Continuous measures of children's speech production: Visual analog scale and equal appearing interval scale measures of fricative goodness.

Kari Urberg Carlson[1], Benjamin Munson[1], and Eden Kaiser[2]

[1]Department of Speech-Language-Hearing Sciences, [2]Program in Linguistics, University of Minnesota

Children acquire speech sounds gradually, yet most assessment tools elicit categorical judgments.

Children acquire speech sounds gradually, but the primary tool used to assess speech development, phonetic transcription, is by definition categorical. This presentation is part of a larger project, one of the goals of which is to develop novel perceptual methods for assessing children's speech production that capture continuity in speech-sound development.

Previous studies in our lab examined different methods of eliciting continuous judgments of children's speech production

Three methods were tested in previous studies: response times in forced-choice identification (FCRT), forced-choice identification followed by direct magnitude estimates of category goodness (FC-DME), and unidimensional visual-analog scaling (U-VAS, i.e., clicking on a double arrow bounded by the text "the 's' sound" and "the 'sh' sound"). In each experiment, listeners rated 400 Fricative-vowel syllables produced by children and adults. The most important criterion for choosing the best method was the extent to which listeners' responses correlated with the acoustic parameters that distinguish the phonemes in question: in this case /s/ and /ʃ/. FCRT and FC-DME correlated modestly with the first spectral moment (Figures 1 and 2), but VAS click location correlated highly with M1.

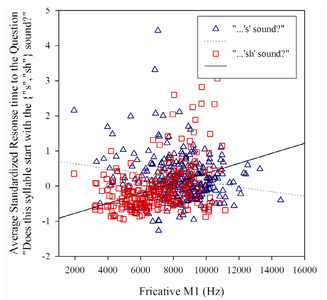

Figure 1: FCRT. Forced choice reaction times to the questions 'is this the ["s"/"sh"] sound' correlated poorly with acoustic parameters.

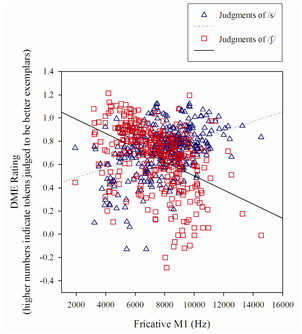

Figure 2: FC-DME. Participants were first asked to identify the phoneme as "s" or "sh". They then used direct magnitude estimation without modulus to rate the goodness of the phoneme. Correlations were better than those for FCRT but worse than those for U-VAS.

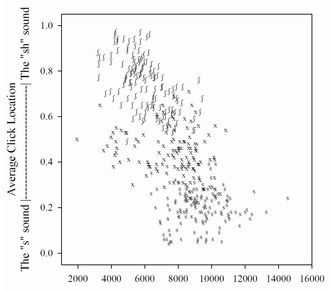

Figure 3: U-VAS. A unidimensional visual analog scale with 'the "s" sound' at one end and 'the "sh" sound' at the other provided the best correlation with the acoustic parameters. It was also the easiest method to administer and score. Sounds labels are taken from the FCRT task. "x" indicates sounds judged to be neither /s/ nor /ʃ/ in the FCRT task.

A unidimensional VAS scale cannot capture judgments of three-way comparisons.

The first set of experiments in the larger project involves examining the perception of two-way contrasts, such as /s/ and /ʃ/ in English, /s/ and /ɕ/ in Japanese. The larger project is also interested in perception of three-way contrasts, such as that in Mandarin among /s/, /ʂ/ and /ɕ/. U-VAS as implemented in the previous experiment is not able to assess the perception of three-way contrasts. Unfortunately, judgments from neither of the forced-choice tasks in the previous experiment were as strongly correlated with the acoustic characteristics of the stimuli as well as those from the U-VAS were. The current experiments examined whether a forced-choice-plus-goodness-rating method using VAS for the goodness ratings (FC-VAS) would yield correlations with acoustics as strong as those for the U-VAS scale. In this experiment, forced-choice /s/-/ʃ/ judgments were followed by a VAS goodness judgment. A second experiment using forced choice followed by a 7-point equal-appearing interval scale of category goodness (FC-EAI) was also tested.

Correlations with the acoustic parameters were much lower for FC-VAS and FC-EAI than for U-VAS.

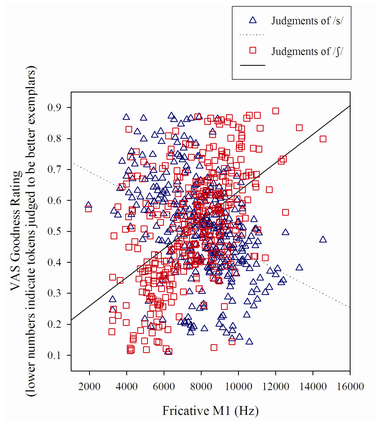

Figure 4: FC-VAS had much lower correlations with acoustic parameters than U-VAS. They were similar to the R^2 values for FC-DME (see fig. 6).

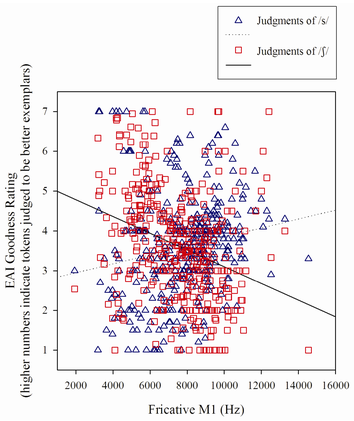

Figure 5: FC-EAI had lower correlations with acoustic parameters than FC-DME or FC-VAS.

