Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

An acoustic investigation of the Swedish child's acquisition of obstruent place of articulation

Jan Van Doorn[a], Fredrik Karlsson[b] and Kirk Sullivan[b]

[a]Umeå University, Department of Clinical Sciences, SE90187 Umeå, Sweden
[b]Umeå University, Department of Language Studies, SE90187 Umeå, Sweden
jan.vandoorn@logopedi.umu.se

Speech produced by children in the initial stages of development does generally not uphold as many phonetic distinctions as speech sounds produced by adults. A child's productions of different target words may therefore have similar acoustic properties and result in homonyms being perceived by the adult observer. This study presents a longitudinal investigation into the development of place of articulation from non-distinctive to distinctive productions in word-initial obstruents produced by 22 Swedish children (aged 18 - 48 months). The data was collected through monthly recordings, approximately one year per child. The acoustic correlates analysed were spectral diffuseness, spectral skewness and spectral tilt for plosives and spectral skewness, spectral kurtosis, spectral variance and F_2 onset frequency for fricatives. The results show a developmental trend in spectral skewness that is indicative of a increasing number of acquired phonetic contrasts. Spectral tilt change, F_2 onset frequency, spectral mean and spectral variance provide evidence of within-category refinement which is argued to be caused primarily by advancements in motor control.

1 Introduction

A frequently observed feature of children's developing speech is that attempted productions of a target obstruent may often be perceived by the adult observers as having a differing phonetic quality regarding place of articulation compared to when uttered by an adult. These homonyms have traditionally been analysed as applications of, for instance, a backing, fronting or dentalisation process [3] based on auditory judgments, and studied for developmental trends though observation of frequency of occurrence at different stages of development.

However, as argued by Oller [14], the adult's perception of segments produced by children is often affected by a *shoehorning* effect, hiding information on possible phonetic and phonological development by the perceptual system of the adult observer. This auditory observation of the details of children's developing speech may be obscured by the phonetic contrasts upheld by the adult observer. Observations of developing child speech made may, therefore, be influenced both by the limitations in the language specific perceptual system of adult observer and the child's actual development.

Acoustic analysis of the child's produced output forms has been proposed by many researchers as a less language and observer biased methodological alternative to auditory observation [9–11, 13, 17]. By using acoustic analysis, complemented by information on which target word was attempted, researchers would be able to go beyond the perceptual boundaries of the observer and provide a detailed investigation into the nature of the child's productions and the developmental path taken by the child.

Appart from establishing of a developmental path in obstruent production in Swedish children, it is also of great importance to form an hypothesis concerning the reason behind an observed developmental trend. Children's productions may progress towards the properties of an adult model production either due to advancements in the child's awareness of distinctive phonetic or phonological features in the attempted target word, or due to increased precision in articulatory movements. Thus, the question may be asked whether the child is developing towards the adult target production, or towards the child's internal representation of the adult target production, which may differ in phonetic features and teased appart using contrastive analysis of target forms.

The aims this paper are thus twofold. First, it aims to provide an acoustic description of Swedish children's productions of obstruents, based on established acoustic correlates of place of articulation in adult productions of obstruents [9–11, 13, 17]. Second, it aims to provide a description of which developmental tends should be regarded as an increasing phonological and phonetic awareness of the details of the target word, and which developmental trends should be regarded as articulatory development.

The general method employed is a contrastive analysis of the acoustic properties of the obstruent manifested in relation to a) the produced phonetic obstruent quality, and b) the intended phonetic obstruent quality. Refinement within produced category not coupled with a refinery within elicited category is analysed as a increased motor control. In contrast, the reversed progression (an refinement with age when dividing the data set by elicited place of articulation that was not coupled with a progression towards the adult target when dividing by produced place) is analysed as a progression from one target towards another due to an increased awareness in the phonetic quality of the target obstruent.

