ACOUSTIC MOMENTS DATA FOR PALATALIZED AND DENTALIZED SIBILANT PRODUCTIONS FROM SPEECH DELAYED CHILDREN WITH AND WITHOUT HISTORIES OF OTITIS MEDIA WITH EFFUSION

Phonology Project Technical Report No. 12

Heather B. Karlsson

Lawrence D. Shriberg

Alison Scheer

Connie J. Nadler

January 2003

Phonology Project, Waisman Center,
University of Wisconsin-Madison

Preparation of this report was supported by research grant DC00496 from the
National Institute on Deafness and Other Communication Disorders,
National Institutes of Health (Lawrence D. Shriberg, P.I.)

BACKGROUND AND PURPOSE

This goal of this technical paper is to provide acoustic data supporting prior research that required acoustic differentiation of sibilant errors perceived as palatalized from those perceived as dentalized. Findings from a prior study (Karlsson, Shriberg, Flipsen, & McSweeny, 2002) indicate that speech delay associated with early recurrent otitis media with effusion and concomitant fluctuant hearing loss may arise from a child's difficulty in reliably perceiving and then producing correct lingual place for sibilant fricatives. The study further indicated that the mean frequencies of sibilant fricatives produced by speakers with speech delay (SD) and OME+ (positive) histories versus SD-OME- (negative) histories were reliably different. Specifically, the average spectral frequencies of sibilants (/s/, /z/, /ʃ/) of children with SD-OME+ histories were lower than those produced by children with SD-OME- histories. In our prior research, this hypothesized articulatory correlate for the lower frequency sibilant value—sibilant production at a more posterior place in the vocal tract—is termed *backing of sibilants*. The present technical report documents the acoustic methods used to identify backing of sibilants that are specific for this error type. That is, these methods differentiate perceptually *backed* productions from those perceived by transcribers as dentalized or *fronted* (i.e., presumed to be associated with more anterior tongue placements).

METHOD

Participants

Table 1 is a summary of descriptive information for a sample of 20 children with speech delay of currently unknown origin. Speech samples and case records of all children were selected from three prior studies of the child speech-sound disorders reported in Shriberg and Kwiatkowski (1994) and in Shriberg, Gruber, and Kwiatkowski (1994). The screening criterion for possible inclusion in these studies was the presence of age-level discrepancies in speech production as sampled in conversational speech (Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997). Conventional exclusionary criteria included evidence supporting any sensory, cognitive, craniofacial, speech-motor, or affective impairment that would warrant an alternative diagnostic label. The 20 participants ranged in age from 38 to 67 months, and the overall sex proportion was 40% male and 60% female.

Additional requirements for inclusion in the present study concerned participants' histories of early recurrent otitis media with effusion and their production of sibilants during a conversational speech task. The primary criterion for inclusion in a subgroup termed SD-OME+ was a history indicating that the child had experienced at least five episodes of OME, with strong parental and/or audiological support data indicating the likelihood of hearing loss during some or all of these episodes. This subgroup of 10 children reportedly "didn't hear well" during episodes of OME, with hearing loss for many children documented by audiological examination reports. To maximize the clinical generalizability of the present findings, the emphasis for the present study was on this type of case history data, typically obtained from parents during intake interviews and follow-up records requests to other health care professionals. Thus, unreliability in such data would increase both false positives and false negatives, as well as decrease the likelihood of rejecting the null hypothesis of no between-group differences. Participants in this group also had 10 or more palatalized /p/ productions during a conversational speech task.

A second group of children was identified from the same cohort of studies noted previously. Criteria for inclusion in the second group were very few (1–2) or no episodes of OME and production of 5 or more tokens of *dentalized* /s/ during a conversational speech task. Furthermore, this group had a maximum of five productions of *palatalized* /s/ sounds during a conversational speech task.

Assessment

The recorded samples used for the present study were selected from two speech tasks. All recordings had been obtained using a Sony 5000 monaural audiocassette recorder and a matching remote microphone monitored at a lip-to-microphone distance of approximately 6 inches. Recording procedures included well-developed monitoring conventions to maximize signal-to-noise ratios.

The first speech task was a conversational speech sample in which children were invited to converse about topics such as their daily activities, friends, and past and upcoming special occasions. The second task was a modified administration of the Photo Articulation Test (PAT: Pendergast, Dickey, Selmar, & Soder, 1984). The research protocol used in the studies from which these participants were selected directed the examiner to obtain two responses to each PAT stimulus word. The examiner first attempted to evoke a response spontaneously, using the pictures in the test book. If spontaneous evocation was not successful, the examiner used delayed imitation (e.g., "That's a ___; say the word") to evoke the correct word. The examiner also obtained a second response to each word using direct imitation. Thus, the data set available for the present analysis included spontaneous responses for most words and imitative responses for all words.

