EFFECT OF ARTICULATION CONDITION ON CHILDREN'S ACOUSTIC CUES FOR BILABIAL PLACE

Megan Hodge, Christine Huang, Janka Hegedus and Melanie Campbell
Department of Speech Pathology and Audiology, University of Alberta, 2-70 Corbett Hall, Edmonton, AB T6G2G4

1. INTRODUCTION

Energy patterns in the speech acoustic signal, and how these patterns change over time as speech muscle groups produce the movements of speech, can reveal information about what the speech mechanism is doing during speech production. Acoustic spectral variations are related to the shape of the vocal tract. Cues contained in the speech spectra provide information about the place and manner of articulation of speech sounds[1]. Cues to place of stop articulation are captured in parameters of the burst spectrum (mean or spectral moment 1 and skewness or spectral moment 3) and in second formant (F2) frequency transitions. It is also well established that voice onset time (VOT) varies with place of articulation[2] .

The purpose of this study was to determine if these four acoustic cues (stop burst spectral moments 1 and 3, F2 onset and VOT) for bilabial place are stable across articulatory condition, and if these are influenced by the presence of /ɹ/ following the stop. We examined measures of four acoustic cues for two stop consonant places (bilabial and lingua-alveolar) from productions of three children who represent three different articulatory conditions: a girl with normal facial movement and typical speech production (CNM) and two girls with facial paralysis, who differed in the nature of their articulatory compensation. One child had congenital, non-progressive facial paralysis and had undergone facial animation surgery at age 7 years (CP1). Prior to this surgery she used a lingua-dental contact for bilabial stop targets. Following the surgery she had been taught to use a labiodental dental contact for bilabial stop targets. The second child (CP2) had a progressive facial paralysis that was complete by age 14 months. She used a lingua-dental contact for bilabial stop targets.

Previous work[3,4] indicated that all three children (CNM, CP1 and CP2) showed F2 CV onsets that fit the typical pattern, that is, lower F2 onset frequencies for bilabial than for lingua-alveolar stops. Analysis of audio recordings from CNM and CP1 revealed that both children showed the expected VOT pattern, that is, longer duration for lingua-alveolar than bilabial stops in both CV and CɹV contexts. However, results for spectral moments 1 and 3 for CP1 showed the reverse of the expected pattern in. Rather than showing a lower spectral mean and a more positive skewness for bilabial than lingua-alveolar stops, her spectral means for bilabial place were higher and spectral skewness less positive than for lingua-alveolar place in both CV and in CɹV contexts. The child with typical speech production (CNM) showed the expected pattern for burst spectral mean and skewness in CV contexts but showed the reverse pattern, like CP1, in CɹV contexts. In this study, recordings from CP2 were analysed and are compared with previous findings for CNM and CP1.

2. METHOD

2.1 Speakers

At the time of recording, the three girls ranged in age from 9-10 years and had hearing within or corrected to within normal limits. All had age appropriate receptive and expressive language abilities and school performance. CP1 had a diagnosis of Moebius syndrome associated with complete congenital paralysis of facial muscles, including the lips. CP2 had a diagnosis of early onset facioscapulohumeral muscular dystrophy, a progressive condition that affects muscles of the face and shoulder girdle primarily but can involve other muscle groups (tongue, soft palate, pelvic girdle) as the disease progresses.

2.2 Recording of Stimuli

The stimulus set consisted of 16 word pairs minimally contrastive in place of stop consonant articulation. These phonemes were contrasted in initial position in the stimulus words: /p/ versus /t/, and /b/ versus /d/. For 8 of the word pairs, the target sounds occurred in singleton context (CV) with no adjacent consonants. Another 8 word pairs contained the target sounds followed by /ɹ/ (C/ɹ/V). The child produced each of the words in the carrier phrase, "I can say ___ again." Recordings of the word pairs were elicited using an imitative model. The child repeated words presented by the examiner while viewing a semantically relevant picture. The recordings were obtained in a sound booth using a Panasonic 455 S-VHS video camera-recorder unit and an external Sony 150T electret condenser microphone that was positioned approximately 6 cm below the child's chin.