1.1 Acoustic correlates of place of articulation in plosives

A number of acoustic correlates of place of articulation for plosives have been proposed. In an early investigation, Stevens and Blumstein [19] considered the relative importance of the burst spectrum at the point of plosive release and the following formant transitions. They concluded that the gross shape of the spectrum of a 26 ms window of the burst serves as a primary cue in the distinction of plosive place of articulation with the following formant transitions as supporting cues. Stevens and Blumstein [19] and Stevens [18] further proposed a classification scheme based on the shape of the release burst spectra. They described the spectrum of velar plosives as narrow with a prominent mid-frequency peak, while alveolar plosive spectra were described as more evenly distributed spectral energy with a rise in energy with increasing frequency ("diffuse rising"). Labial plosives were described as having an evenly distributed energy in the spectrum similar to that of alveolar plosives, but with either a flat or falling shape (diffuse flat or diffuse falling).

The static burst classification scheme of Stevens and Blumstein was extended by Kewley-Port [6] , using a set of acoustic criteria (tilt of the burst, presence of a late onset of low-frequency energy and the presence of a mid-frequency peak) that described changes in the spectrum over time. Using visual displays of running spectra the participants in [6] study correctly identified 78Ð92% of the cases. In a followup study, Kewley-Port et al. [7] concluded, in contrast to the previous statement made by Stevens and Blumstein [19] , that the primary acoustic correlates of place of articulation is in the dynamic and context dependent spectral properties after the release, rather than the static and context independent properties of the burst spectrum. The perceptual effect of the burst spectrum identified by Stevens and Blumstein [19] was argued to be a secondary cue that to a large degree depends on the presence of F_1 transitions.

The results of Kewley-Port et al. [7] were also in agreement with the results from an investigation of non-English speakers. In a cross-language investigation (Malayalam, French and English) Lahiri et al. [8] reported an inability to differentiate labial and dental plosives based on Stevens and Blumstein [19] classification of burst shapes. Instead, they proposed that the relative tilt of the spectrum at the burst at the onset of voicing (spectral tilt change) would afford an acoustic differentiation between labial and alveolar plosives. They obtained high levels of recognition based on the fraction between the spectral tilt at the release and the spectral tilt before the vowel, leading them to conclude that the proposed metric could serve as a productive quantification of the difference between plosives with spectral shapes described by Stevens and Blumstein [19] as diffuse using gross dynamic properties of the plosive.

1.2 Acoustic correlates of place of articulation in fricatives

Forrest et al. [2] suggested that classification of fricatives produced by adults may achieve a 92% or higher classification rate using the first (mean), third (skewness) and fourth (kurtosis) spectral moments. More recently, Jongman et al. [4] . were able to show significant main e?ects of spectral mean, spectral variance, skewness and kurtosis in the perception of fricatives. Measurements of spectral skewness were able to separate all places of articulation. For spectral mean and variance, with the exception of the bilabial – labiodental distinction, all place distinctions were separated by the acoustic cues. Kurtosis was able to distinguish between labiodental and dental fricatives. These investigations have shown that the static cues of spectral moments are a solid cue to the place distinction in adult speech.

In a study examining dynamic cues to place in fricatives, Jongman et al. [4] showed a significant e?ect of the onset frequency of F_2 but no significant effect of the calculated locus equation. Jongman et al.. therefore argued that it is the onset frequency of F_2, not the following F_2 trajectory, that provides the significant effect on perception of place for the fricative. In contrast, Soli [17] presented results indicating that the frequency of F_2 within the fricatives preceding a vowel varies both with place of articulation of the and with Place of Articulation of the vowel.

Although the data presented by Soli [17] show the largest effect due to the Place of Articulation of the fricative on the F_2 frequency manifested in the fricative, the e?ect of vowel context presented indicated that the transition between fricative and vowel may serve as a significant cue to the place distinction. From a developmental perspective, the cues of spectral moments and the onset frequency of formants have been shown to be weighted differently by children and adults. Nittrouer et al. [13] argued that a shift from a focus on the formant cues to the spectral cues from children to adults in the productions of fricatives. Nittrouer et al. [13] proposed that this shift indicates a transition from a focus on articulatory gestures towards a more developed awareness of the individual segments. Similar results were obtained by Nittrouer and Studdert-Kennedy [12] and Nittrouer and Miller [10, 11] from production experiments contrasting child and adult productions.

sV(C)	C^{h}V(C)	C̥V(C)	CCV(C)
sår [soːr]	*kal* [kʰɑːl]	*gal* [gɑːl]	*skal* [skɑːl]
sal [sɑːl]	*pår* [pʰoːr]	*bår* [boːr]	*spår* [spoːr]
så [soː]	*tå* [tʰoː]	*då* [doː]	*stå* [stoː]
sak [sɑːk]	*pak* [pʰɑːk]	*bak* [bɑːk]	*spak* [spɑːk]

Table 1: The corpus of elicited target word quadruples.