Acoustic Analyses

Preliminary analyses. Acoustic analyses of backing in the two groups of speakers were completed on a set of 9 PAT words, from which spontaneous (i.e., citation) responses had been obtained. The analyses were preceded by two scans through the raw data to eliminate responses that would not be appropriate for further analysis. First, tokens were eliminated if the wrong word had been produced, or if examiner overtalk or extraneous environmental noise had obscured the target sound(s) in the word. Data loss from such factors was approximately 6% (i.e., 12 of the original 200 [20 x 10] tokens were eliminated).

The second exclusionary analysis was based on a review of the narrow phonetic transcription of each participant's responses. Responses were classified into seven categories based on the participant's production of the target segment in the word as symbolized in narrow phonetic transcription. The seven categories identified whether the target was (1) produced correctly, (2) replaced by another consonant of the same manner but at a more posterior place (i.e., backed), (3) backed, as indicated by a backing diacritic, (4) fronted, as indicated by a fronting diacritic, (5) deleted, (6) replaced by a consonant of a different manner, or (7) replaced by a vowel. Only those responses meeting criteria for any of the first four categories were retained for analyses. The result of this analysis was the elimination of an additional 44 responses from the 188 remaining eligible responses. Thus, the final data set appropriate for acoustic analyses included 144 (or 72%) of the original 200 responses to the nine PAT words.

Nine of the PAT words yielded data on three target sounds for the acoustic analyses. One word, *scissors*, yielded a sufficient number of tokens for analyses of two sounds. The target words included the alveolar fricatives (/s/, /z/) in *sandwich*, *saw*, *scissors*, *house*, *zipper*, and *keys*, and the palatal fricative (/ʃ/) in *shoe*, *brush*, and *fish*. Each of these three sibilants can be articulated with more anterior or more posterior lingual positioning, as indexed acoustically by the first spectral moment at the temporal middle of the fricative noise. The temporal middle of the fricative noise was chosen for acoustic analysis, because in that position there are assumedly reduced effects of phonetic context.

Procedures. Acoustic analyses of the sibilant segments from each of the target words noted above were completed. Eligible responses were digitized using a Tascam 112MK II tape deck connected to a Kay Elemetric Computer Speech Lab 4300B (CSL4300B) station. The signal was sampled at 20 kHz with 16 bits of quantization using the CSL 4300B record facility.

Files digitized in CSL were reopened in TF32 (Milenkovic, 2000). Fricatives were identified by the presence of strong aperiodic energy as evidenced in both the waveform and the spectrogram. The left cursor was placed at the onset of strong aperiodic energy and the right cursor at the offset of strong aperiodic energy. Moments files were subsequently created and analyzed in TF32 with a 20 ms Hamming window and 10 ms step. The middle time slice (i.e., the middle 20 ms) was identified and isolated. If a file contained an even number of analysis windows, one of the two middle windows was selected randomly.

RESULTS AND DISCUSSION

Acoustic Analyses and Discussion

Acoustic findings. Figure 1 is a display of the first spectral moment (M1) findings for the sibilants in the 10 words arranged by sound class (alveolars, palatal) and increasing M1 (kHz) values for children in the SD-OME+ group (see below). Based on the prior research, the children in the SD-OME+/*palatalized* group were expected to have lower M1 values on the three sibilant fricatives. Examination of Figure 1 is consistent with this expectation, illustrating a trend for children with SD-OME+ and *palatalized* /s/ histories to have lower M1 values than their counterparts with SD-OME- and *dentalized* /s/ histories.

As indicated in Table 2, statistically significant between-group differences were obtained for only one word—*shoe*. However, calculation of effect size yielded a very large effect size for *shoe*; a large effect size for *keys*; medium effect sizes for *house, saw, scissors*, and *zipper*; and a small effect size for *fish*. Effect sizes for the remaining target words (*sandwich, scissors, brush*) were negligible. One potential explanation for the larger effect sizes associated with *shoe* and *keys* might be that there is relatively little movement of the tongue between the sibilant and the preceding or following vowel for each of these two words. That is, in this context, the sibilant is less influenced by place features of adjacent vowels, which in these two words were also high back and high front, respectively.

