CHARACTERIZING THE ADULT TARGET: ACOUSTIC STUDIES OF SWEDISH AND AMERICAN ENGLISH /t/ AND /p/.

E. H. Buder[a], K. Williams[b], and C. Stoel-Gammon[a]
[a]University of Washington, Seattle, WA, USA
[b]Stockholm University, Stockholm, Sweden

ABSTRACT

Work by our group has investigated phonetic development of language-specific segments, including study of differences in place of /t/ articulation (Swedish: dental, American English: alveolar). We have reported acoustic measures showing language-specific spectral shapes for these bursts in both adults and 30-month-old children [1]. In this paper we examine factors that support the interpretation that these differences are indeed due to place.

INTRODUCTION

A recent study [1] revealed perceptual and acoustic differences between Swedish (S) dental and American English (AE) alveolar /t/ bursts. Listeners were able to categorize 15 ms burst portions of word-initial /t/s in both adults and 30-month-old children as dental and alveolar. Acoustic measurements indicated that spectral diffuseness (measured as std. deviation in Hz of the burst spectrum (SD)) and burst intensity (measured in dB difference from the following vowel) were significantly shorter across languages. Bursts were more diffuse and less intense in S than in AE. Shorter VOTs were also typical of Swedish /t/-initial tokens in both adults and children.

This paper reports similar differences in S and AE adults' word initial /p/ bursts from the two languages, suggesting that some language-specific acoustic features are shared among the stop consonants in our data. The central question guiding the present investigation is whether the spectral shape measures we have applied to /t/ bursts are uniquely associated with the alveolar/dental place distinction. Before continuing to investigate this topic in development, we need to characterize the adult targets using measures that uniquely capture this distinction. We therefore present data addressing two factors that may have affected earlier results: 1) that some of the differences in spectral shape may be due to differences in recording (e.g. equipment and room acoustic differences), and 2) that some of the spectral shape differences in both /t/ and /p/ are related to lower intensities and shorter VOTs of Swedish word-initial bursts.

RECORDING EFFECTS ON SPECTRAL SHAPE

Spectral shape measures of stop bursts based on a "moments" analysis of Fourier spectra have been of central interest in our work, following earlier work on the technique [2], [3]. Briefly, the spectrum can be characterized by its mean (M) and std. deviation (SD) in Hz and also by the higher moments-based dimensionless coefficients of skewness and kurtosis. Stoel-Gammon et al. [1] reported differences in all these measures when comparing adults' and children's /t/ bursts in S and AE, but acknowledged that some differences may have been due to recording effects such as room acoustics, microphone types, and the different standards of videotape recording media used (PAL in S, and NTSC in AE). Informal calibration efforts led us to suspect artifactual influences on the absolute validity of spectral M and the higher moments, but the effects in SD seemed too large to be artifactual. Further acoustic investigations of /p/ bursts revealed that SD differences were in the same direction as for the /t/ bursts.

In order to better calibrate recording effects, synthetic burst tokens were developed using filtered and dynamically shaped white noise to create 15 ms transients centering at four frequencies (1, 2, 3, and 4 kHz) and with two different bandwidths to emulate the alveolar/dental contrast in diffuseness. These burst tokens were recorded by playback over identical versions of Kay CSL software and the same speaker (JBL Pro-III) in both recording environments, with the same equipment used to record the actual speech data under investigation. Analysis of these data then proceeded using the same methods as Stoel-Gammon et al. Results indicated small yet systematic differences between S and AE recording environments in spectral M and SD, and larger differences in skewness and kurtosis. Because M and SD are the most powerful and interpretable in our speech data we focused on calibrations of these measures (see [3] for further discussion of difficulties with interpretation and statistical analysis of the higher moments). See Table 1 for adjustment values obtained from the calibration data.

Table 1. Adjustments to spectral mean and SD due to recording differences.

	Spec. M	Spec. SD
S-AE difference	+ 143 Hz	+ 73 Hz

Adjusting our adult /t/ and /p/ burst measures accordingly, the calibrated data were used to measure and statistically analyze (by *t*-tests) the language differences reported in Table 2. As can be seen, VOT and Burst intensity are significantly different across languages for both /t/ and /p/. However, the cross-language differences between spectral shape measures change when the measures are calibrated: an apparently non-significant difference between /t/ burst spectral Ms becomes significant, and a nearly significant difference between /p/ burst spectral Ms becomes non-significant. Regarding spectral SDs, the significant language difference in /t/s remains significant, and a significant difference in /p/ SDs is reduced, but still significant, when the measures are calibrated. The next section of this report investigates this further by assessing relations among VOT, burst intensity, and calibrated spectral shape measures.