2 Method

The data set investigated in this paper was extracted from the Umeå Child Consonant Clusters Corpus (UC^3 Corpus) [5].

2.1 Participants

Twenty-one monolingual Swedish children, 12 male and 9 female, were recruited for the study. Information regarding the children's hearing status and dialect of Swedish spoken in the home environment was gathered from the parents/care-takers before the child began the recording phase of the investigation. All the children were growing up in a monolingual home environment and there were no known hearing problems. Each participant was recorded at monthly intervals for a period of twelve months. The recording phase of the investigation resulted in 229 acoustically satisfactory recordings. The children were between 77 and 218 weeks old at the time of the recording sessions.

2.2 Speech material

The target words elicited were four triplets of monosyllabic words. Each triplet contained target words with an identical syllable nucleus and coda. The onsets in each triplet were a voiced plosive, a voiceless aspirated plosive and an s+plosive consonant cluster in which the plosive is voiceless unaspirated in adult speech. The target words are presented in Table 1.

2.3 Procedure

The recordings were made in a sound treated room. The target words were elicited by the parent or care-taker accompanying the child to the recording session. The target words were elicited using black-and-white picture prompts. Each parent or caretaker was informed that they should aim for spontaneous productions. When it proved impossible to elicit a spontaneous production of a target word, they were instructed to elicit the target word through a delayed repetition task by providing the target word in non-final position in a carrier sentence. In the cases where a spontaneous production was not obtained, the parents and caretakers were asked to elicit the target word through direct repetition of an adult model production.

2.4 Acoustic markup and analysis

The children's productions of the target words were initially identified and extracted from the recorded sessions. A child's production was extracted from the session recording and included into the set of analyzed productions iff the context in which the production was made and the parent's reaction towards the production indicated that the child had attempted to produce the elicited target word. Productions that co-occurred with an external sound, such as banging on the table or an utterance made by someone other than the child, were not included in the analyzed set.

The extracted productions were then labeled at the phoneme level in WaveSurfer [15]. For each produced plosive, the release burst's start and end times as well as F_2 onset time and time of reaching a steady state of the following vowel were also marked. The F_2 automatically extracted formant track was verified and corrected manually. From the acquired label files, acoustic measurements of spectral diffuseness, spectral skewness and spectral tilt (for plosives) and spectral skewness, spectral kurtosis, spectral variance and F_2 onset frequency (for fricatives) were produced using custom scripts based on the AGTK [1] toolkit.

The acoustic quantfication of spectral tilt change from the release burst to the onset of the vowel proposed by Lahiri et al. [8] was caclucated from the amplitude found at 1500 Hz and 3500 Hz in the spectrum of the release (a and b) and at 1500 Hz and 3500 Hz at the onset of voicing (c and d). Based on these acoustic measurements, the q fraction was calculated as $q = (d - b)/(c - a)$.

The extracted data set was analysed by a contrastive analysis of the developmental trends observed when dividing the data set by transcribed phonetic quality (perceived Place of Articulation) with the developmental trend observed when dividing the set by elicited phonetic obstruent quality (target Place of Articulation).

3 Results

Example plots of the progression in the extracted acoustic correlates of place of articulation are presented in Figures 1-2,5-7 in the form of Box-and- whiskers plots, one box for each investigated age (in months). A linear regression line (solid line) and a dashed LOESS line (span=1) indicate the overall and local developmental trends in the data.

Classifications of plosives according to the q quotient are shown in Figures 3-4.