2.3 Measurement of Acoustic Cues

CSpeechSP[5] was used to measure the four acoustic cues of interest. To measure VOT and F2 transition onset, digital audio files of the child's recordings of each stimulus phrase were made using a sampling rate of 22 kHz with a quantization size of 16 bits and a low pass cut-off filter of 9 kHz. The target word was extracted from the files with a 50-ms lead. VOT was obtained by measuring the time (ms) from the first transient noise burst for the stop release to the zero-crossing of the first positive-going, large amplitude, well-formed repetitive

wave shape indicating vocal fold vibration. F2 transition onset was measured by visually estimating the centre of the second formant on a wideband spectrogram, with reference to a narrowband spectrogram, fast Fourier transform (FFT) spectra, and linear predictive coding (LPC) spectra as needed. Measurements were taken at the first glottal pulse after the stop burst that excited at least the first two formants of the vowel[3]. A 600-Hz analysing bandwidth was used for the wideband spectrogram so that it was slightly greater than twice the child's mean fundamental frequency of 248 Hz[3]. Audio files for spectral moments analysis were prepared as for VOT and F2 measures but using a low pass cut-off filter of 10 kHz. Each word waveform was edited to include the interval from the onset of the first burst to the end of the third pitch period with a 10-ms leader. The spectral moment values for all audio files were calculated using CSpeechSP, which automatically generates FFTs on a series of overlapping 20-ms hamming windows at 10-ms intervals. The initial window was centred on the stop burst and the final window was centred at the end of the third pitch period of the vowel. Burst spectral mean and burst spectral skewness were computed from the resulting probability distributions. Results are reported for the first 20 ms interval, centred on the stop burst.

3. RESULTS

Effects of the two independent variables, place and context, were tested statistically using two-way ANOVAs with repeated measures for each acoustic cue. The results are summarized in Table 1. Significance or size of differences in means and the direction of those differences are listed for each cue.

Table 1. Relationship of bilabial and alveolar means for four acoustic cues.

Acoustic Cue	Place	Context	P x C
VOT	B<A (p = .008)	S</ɹ/ (p = .002)	No (p = .161)
F2 Onset	B<A (p = .004)	S>/ɹ/ (p = .117)	Yes (p = .035)
Spectral Mean 1st 20-ms	B<A (p = .04)	S>/ɹ/ (p = .38)	No (p = .99)
Spectral Skewness 1st 20-ms	B>A (p = .04)	S</ɹ/ (p = .03)	No (p = .49)

Note. B = Bilabial; A = Alveolar; S = Singleton; /ɹ/ = /ɹ/-Context

Significant main effects were found for both place and phonetic context for VOT and F2 onset. Mean VOT and F2 onset values were lower for bilabial than for alveolar place in both contexts. For the first 20-ms segment of the burst, a significant main effect of place was found for moment 1 and moment 3. Moment 1 values were lower and moment 3 values were more positive for bilabial than lingua-alveolar place. Moment 1 did not differ significantly by context but mean values were lower in singleton context for both places of articulation. A significant main effect for context was found for moment 3 values, with lower values in singleton context.

4. DISCUSSION AND CONCLUSIONS

The child with normal facial movement and both children with facial paralysis showed the expected pattern of VOT and F2 onset values for bilabial vs. lingua-alveolar stop place in both CV and CɹV contexts. For both children with facial paralysis, these acoustic distinctions in consonant place were greater in/ɹ/-context. However, each of the three children showed a different result for the two stop burst cues. The child with normal facial movement and typical articulation showed the expected pattern of a lower mean and more positive skewness for bilabials in a CV context but the reverse of this pattern in the CɹV context. CP1 showed the unexpected pattern of a higher mean and a less positive skewness for bilabials in both CV and CɹV contexts. CP2 showed the pattern of a lower mean and a more positive skewness for bilabials in both CV and CɹV contexts. These varied results for spectral burst cues suggest that they are sensitive to articulatory differences in stop articulation while VOT and F2 onset appear to be more stable place cues, regardless of articulatory differences. Further research with larger numbers of speakers is necessary to clarify the effect of /ɹ/ on spectral burst place cues in typical speakers and to determine the relative importance of different cues on place identification for speakers with typical and atypical stop consonant articulation.

5. REFERENCES

1. Pickett, J. (1999). *The acoustics of speech communication: Fundamentals, speech perception theory, and technology.* Needham Heights, MA: Allyn & Bacon.

2. Cho, T. & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*, 207-229.

3. Nelson, M.A., & Hodge, M.M. (2000). Effects of facial paralysis and audiovisual information on stop place identification. *Journal of Speech, Language and Hearing Research, 43*, 158-171.

4. Hodge, M., Hegedus, J., Kenny, M., & Souster, C. (2001, June). *Influence of /ɹ/ on burst spectra for stop place.* Poster session presented at the 141st meeting of the Acoustical Society of America, Chicago, IL.

5. Milenkovic, P.H., & Read, C. (1997). CSpeechSP, Laboratory Version [Computer Program]. Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI.