/fa,sa/	97%	75%
/va,za/	91%	75%
/fi,si/	84%	78%
/vi,zi/	**97%**	91%
/fu,su/	**100%**	75%
/vu,zu/	**97%**	75%
/f,s/	**90%**	68%
/v,z/	**90%**	70%

Table 4: Percent correct classification using Av-Anoise or Noise Duration.

	Amid-Avmid	Av-maxA23	Burst Duration
/ba,da/	78%	81%	81%
/pa,ta/	75%	66%	**97%**
/bi,di/	84%	78%	84%
/pi,ti/	75%	72%	81%
/bu,du/	78%	72%	59%
/pu,tu/	88%	75%	69%
/b,d/	76%	74%	72%
/p,t/	77%	60%	68%

Table 5: Percent correct classification using Amid-Avmid, Av-maxA23, or Burst Duration.

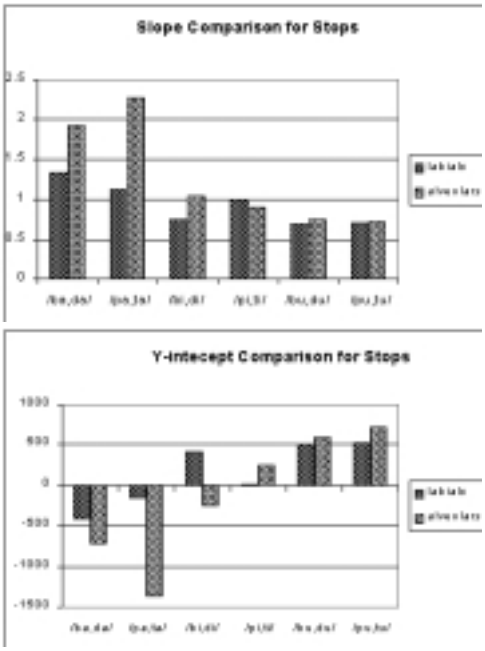

Figure 1: Slopes and Y-intercept Values for Plosives

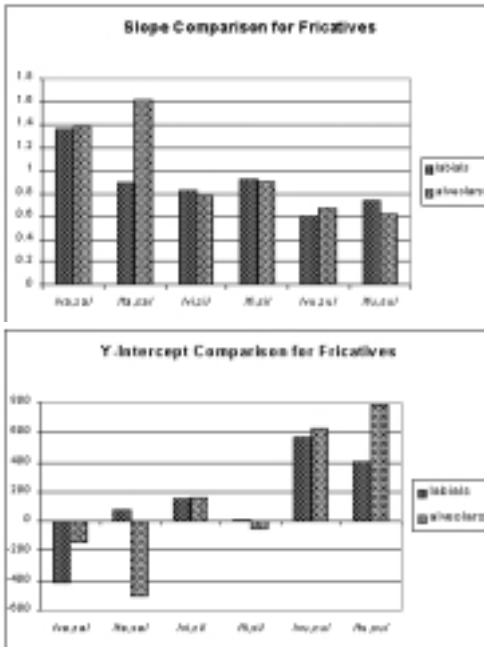

Figure 2: Slopes and Y-intercept Values for Fricatives

	/b/	/d/	/p/	/t/
slope	0.8232	0.524	0.9476	0.8458
Y-intercept	220.62	1030.0	113.04	409.81
	/v/	/z/	/f/	/s/
slope	0.7218	0.4315	0.8559	0.5890
Y-intercept	378.75	1008.5	171.92	754.34

Table 6: Slope and Y-intercept Values for Plosives and Fricatives when Considering all Vowel Contexts

measure [8] for plosives are robust cues for place of articulation for each vowel context and across all vowel contexts. Formant information appeared to cue place, to some extent, only for /Ca/ and /Cu/ syllables but not for /Ci/ syllables for both plosives and fricatives. Locus equation information [9] cues place for voiced plosives by slope values and voiceless plosives and voiced fricatives by Y-intercept values. However, when analyzed across all vowel contexts, results showed that slopes are flatter and Y-intercepts are higher for alveolars for these syllable initial fricatives and plosives regardless of voicing.
Classification scores shown in Tables 1- 5 can be further analyzed by separating out male and female speech. Future studies will examine the perceptual role of these acoustic cues in signaling place and examine a larger database of utterances, talkers, and speaking efforts.

Acknowledgments

This work was supported in part by NIH-NIDCD grant 1R29-DC02033-01A1. We thank Marcia Chen for her help in data analysis.

References

[1] S.E. Blumstein and K.N. Stevens. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *JASA*, 66:1001–1017, 1979.

[2] P.C. Delattre, A.M. Liberman, and F.S. Cooper. Acoustic loci and transitional cues for consonants. *JASA*, 27:769–773, 1955.

[3] M. Hedrick and R. N. Ohde. Effect of relative amplitude of frication on perception of place of articulation. *JASA*, 94(4):2005–2026, Oct. 1993.

[4] M.S. Hedrick, L. Schulte, and W. Jesteadt. Effect of relative and overall amplitude on perception of voiceless stop consonants by listeners with normal and impaired hearing. *JASA*, 98(3):1292–1303, Sep. 1995.

[5] D. Kewley-Port. Time varying features as correlates of place of articulation in stop consonants. *JASA*, 73:322–335, 1983.

[6] C.H. Shadle and S.J. Mair. Quantifying spectral characteristics of fricatives. *ICSLP Proc.*, 3:1521–1524, 1996.

[7] K.N. Stevens, A.S. House, and A.P. Paul. Acoustical description of syllabic nuclei: an interpretation in terms of a dynamic model of articulation. *JASA*, 40:123–132, 1966.

[8] K.N. Stevens, S.Y. Manuel, and M. Metthies. Revisiting place of articulation measures for stop consonants : Implications for models of consonant production. *ICPhS Proc. 1999*, 2:1117 1120, 1999.

[9] H.M. Sussman, H.A. McCaffrey, and S.A. Matthews. An investigation of locus equations as a source of relational invariance for stop place categorization. *JASA*, 90(3):1309–1325, 1991.

[10] V. Zue. *Acoustic characteristics of stop consonants: a controlled study.* PhD thesis, MIT, 1976.