Acoustic differences by perceptual diacritic. Detailed analyses of M1 frequencies by diacritic (i.e., *dentalized* or *palatalized*) were undertaken. Findings indicated no reliable differences or trends in specific frequencies for specific sibilant errors perceptually transcribed as *dentalized* or *palatalized*. This could be due to a number of factors including the following two considerations. First, a transcriber perceptually considers the entire fricative production as well as context cues to make a judgment about a specific modifying diacritic. In contrast, the moments analysis procedure considered only a 20 ms time slice from the midpoint of fricative production. Second, the relevant acoustic differences could be subperceptual, such that a wide range of tokens perceived as "normal," may be, in fact, acoustically distinguishable.

CONCLUSION

The findings of this study are viewed as support for the possibility of identifying acoustic markers to classify etiological subtypes of speech delay of currently unknown origin. Although individual productions of *palatalized* or *dentalized* /s/ could not reliably be distinguished acoustically, the children who had histories of the respective error patterns were acoustically distinguished by group on many of the target words. As in a prior study (Karlsson et al., 2002), lower mean spectral frequency was the distinguishing characteristic of sibilants for children with SD-OME+ history and *palatalized* /s/ productions. In contrast, children with SD-OME- histories and perceptually *dentalized* /s/ productions had higher spectral mean frequencies.

REFERENCES

Karlsson, H. B., Shriberg, L. D., Flipsen, P. Jr., & McSweeny, J. L. (2002). Acoustic phenotypes for speech-genetics studies: toward an acoustic marker for residual /s/distortions. Clinical Linguistics and Phonetics, 16, 403–424.

Milenkovic, P. (2000). Time-frequency analysis for 32-bit Windows [Computer software]. Madison, WI: Department of Electrical Engineering, University of Wisconsin-Madison.

Pendergast, K., Dickey, S., Selmar, J., & Soder, A. (1984). The Photo Articulation Test (Rev. ed.). Dansville, IL: Interstate Printers and Publishers.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The Percentage of Consonants Correct (PCC) metric: Extensions and reliability data. Journal of Speech, Language, and Hearing Research, 40, 708–722.

Shriberg, L. D., Gruber, F. A., & Kwiatkowski, J. (1994). Developmental phonological disorders III: Long-term speech-sound normalization. Journal of Speech and Hearing Research, 37, 1151–1177.

Shriberg, L. D., & Kwiatkowski, J. (1994). Developmental phonological disorders I: A clinical profile. Journal of Speech and Hearing Research, 37, 1100–1126.

