Assessment of Phonetic Skills in Children 2:

Testing Gradient Measures of Children's Productions

Kari Urberg Carlson[1], Eden Kaiser[2] , and Benjamin Munson[1]
[1]Department of Speech-Language-Hearing Sciences, [2]Program in Linguistics, University of Minnesota

Children acquire speech sounds gradually.

•Cross-sectional and longitudinal studies of speech-sound development often show an apparently discontinuous acquisition of speech sounds (e.g., Smith, 1973).
•Acoustic and articulatory studies suggest that children's speech sound acquisition involves the *gradual* acquisition of speech-sound contrasts. Some representative studies of this phenomenon are:

- Macken and Barton (1980): Children gradually learn to differentiate English voiced and voiceless initial stops along the voice-onset time dimension.
- Edwards, Gibbon, and Fourakis (1997): children with velar fronting patterns may differentiate between target /k/ and [t] for /k/ substitutions acoustically.
- Li, Beckman, and Edwards (2008): Children gradually differentiate between /s/ and /ʃ/ along two acoustic parameters: spectral center of gravity and frequency of the second formant at the onset of the following vowel.

Most assessment tools elicit categorical judgments.

•In most clinical assessment regimens and experimental research protocols, adults listen to a child and transcribe their production using a phonetic symbol and an optional series of diacritic markings. These transcriptions do not allow us to directly examine children's gradual attainment of speech sounds. The widespread use of phonetic transcription is arguably the by-product of two factors:
1. The apparent ease and simplicity of the task of phonetic transcription
2. The belief, based on studies of categorical perception, that people cannot perceive fine phonetic detail within a phoneme category.

Can listeners distinguish fine phonetic detail?

•A finer-grained method of rating children's productions would offer clinicians a way to document progress in children who have not yet mastered a target phoneme, and researchers the ability to examine children's exploration of the articulatory space. An ideal measure of children's production would:
1. capture gradient change in production,
2. correlate well with acoustic data,
3. be fast, and easy both to administer and to score.

In this study, three methods of collecting continuous ratings were assessed relative to these criteria. They were:

1. Reaction time to a forced-choice task (F-CRT)
2. Direct Magnitude Estimation without Modulus (DME)
3. Visual Analog Scaling (VAS).

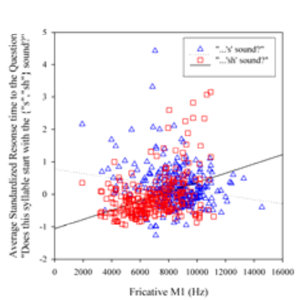
•In each of 3 experiments, English-speaking listeners were presented with 400 productions of CV syllables. The stimuli were elicited using a real-word repetition task, where to-be-repeated prompts were presented concurrent with pictures of the objects that they were naming. The speech tokens were collected as part of the παιδολογος project (http://www.ling.ohio-state.edu/~edwards, Jan Edwards, PI). The words were produced by native, monolingual English- and Japanese-acquiring children, and adults.
• Listeners' ratings were correlated with six acoustic parameters associated with fricative identification: The first four spectral moments (M1, M2, M3, M4) of a 40 ms slice of the middle of the fricative, onset F2 frequency of the following vowel, and relative intensity of the fricative and the vowel.

Reaction times from the forced-choice identification task did not correlate well with acoustic measurements.

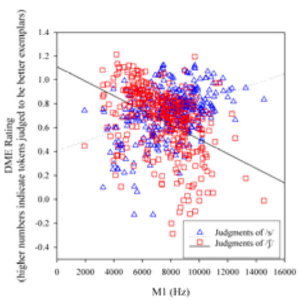
•For the forced-choice reaction time task, listeners were first asked to identify whether the target phoneme was /s/ or not. They then heard the sounds again and were asked whether they had heard /ʃ/ or not. Their reaction time for "yes" responses was recorded. Presumably, listeners would take longer to identify a poorer example of a target sound than a more canonical one.
•This was not borne out in the data. The acoustic parameters together accounted for about 15% of the variance in the reaction times to both questions. M1 (the parameter that most robustly distinguished between /s/ and /ʃ/) predicted for the most variance, at about 10%.

Direct magnitude estimation correlated moderately well with the acoustic parameters, but was difficult to administer and to score.

•In the direct magnitude estimation experiment, 20 listeners were first asked to identify the phonemes as /s/ or /ʃ/. They then heard the item a second time. For the first item, they were asked to give a number that represents how good an example that item was of the phoneme they chose. They were instructed that if the second item was twice as good, they should give it a number twice as high. If it was half as good, they should give it a number half as high. The order of the items was randomized across subjects.
•The acoustic parameters accounted for 39% of the variance in log-transformed goodness ratings for phonemes identified as /s/ and 33% of the variance in goodness ratings for phonemes identified as /ʃ/. DME ratings require more steps to analyze than the other methods we tested, as the ratings must be modulus-equalized to adjust for the differences in scale used by the different participants (Lane, Catania & Stevens, 1961).
•A drawback to the DME method were that it required careful instructions to participants. Participants who responded too quickly had only part of their response recorded, and these errors were difficult to identify and correct.



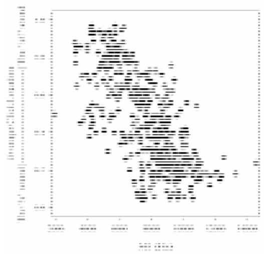
Forced-choice reaction times in the phoneme categorization task vs. M1



Direct magnitude estimates of category goodness vs. fricative M1

The visual analog scale was easy to administer and analyze, and correlated well with acoustic parameters.

•In the visual analog scale experiment, 24 listeners were asked to click the mouse on a line on the screen. The line had the words 'the "s" sound' at one end and 'the "sh" sound' at the other. Listeners were asked to click on the point on the line that corresponds with how close the sound they heard was to either endpoint.



•When compared to acoustic parameters, M1 accounted for 46.4% of the variance in click location, and the plot of M1 vs. click location appears fairly linear. Regressions for individual listeners showed similarly strong relationships between VAS ratings and acoustic measures.
• Individual differences among listeners in performance on this task are discussed in Munson, Kaiser, and Urberg-Carlson, 2008 (this conference).

We conclude that VAS rating is the best method for eliciting gradient judgments of phoneme goodness.

•VAS rating scales have great potential for application to both clinical and research studies of speech-sound acquisition. They are easy to administer and to score, and are strongly related to acoustic characteristics of stimuli.

Future experiments will test equal-appearing interval scales, and a forced-choice of phoneme followed by a visual analog scale of category goodness.

•The equal-appearing interval scale is easy to administer and score, and because it is commonly used for psychometric experiments, it is familiar to participants. Some studies have found that although more fine-grained scales give slightly better fits, the difference is not large (e.g. Toner, 1989).
•For some of our experiments—and in most potential clinical applications—we will want to compare listener's judgments of contrasts between more than 2 phonemes (e.g. /s, ʂ, ɕ/). For this reason, we will compare the VAS results presented here to a new experiment in which the participants first make a forced-choice identification of the stimulus, and then hear the stimulus again, and make a VAS judgment where the ends of the scales are "Very Bad" and "Perfect".

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 Zittnan and Celina Marnie for help running subjects. We thank Ilana Heintz, Eric Fosler-Lussier, Mary Beckman and Jan Edwards for their extremely useful input on this work.

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.