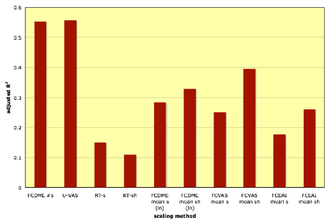

Figure 6: Values of adjusted R^2 for each of the experiments.

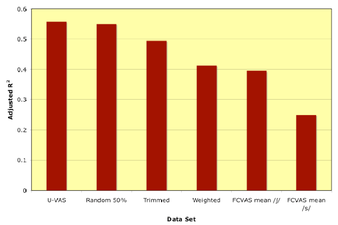

Figure 7: Reduction in R^2 due to changes to the data set.

Four factors may explain the decrease in adjusted R^2 relative to the U-VAS experiment.

1. The data set was smaller.

Once the stimuli in the forced-choice experiments were identified as /s/ or /ʃ/, they were separated into two groups, each with roughly half the number of data points as the U-VAS experiment. To simulate this reduction in data, half of the data points were randomly removed from the U-VAS data set. The value of adjusted R^2 dropped from .56 to .55, suggesting that the smaller data set was not a large factor in the reduction of variance accounted for in the forced-choice experiments. This case is labeled "Random 50%" in figure 7.

2. The range of the independent variable was reduced.

The stimuli at the edges of the data set with respect to the acoustic parameters were more likely to be eliminated from the forced-choice experiments because all of the participants labeled them as the same phoneme. There was some reduction in the range of the independent variable,. To simulate this condition, all of the stimuli that were never identified as /s/ in the FC-VAS experiment were eliminated from the U-VAS data set. The value of adjusted R2 dropped from .56 to .5, suggesting that the reduction in range may have contributed somewhat to the reduction in variance accounted for. This case is labeled "Trimmed" in figure 7.

3. The number of judgments per stimulus was variable.

In the U-VAS measurements, and the ratio of participants identifying a phoneme as /s/ or /ʃ/ (s:ʃ ratio), the average listener rating for each stimulus is based on judgments from all 20 participants. In the forced-choice and FCRT experiments where the data is split by phoneme identification, the number of judgments varies greatly between stimuli. To simulate this condition, listener ratings were randomly eliminated from the U-VAS data set based on the number of participants who identified each stimulus as /s/ in the FC-VAS experiment. Adjusted R^2 dropped from .56 to .41, although some of that change is due to the reduction in range that also occurred. The value of adjusted R^2 in this condition is very similar to the value of .40 for the /ʃ/ condition of the FC-VAS experiment. This suggests that together, reduction in range and low numbers of ratings for some stimuli may be responsible for the decrease in adjusted R^2 for the /ʃ/ condition, but cannot by themselves account for the decrease for the /s/ condition. This case is labeled "Weighted" in figure 7.

4. There were more degrees of freedom.

In the U-VAS experiment subjects were forced to compare /s/ with /ʃ/. In the goodness judgment ratings, there were no restrictions on the criteria participants could use to judge the quality of the stimuli. That the correlations were lower than for FC-VAS than U-VAS suggests that there may have been other criteria used to judge goodness. This is supported in that the ratings for /ʃ/ correlated better than those for /s/. /ʃ/ is generally confused only with /s/, while /s/ can also be confused with /θ/ and /f/ as well as with /ʃ/. If this is true, we would expect to see cross-linguistic differences in these tendencies. For example, in Japanese /s/ is confused only with /ɕ/, while /ɕ/ can be confused with /s/ and /ç/. We would expect to see stronger correlations stimulus acoustics and category-goodness judgments for /s/ than for /ɕ/.

References

Edwards, J., Gibbon, F, and Fourakis, M. (1997). On discrete changes in the acquisition of the alveolar/velar stop consonant contrast. *Language and Speech, 40,* 203-210

Lane, H. L., Catania, A.C. and Stevens, S. S. (1961). Voice level: autophonic scale, perceived loudness, and effects of sidetone. *Journal of the Acoustical Society of America, 31,* 160-167.

Li, F. (2008). Dissertation in Progress. Department of Linguistics, Ohio State University.

Li, F., Beckman, M., and Edwards, J. (2008, in press). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. In press in *Journal of Phonetics*.

Macken, M., and Barton, D. (1980). A longitudinal study of the acquisition of the voicing contrast in American English word-initial stops, as measured by voice onset time. *Journal of Child Language, 7,* 41-74.

Smith, N. (1973). *The Acquisition of Phonology: A Case Study*. Cambridge: Cambridge University Press.

Toner, M. A. and F. W. Emanuel (1989). Direct magnitude estimation and equal appearing interval scaling of vowel roughness. *Journal of Speech and Hearing Research, 32(1),* 78-82.

Tyler, A., Figourski, G., and Langsdale, T. (1993). Relationships between acoustically determined knowledge of stop place and voicing contrasts and phonological treatment progress. *Journal of Speech, Language, and Hearing Research, 36,* 746-759.

Urberg-Carlson, K., Kaiser, E., & Munson, B. (2008). *Assessment of Children's Speech Production 2: Testing Gradient Measures of Children's Productions*. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association.

Acknowledgements

This research was supported by NSF Grant BCS0729277 to Benjamin Munson, and by NIH grant R01 DC02932 to Jan Edwards. We thank Fangfang Li for conducting the acoustic analysis of the tokens. We thank Alysse Zittman and Celina Marnie for help running subjects and Marie Meyer and Sarah Mans for preparation of the data. We thank Ilana Heintz, Eric Fosler-Lussier, Mary Beckman and Jan Edwards for their extremely useful input on this work.