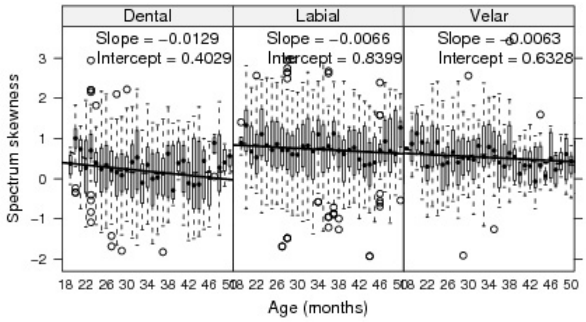

Figure 1: Skewness of the release spectrum against age of the participant. The data set is divided by elicited Place of Articulation.

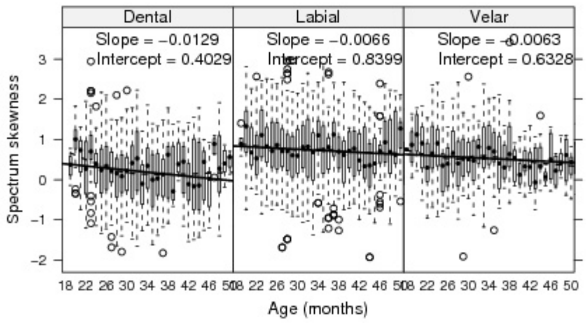

Figure 2: Skewness of the release spectrum against age of the participant. The data set is divided by perceived Place of Articulation transcribed.

4 Discussion

This paper aimed to provide a description of developmental trends in the acoustic manifestations of obstruents produced by children aged 18 – 48 months. Further, it aimed to provide insight into nature of the phonological awareness present in the child. In order to facilitate this, a general methodology of categorization of the results by both target Place of Articulation and perceived Place of Articulation was employed. Thus, in the case a progression may be observed in the categorization by target place towards an adult target value, children are viewed to have perceived the target correctly but are unable to produce it. In this case, the variance observed is caused by children's failure to meet the demands of adult perception. Thus, the imperfect productions were viewed as caused by articulatory factors and the variance in perceptual quality was viewed as an artifact of the adult perceptual system.

In contrast, developmental trends observed in the acoustic measurements divided by perceived Place of Articulation are argued to provide evidence of the child having attempted productions at more than one place of articulation. In turn, this suggests that the child may have an imperfect phonological representation of the target, possibly caused by perceptual processes. Thus, the consonants produced may be closer to an adult output form of the produced consonant than the adult output form of the consonant present in the target word elicited. Variation in perceptual quality was in this case viewed as representative of the child's intended production.

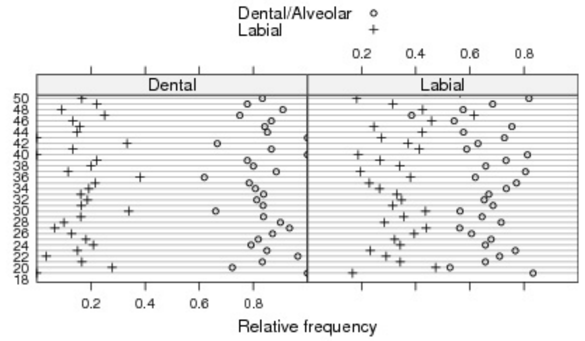

Figure 3: Categorization of plosives by the spectral tilt changefrom the release to the onset of voicing by the `q` quantity, divided by the Place of Articulation of the target plosive.

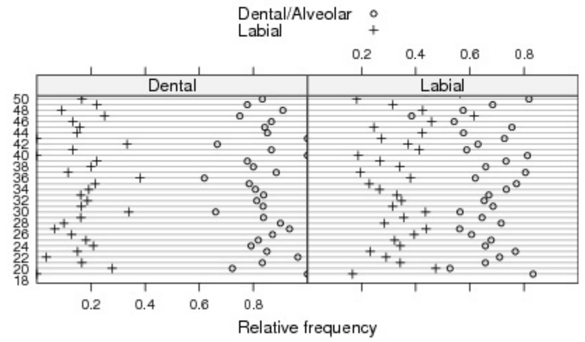

Figure 4: Categorization of plosives by the spectral tilt change in from the release to the onset of voicing by the `q` quantity, divided by perceived Place of Articulation of the plosive.

