

The role of clinical experience in listening for phonetic detail in children's speech

Julie M. Johnson¹, Benjamin Munson¹, and Jan Edwards²

Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis, Department of Communicative Disorders, University of Wisconsin, Madison



Moving beyond phonetic transcription alone...

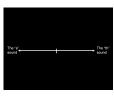
- Phonetic transcription is the 'standard-of-care' tool for the assessment of speech-sound disorders. It is fast and relatively easy to implement.

 Normative databases for the assessment of speech-sound disorders reference the number of sounds
- orrectly produced.
- The field has survived for 70+ years with this as the primary tool that we use
- However, it is ill-suited to assess the full range of productions that children produce because
- Children gradually master adult-like productions of phonemes, and often produce sounds that are
- intermediate between adult endpoints (Baum & McNutt, 1990; Li, Edwards, & Beckman, 2008).

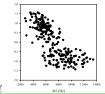
 Phonetic transcription forces users to characterize continuous development with a finite set of symbols and
- Phonetic transcriptions are discrete and erroneously suggest that phonetic development is discontinuous. Furthermore, it is clinically relevant whether or not a child is producing intermediate forms, as this predicts patterns of generalization in treatment for speech-sound disorder (Tyler et al., 1993).

...to transcription plus continuous rating scales

- Urberg-Carlson et al. (2008) and Schellinger et al. (2008) proposed that this problem can be solved by supplementing phonetic transcription with visual-analog scaling (VAS) measures (Figure 1).
- Stimuli in these studies were fricative-initial CV sequences that had been produced by children participating in the παιδολογος project (http://www.ling.ohio-state.edu/~edwards) in response to pictures of words beginning
- Urberg-Carlson et al. showed that VAS ratings of /s/-/S/ were well correlated with acoustic measures of these sounds (Figure 2).
- Other studies have shown that VAS click locations correlate with acoustic characteristics of the /s/-/T/ /d/-/D/



esponds with their perception of the sound's proximity to the trendpints



Does clinical experience affect the perception of children's speech?

- Only a small number of studies have examined whether clinical experience affects adults' perception of fine
- phonetic detail in children's speech.

 For example, Wolfe et al. (2003): clinical experience is associated with better perception of the /r/-/w/ contrast in children's speech
- It is possible that experience leads clinicians to hear more phonetic detail in children's speech
- It is also possible that the clinical practice of representing speech with the discrete labels provided by phonetic transcription make them poorer perceivers of children's speech.

Our research questions

- Do clinicians show better intra-rater reliability for VAS ratings of children's speech than do laypeople?
 Is there a closer relationship between the acoustic characteristics of children's speech and VAS ratings for clinicians than for lavneonle?

Stimuli

- The experiment comprised three tasks:
- rating 111 /t/-/k/ stimuli rating 135 /d/-/g/ stimuli
- rating 200 /s/-/T/ stimuli
- All stimuli were CVs, excised from natural productions of real words or nonwords elicited in a picture-prompted repetition task (Edwards & Beckman, 2008a, 2008b) Each task used stimuli that had been transcribed by skilled phonetic transcribers in the following six categories
- correct productions (i.e., [d] for target /d/, [s] for target /s/,etc.) substitutions ([d] for target /g/, [s] for target /T/, etc.),

- 3. productions coded as 'intermediate' between the endpoints (notated as $\{1\}[k]$, meaning 'intermediate between ll and lkl but more similar to $ll^{\prime}l$). Acoustic measures of the stimuli were also taken. These were measures of the peak frequency, the distribution of energy, and the total energy in the stop burst (for ll, ll^{\prime} , ll^{\prime}). k/, and /g/) or middle portion of the frication interval (for /s/ and /T/)

Listeners

There were two groups of listeners:

Four analyses were conducted:

- Laypeople were 20 members of the University of Minnesota community.
- Clinicians were 21 experienced (at least 2 years of professional experience after graduate school) speech-language pathologists who worked in a variety of settings with children with wide variety of communication disorders.
- Listeners in both groups were native speakers of English who had no current speech, language, or hearing impairment

Procedures

- The order of the three tasks was randomized across participants.
- In each task, listeners were presented with randomly-ordered CVs
- After each CV, they clicked on a VAS rating scale and their click location was logged automatically

- a analyses were conducted.

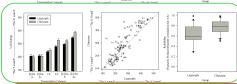
 We examined the two groups' mean VAS rating as a function of transcription category

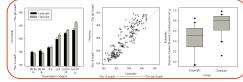
 We examined the average ratings for each stimulus, separated by groups. These first two analyses allowed us to examine gross differences in the types of ratings the
- groups provided.