VOT, BURST INTENSITY, AND SPECTRAL SHAPE MEASURES

It is possible that some degree of spectral shaping is related not to place of articulation but instead to burst intensity and VOT. In terms of intensity, the turbulence noise of a /p/ burst may become lower in central frequency and more compact when the stop is released with greater pressure as it appears to be in AE. In terms of VOT, a release burst may become higher in frequency and more compact when the longer VOT of AE yields a stop release that is followed by more aspiration. These possibilities seem supported by regression analyses investigating the effects of VOT and intensity on our spectral shape data. Burst intensity is a significant predictor of SD in S /p/s ($p<0.001$) and in AE /p/s ($p<0.05$), lower SDs correlating with higher intensities. Higher intensity is also a significant predictor of lower spectral means in S /p/s ($p<0.01$) and in AE /p/s ($p<0.001$). In AE /t/s, longer VOT is correlated with higher spectral means ($p<0.05$) and with lower spectral SDs ($p<0.05$).

Table 2. Acoustic measures of Swedish (S) and American English (AE) adult /t/ and /p/ bursts, with t-test comparisons across languages.

	S average	AE average	t	p
/t/				
VOT, ms	49	74	7.163	<.001
Intensity, dB below vowel	17.7	12.8	-6.376	<.001
Uncalibrated, kHz				
Burst Mean	5.158	5.501	1.793	n.s.
Burst SD	2.127	1.194	-9.477	<0.001
Calibrated, kHz				
Burst Mean	5.015	5.501	2.555	<0.05
Burst SD	2.054	1.149	-8.747	<0.001
/p/				
VOT, ms	41	66	7.061	<0.001
Intensity, dB below vowel	20.5	18.1	2.287	<0.05
Uncalibrated, kHz				
Burst Mean	2.957	2.604	-1.833	<0.10
Burst SD	2.015	1.689	-3.794	<0.001
Calibrated, kHz				
Burst Mean	2.814	2.604	-1.090	n.s.
Burst SD	1.942	1.689	-2.945	<.01

To investigate the effects among all these variables in a multiple regression framework, a logistic model can be used [4]. This type of regression model must be used when the dependent variable is categorical; here the continuous acoustic variables of VOT, intensity, burst spectral M and burst spectral SD can be entered as continuous predictors of the categorical language variable (S/AE). The logistic regression model is also appropriate because the model assesses the percentage of observations that are successfully classified according to language. By comparing the successful classification of a model incorporating VOT and intensity alone with the improvement in classification of a model that also incorporates the calibrated spectral shape measures, we see the extent to which spectral shape measures uniquely improve language discrimination above and beyond VOT and intensity alone. Table 3 lists these models and their classification success in the /t/ and /p/ bursts of our adult dataset. In the /p/ bursts, the marginal increment of classification success in the model incorporating spectral shape measures indicates that language differences in spectral shape do not contribute much additional predictiveness. In the /t/ bursts however, the addition of spectral shape measures contributes predictiveness that clearly goes beyond VOT and intensity. This result is consistent with the hypothesis that our spectral shape measures of /t/ bursts correlate with the alveolar/dental place distinction, and helps to explain the spectral shape differences in /p/ bursts as epiphenomenal to VOT and intensity differences.

Table 3. Language classification success percentages for logistic regression models with and without calibrated spectral shape measures for S and AE /t/ and /p/ bursts.

/t/	Classification Success
VOT and intensity alone	75%
VOT, intensity, burst mean and SD	85%
/p/	
VOT and intensity alone	71%
VOT, intensity, burst mean and SD	73%

CONCLUSIONS

After 1) calibrating spectral shape measures for effects of recording environment and 2) demonstrating effects of VOT and burst intensity on spectral shape, we conclude that burst spectral M and SD can be used as measures for language specific aspects of /t/, presumably correlating with place of articulation. Based on these demonstrations with adult speech samples, our future research will continue to use spectral shape measures to examine the development of place of articulation in children's speech.

REFERENCES

[1] Stoel-Gammon, C., Williams, K., and Buder, E.H. (1994). Cross-language differences in phonological acquisition: Swedish and American /t/. *Phonetica*, vol. 51, pp. 146-158.

[2] Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R.N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *Journal of the Acoustical Society of America*, vol. 84, pp. 115-123.

[3] Buder, E.H., Kent, R.D., Kent, J.F., Milenkovic, P., and Workinger, M. (in press). FORMOFFA: an automated formant-moment-fundamental frequency-amplitude analysis of normal and disordered speech. *Clinical Linguistics and Phonetics.*

[4] Steinberg, Dan and Colla, P. (1991). *LOGIT: A supplementary module for SYSTAT.* Evanston, IL: SYSTAT, Inc.