Of the four acoustic cues of Place of Articulation for plosives proposed in the literature, only spectral skewness and the spectral tilt change cue were shown to display a developmental effect towards the adult target. Skewness was shown to develop towards an increased separation in distribution skewness that agreed with what was predicted from data gathered from adult speakers of English. At the end of the investigated age interval, dental plosives most strongly agreed with a *diffuse-rising* spectral description, labial plosives agreed best with the *diffuse-falling* description and velar plosive skewness showed the most centred spectral distribution.

As this developmental trend was observed both in the data set when divided by target plosive place (Figure 1) and by perceived place (Figure 2), it is argued that this effect is caused by the overlap between two categorizations of the data points. However, a detailed analysis of the transcriptions a high rate of adult-like production of Place of Articulation was achieved by the children at an early age. This observation is interpreted as an indication that enough separation in terms of release spectrum skewness is achieved earlier than the earliest ages investigated here. Thus, place differences may in many cases is perceived using only the minimal separation displayed by the small differences displayed in Figures 1–2 at the age of 18 months. The developmental trends observed in Figures 1–2 are thus viewed as within-category refinement in order to take care of ambiguous cases.

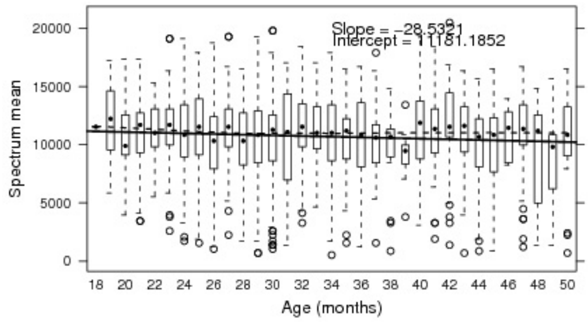

Figure 5: Spectral mean of the fricative spectrum against age of the participant. The data set is divided by Place of Articulation of the target plosive.

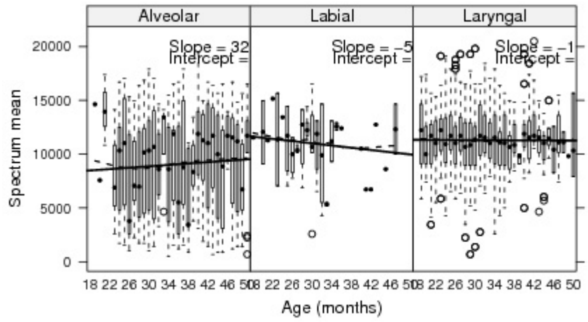

Figure 6: Spectral mean of the fricative spectrum against age of the participant. The data set is divided by Place of Articulation of the transcribed plosive.

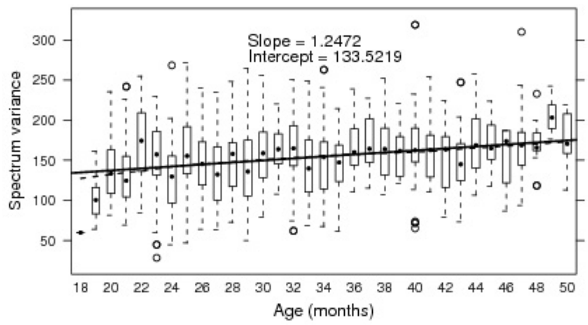

Figure 7: Variance of the fricative spectrum against age of the participant.

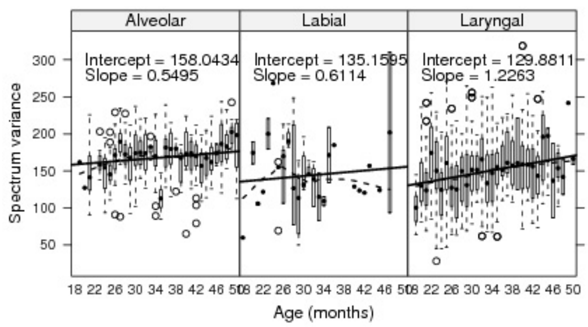

Figure 8: spectral variance of the fricative spectrum against age of the participant. The data set is divided by Place of Articulation of the transcribed plosive.