 We examined differences in reliability measures (Pearson's product-moment correlations between the first and second ratings of the items repeated to assess reliability) between groups.

 We performed a series a linear mixed-effects models with crossed random effects for subjects and items (Baayen, Davidson, & Bates, 2008) to examine group
- differences in the relationship between acoustic measures and VAS ratings.

VAS Ratings Reliability by Group





Results

- ANOVA on Listener Averages The first column shows bar charts plotting mean VAS ratings by transcription category for clinicians and laypeople's
- perception of /s/-/T/, /t/-/k/, and /d/-/g/.

 ANOVA results: a significant main effect of transcription category, and a significant interaction between listener group and transcription category for all three tasks.
- There was a significant main effect of group for /t/-/k/ Interaction for /s/-/T/: the groups did not differ in their rating of the two intermediate categories and the [s] for
- T/ tokens, but did differ in the other three categories Interaction for /t/-/k/: there were significant group differences in the three /k/-like categories, but not in the
- three /t/-like categories.

 Interaction for /d/-/g/: the clinicians rated the [g] for /d/ tokens differently from the [g] for /g/ tokens, while the laypeople did not, and because the two groups differed
- significantly only for the three /g/-like categories.

 The averages for each item are shown in the middle columns These scatterplots show that the clinicians used more of the VAS scale than did the laypeople. The clinicians rated more

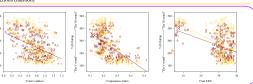
tokens on the /T/, /k/, and /g/ ends of the scale. Mann-Whitney U Tests on Reliability Measures The third column shows reliability data. Three Mann-

- Whitney U tests showed all group differences to be significan (p's ≥ 0.015). In all three tasks, the clinicians had higher intrarater reliability than the laypeople.
- Clinicians also had less within-group variation than the laypeople, particularly for the /d/-/g/ stimuli.

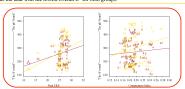
Linear Mixed-Effects Models

- Five LME models with crossed random effects for subjects and items (Baayen et al., 2008) were conducted to analyze the /s/-/T/, /t/-lk/, and /d/-/g/ ratings. Separate models were fit for the stops in front- and back-vowel contexts, as different acoustic measures discriminate between velar and alveolar stops in those contexts (Arbisi-Kelm et al., 2008).
- Fixed factors were group, the three acoustic measures, and the interaction between group and the acoustic measures. A significant interaction between group and the acoustic measures would indicate that the two groups weighted acoustic
- measures differently when making ratings.

 For the s/-/T/ stimuli, all three acoustic measures were weighted differently by the two groups: the slopes of the function relating clinicians' ratings (gold) to the acoustic measures were stronger than those for the laypeople's



- For the /d/-/g/ stimuli, there were significant interactions between all three acoustic measures in the back-vowe contexts (see below), and for two of the three measures in the front-vowel context.
- Though LME models do not provide measures of variance accounted for, regressions on items' mean ratings showed that more variance in the clinicians' ratings could be accounted for in all regressions except /t/-/k/ in front vowe contexts. This was the task with the lowest overall R2 for both groups.



Clinical experience does affect the perception of children's speech

- Clinicians.
- ...show better intra-rater reliability than laypeople when making VAS ratings of children's speech.
- ... provide ratings that are more closely related to the acoustic characteristics of stimuli than laypeople
- ...are more likely than laypeople to rate sounds as /k/, /g/, or /T/.

 Both clinicians and laypeople's VAS ratings discriminated among more than just the endpoint categories
- · Productions that were transcribed as intermediate by a single between the endpoint categories were almost always
- In many cases, sounds that were transcribed as identical were given systematically different VAS ratings;
- [s] for /s/ and [s] for /T/ substitutions
 [k] for /k/ and [k] for /t/ substitutions
- [q] for /q/ and [q] for /d/ substitutions
- One limitation of this study is that the transcription categories could not always be validated by acoustic measures, particularly for /t/-/k/.
- Clinicians' tendency to rate sounds as more /T/-, /k/- or /g/-like may reflect their greater exposure to children's emerging productions of these sounds.
- Other research in our lab (Julien et al., 2010) show that people produce hyperarticulated speech in response to hearing sounds that they perceive as inaccurate. Given this, we predict that clinicians and laypeople would provide different feedback to children's productions

This poster is available online at http://www.tc.umn.edu/~munso005