The Role of Clinical Experience in the Perception of Children’s Speech in the Presence of Background Noise

Erin K. Diamond and Benjamin Munson

Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis

Introduction

The acquisition of /s/ and /ʃ/ involves their gradual articulatory and acoustic differentiation (Figure 1, Li, 2012). The standard tool for evaluating speech sound disorders is phonetic transcription. The categorical nature of phonetic transcription does not allow clinicians to measure fine phonetic distinctions within categorical responses. If Li’s findings extend to the learning of speech sounds in treatment, then the accurate assessment of fine phonetic detail is crucial to evaluating children’s progress toward proper speech sound production. Accurate and articulatory measures can accurately track current progress, but they are currently ineffective as clinical tools due to the time it takes to complete them.

Visual Analog Scaling (VAS, Figure 2) is a promising technique for performing gradient measures of children’s speech production. VAS ratings correlate strongly with within-category acoustic variation (Julien & Munson, 2012; Urberg-Carlson & Munson, 2013) and have been shown to be more strongly correlated when made by speech-language pathologists (SLPs) than by laypeople (Munson, Johnson, & Edwards, 2012, Meyer & Munson, 2013, this conference).

Previous research has examined perception in quiet. For these measures to be useful clinically, they must be robust in the presence of background noise.

Research Questions

This study evaluates the effect of background noise on SLPs’ and laypeople’s VAS ratings of children’s /s/ and /ʃ/ productions in quiet and in background noise. Ratings are evaluated in terms of the effect of clinical training and background noise on the reliability of VAS ratings, the correlation of VAS ratings with sounds’ acoustic characteristics, the extent to which VAS ratings are biased toward /s/ or /ʃ/, and the extent to which individuals provide a continuous VAS rating. We also examine whether measures of hearing acuity (pure tone thresholds and sentence intelligibility in noise) predict the effect of noise on VAS ratings.

Methods

Participants

Inexperienced Listeners: 20 adult listeners (11 women, 9 men), 18-50 years of age
Experienced Listeners: 10 adult listeners (9 women, 1 man, recruitment ongoing)

Visual Analog Scaling Task

Stimuli

A range of child productions of the fricatives /s/ and /ʃ/ were taken from a corpus of 2-5 year old children’s real-word-repetitions elicited by picture prompts in the /ṣn/ /ʃn/ project (http://www.learningtotalk.org/?q=node/24).

Initial CVs were excised from real-word productions to minimize lexical bias.

The peak frequency in the fricative (in ERB units) and the F2 frequency of the following vowel were measured. Li (2012) showed that these change throughout development.

Procedures

The speech sound stimuli were presented in quiet and in background noise at three SNRs, -3 dB, 0 dB, and +6 dB SNR (i.e., Figure 3).

Participants were asked to rate the consonant along a visual analog scale labeled “the ‘s’ sound” on one end and “the ‘ʃ’ sound” on the other (Figure 2).

Sentence Intelligibility in Noise

Listeners were presented with 20 sentences from the Basic English Lexicon (Calandruccio & Smuljanic, 2012) at a -3 dB SNR using the same babble as was used for the stimuli. Their responses were binned by the first author who wrote them. Percent keywords correctly repeated was noted. The groups did not differ in this measure (Mexp = 78%, Minc = 75%).

Pure Tone Thresholds

Standard audiometric procedures were used to calculate pure tone averages (PTAs) bilaterally at 0.5, 1, 2, and 4 kHz. The groups did not differ statistically significantly in left PTA (Mexp = 37.3 dB HL; Minc = 3.0 dB HL) or right PTA (Mexp = 33.3 dB HL; Minc = 3.0 dB HL).

Analysis and Results

Reliability was assessed by examining the correlation between first and second ratings as a function of group and SNR. The experienced listeners were statistically significantly more reliable than the inexperienced listeners (Mexp = 0.78, Minc = 0.33).

- Experienced listeners make more reliable ratings than inexperienced listeners. This does not interact with SNR.

- A series of quasipoisson regression analyses predicts VAS ratings from fricatives peak frequency and onset F2 were conducted separately for each listener at each SNR. The R2 from those regressions is shown in Figure 5. The R2 increased as a function of SNR and was higher for experienced listeners than for inexperienced ones. The β coefficients indicated that onset F2 influenced judgments more strongly in the quiet and +6 dB SNR condition than in the -3 dB SNR condition.

Ratings were more accurate in quiet than in noise. Experienced listeners’ ratings are more strongly predicted by the acoustic characteristics of the stimuli than are inexperienced listeners’. This does not interact with SNR.

Gaussian mixture models were used to decompose individual participants’ ratings into three underlying response distributions (Figure 6). The mean and standard deviation of those distributions was tallied. The means of the experienced listeners’ leftmost and rightmost (i.e., most “s”- and most “ʃ”-like) distributions were more extreme than those for the inexperienced listeners. Moreover, their ratings for the middle distribution was more dispersed than were those for the inexperienced listeners.

- Experienced listeners use a wider range of the visual analog scale than do inexperienced listeners, and they label a wider range of stimuli as intermediate between /s/ and /ʃ/. This does not interact with SNR.

None of the dependent measures above correlated significantly with any of the audiometric measures (left PTA, right PTA, or keywords correct).

Conclusion: Visual Analog Scale ratings of children’s /s/ and /ʃ/ are very robust to the presence of background noise. Clinical training is associated with ratings that are superior to those of inexperienced listeners in many key ways.

Acknowledgments

This research was supported by NIH Grant R01 DC02932 to Jan Edwards, Mary E. Beckman, and Benjamin Munson, and an award to Erin K. Diamond by the Roya Brynberg fund in the Department of Speech-Language-Hearing Sciences at the University of Minnesota.

http://www.learningtotalk.org