The cue of spectral tilt change from the release burst to the vowel proposed by Lahiri et al. [8] also provided evidence of a developmental trend. This trend was observed only in the division of the data set by Place of Articulation perceived by the adult observer. This developmental trend was not observed in the division of data by target Place of Articulation. Thus, it is proposed that the cue of spectral tilt indicates within-category adjustment of the cue rather than an attempt to achieve an increased adult target likeness.

For fricatives, three out of the four investigated acoustic cues of Place of Articulation were shown to change in usage with development. The spectral mean cue showed a developmental trend towards increasing adult values. This progression was, however, only observed in the division of the data set by perceived Place of Articulation, indicating a within-category refinement of the produced Place of Articulation. The lack of development towards the adult model found in the division by place of the target indicates that children do not use the cue of spectral skewness in order to push the output form closer to that of the target; thus, children may not be aware of the difference between how the target form and the output form produced are perceived by the adult observer.

5 Conclusion

This paper summarizes the findings of an investigation into the manifestations of acoustic correlates of place of articulation used by children from at ages 18 – 48 months. The data obtained suggest a developmental trend indicative of phonological maturation in spectral skewness of the release spectrum of plosives. Spectral tilt change, F_2 onset frequency, spectral mean and spectral variance provide evidence of within-category refinement which is argued to be caused primarily by advancements in motor control.

References

[1] S. Bird and M. Liberman. Annotation graphs as a framework for multidimensional linguistic data analysis. In *Towards Standards and Tools for Discourse Tagging, Proceedings of the Workshop*, pages 1–10. Association for Computational Linguistics, (1999).

[2] K. Forrest, G. Weismer, P. Milenkovic, and R. N. Dougall. Statistical analysis of word-initial voiceless obstruents: Perliminary data. *Journal of the Acoustical Society of America*, 84(1):15–123, (1988).

[3] P. Grunwell. *Clinical Phonology.* Croom Helm Ltd., 2 edition, (1987).

[4] A. Jongman, R. Wayland, and S. Wong. Acoustic characteristics of english fricatives. *Journal of the Acoustical Society of America*, 108(3):1252–1263, (2000).

[5] F. Karlsson. *The acquisition of contrast: a longitudinal investigation of initial s+plosive cluster development in Swedish children.* PhD thesis, UmeÅ University, (2006).

[6] D. Kewley-Port. Time-varying features as correlates of place of articulation in stop consonats. *Journal of the Acoustical Society of America*, 73 (1):322–335, (1983).

[7] D. Kewley-Port, D. B. Pisoni, and M. Studdert-Kennedy. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, 73(5):1779–1793, (1983).

[8] A. Lahiri, L. Gewirth, and S. E. Blumstein. A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study. *Journal of the Acoustical Society of America*, 76(2):391–404, (1984).

[9] S. Nittrouer. Children learn separate aspects of speech production at diffeent rates: Evidene from spectral moments. *Journal of the Acoustical Society of America*, 97(1):520–530, 1995.

[10] S. Nittrouer and M. E. Miller. Developmental weighting shifts of noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America*, 102(1):572–580, (1997).

[11] S. Nittrouer and M. E. Miller. Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America*, 102 (4):2253–2266, (1997).

[12] S. Nittrouer and M. Studdert-Kennedy. The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research*, 30:319–329, (1987).

[13] S. Nittrouer, M. Studdert-Kennedy, and R. S. McGowan. The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, 32:120–132, 1989.

[14] K. D. Oller. *The Emergence of the Speech Capacity.* Lawrence Erlbaum Associates, (2000).

[15] K. Sjölander and J. Beskow. Wavesurfer - an open source speech tool. In *Proceedings of ICSLP 2000*, volume 4, pages 464–467, Beijing, China, (2000).

[16] K. Sjȍlander and J. Beskov. The snack toolkit. 2000. URL `http://www.speech.kth.se/snack/`.

[17] S. D. Soli. Second formant in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, 70 (4):976–984, (1981).

[18] K. N. Stevens. Acoustic correlates of some phonetic categories. *Journal of the Acoustical Society of America*, 86(3):836–842, (1980).

[19] K. N. Stevens and S. E. Blumstein. Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 64(5):1358–1368, (1978).