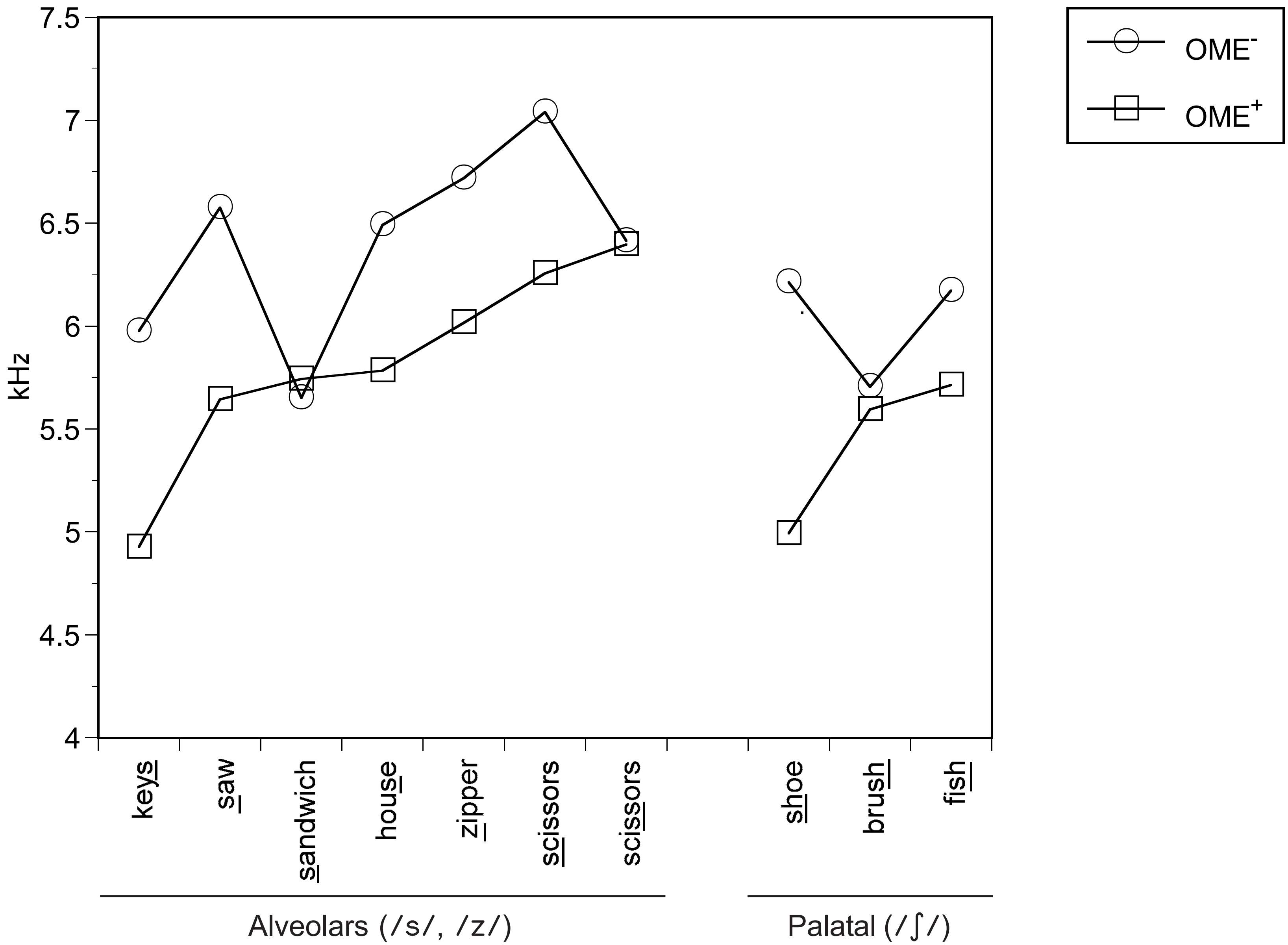
Table 1. Description of the 20 participants divided into two groups based on their histories of early recurrent otitis media with effusion (OME).

Group				**Age (in mos.)**			**Sex**		**PCC (%)**			**PCC-R (%)**			**II (%)**		
Number	***n***	**OME History**	**/s/ Error Status**	***M***	***SD***	**Range**	**% Male**	**% Female**	***M***	***SD***	**Range**	***M***	***SD***	**Range**	***M***	***SD***	**Range**
1	10	Positive	Palatalized	50.9	8.9	38–67	50.0	50.0	63.3	9.2	49.2–78.1	72.0	9.7	54.7–82.8	88.7	9.4	80.0–99.1
2	10	Negative	Dentalized	45.8	3.3	40–50	30.0	70.0	67.0	7.1	54.7–82.1	78.2	7.0	63.8–90.4	91.2	8.5	72.0–99.5
Total	20			48.4	6.1	38–67	40.0	60.0	65.2	8.2	49.2–82.1	75.1	8.4	54.7–90.4	90.0	9.0	72.0–99.5

Table 2. Mean frequencies (in kHz), Mann-Whitney comparison *p* values, and effect sizes for the two study groups.

	OME+		OME-				
Word	***M***	***SD***	***M***	***SD***	**Mann-Whitney**	**Effect size**	**Descriptor[a]**
shoe	4.993	0.856	6.214	0.869	.03	1.42	Very Large
keys	4.927	1.237	5.976	1.309	.06	0.82	Large
saw	5.644	1.380	6.576	1.341	.42	0.69	Medium
scissors	6.526	1.145	7.040	1.405	.20	0.61	Medium
house	5.784	1.320	6.492	1.150	.45	0.57	Medium
zipper	6.017	1.191	6.719	1.306	.40	0.56	Medium
fish	5.714	1.198	6.173	1.705	.95	0.32	Small
sandwich	5.743	0.950	5.652	0.990	.75	0.10	Negligible
brush	5.595	0.711	5.706	1.554	.96	0.10	Negligible
scissors	6.397	0.915	6.414	2.353	.74	0.01	Negligible

[a]Descriptors for effect sizes: *Negligible* (0.00–0.19); *Small* (0.20–0.49); *Medium* (0.50–0.79); *Large* (0.80–0.99); *Very Large* (≥ 1.00).



7.5
7
6.5
6
5.5
5
4.5
4
kHz
keys
saw
sandwich
house
zipper
scissors
scissors
shoe
brush
fish
Alveolars (/s/, /z/)
Palatal (/ʃ/)
OME-
OME+