Published in final edited form as:
Am J Speech Lang Pathol. 2012 May ; 21(2): 124–139. doi:10.1044/1058-0360(2011/11-0009).

The Role of Clinical Experience in Speech-Language Pathologists' Perception of Subphonemic Detail in Children's Speech

Benjamin Munson[1], **Julie M. Johnson**[1], and **Jan Edwards**[2]

[1]Department of Speech-Language-Hearing Sciences, University of Minnesota, Twin Cities

[2]Department of Communicative Disorders, University of Wisconsin, Madison

Abstract

Purpose—This study examined whether experienced speech-language pathologists differ from inexperienced people in their perception of phonetic detail in children's speech.

Method—Convenience samples comprising 21 experienced speech-language pathologist and 21 inexperienced listeners participated in a series of tasks in which they made visual-analog scale (VAS) ratings of children's natural productions of target /s/-/θ/, /t/-/k/, and /d/-/g/ in word-initial position. Listeners rated the perception distance between individual productions and ideal productions.

Results—The experienced listeners' ratings differed from inexperienced listeners' in four ways: they had higher intra-rater reliability, they showed less bias toward a more frequent sound, their ratings were more closely related to the acoustic characteristics of the children's speech, and their responses were related to a different set of predictor variables.

Conclusions—Results suggest that experience working as a speech-language pathologist leads to better perception of phonetic detail in children's speech. Limitations and future research are discussed.

A great deal of what we currently know about both typical and atypical speech-sound development comes from studies that used phonetic transcriptions of children's speech. The validity of the data in those studies rests on the ability of the people making those transcriptions to accurately perceive and denote what children have said. Thus, it is vitally important that clinicians and researchers who assess, treat, and study children's speech production make accurate judgments and measures of children's speech. An understanding of how adults perceive children's speech begins with an understanding of how children's speech develops. Speech sound development is a gradual process that starts very early in a child's life. Toddlers' early word productions have widespread mismatches with the adult targets. Large-scale normative studies using phonetic transcription of children's speech show that productions are generally transcribed as matching transcriptions of adults' production around six years of age (Smit, Freilinger, Bernthal, Hand, & Bird, 1990).

In contrast to the results of studies of transcribed speech, acoustic and kinematic investigations suggest that speech development continues well beyond the point at which transcriptions of children's speech match those of adults' speech. Speech rate, duration, and within-subject variability in temporal, spectral, and kinematic measures decrease past the

Address Correspondence to: Benjamin Munson Department of Speech-Language-Hearing Science University of Minnesota 115 Shevlin Hall 164 Pillsbury Drive, SE Minneapolis, MN 55455 Munso005@umn.edu +1 612 624 3322 Fax: +1 612 624 7586.

end of the first decade of life (i.e., Lee, Potamianos, & Narayanan, 1999; Walsh & Smith, 2002). Part of the discrepancy between the findings of different studies relates to the nature of the dependent measures that are used. While acoustic and articulatory records of speech sounds are continuous signals (or constrained only by the limitations of the recording and storage media), transcription is categorical. That is, phonetic transcription involves parsing the inherently continuous speech signal into categories, and denoting it with a discrete—and relatively small—set of symbols. Since transcription parses all continuous variation in sounds into discrete categories, the transcriber is often required to 'round off' some sounds to their closest phonetic symbol. That is, a range of articulatory and acoustic values are denoted by the same phonetic symbol. This process of 'rounding off' is seen not only in phonetic transcription, but in many other tasks in which listeners must parse continuous variation. For example, the categorical identification functions that result when listeners are asked to identify continuously varying signals as members of a category are also the result of a 'rounding off' process.

By its nature, then, phonetic transcription results in the loss of information, as much of the phonetic detail within a transcribed category is not denoted. Despite this limitation and others (as discussed by Kent, 1996), phonetic transcription has several benefits that make it a useful analysis tool for clinicians and researchers. Transcription is a standardized means to record speech sounds and communicate them among professionals. Clinicians can use it to communicate information about clients to other clinicians as well as track clients' progress during therapy. Without a generally accepted procedure for documenting the characteristics of speech sounds, clinicians and researchers would be forced to use *ad hoc* descriptions, to acoustically analyze audio recordings, or to make direct articulatory measures. While acoustic analyses of audio recordings and articulatory measures have the potential to measure phonetic detail, both techniques have problems. Current articulatory measures, such as ultrasound and electropalatography (Bernhardt, Bacsfalvi, Gick, Radanov, & Willains, 2007) are prohibitively expensive for many clinical practices. Acoustic analysis can be time consuming and difficult to standardize. For example, consider the acoustic analysis of the voicing contrast in initial stop consonants. Word-initial voiced and voiceless stops differ in numerous acoustic parameters, including the intensity of stop burst, the voice-onset time (VOT, the interval between the release of a stop consonant and the onset of voicing in the following vowel), and the f0 and voice quality of the following vowel at its onset, among others. Standardizing acoustic analysis would require quantification of all of these parameters. On the other hand, phonetic transcription requires only training in transcription, a functioning hearing mechanism, a writing implement, and something to write on. Moreover, phonetic transcription is incorporated in standard assessment instruments for speech-sound disorder, like the Goldman-Fristoe Test of Articulation-2 (GFTA-2, Goldman and Fristoe, 2000).

The limitations of phonetic transcription are illustrated well by considering the development of contrasts among sounds. Children start out with a smaller set of contrasts than adults, and gradually transition to adult-like contrasts. The acquisition of contrast is a process of differentiation. This continuous differentiation results in the production of forms that are intermediate between adult targets. Consider the acquisition of adult-like VOTs in initial stop consonants, which was studied by Macken and Barton (1980). Macken and Barton demonstrated that as children progress from producing identical VOTs for target voiced and voiceless stops to producing adult-like values, they progress through a stage in which they produce different VOTs for target voiced and voiceless stops that are imperceptible to naïve listeners. These intermediate forms can be called *covert contrasts* between sounds. They are *contrasts* in that there is a measurable acoustic difference between targets. They are *covert* because they are not readily perceptible to inexperienced listeners. Subsequent studies found covert contrasts in the production of the /t/-/k/ contrast (Forrest et al., 1988), the contrast

between clusters and singleton stops (Scobbie et al., 2000), the contrast between /s/ and /θ/ (Baum & McNutt, 1990), and the place contrast between anterior sibilant fricatives (Li, Edwards, & Beckman, 2009). A study by Tyler, Figurski, and Langsdale (1993) demonstrated that covert contrasts have clinical significance. Tyler et al. found that children who demonstrate a covert contrast between targets (as assessed acoustically) progress through therapy more quickly and generalize correct production more readily than children who do not. Hence, it is potentially important for clinicians and researchers to determine whether children demonstrate covert contrasts so that they can effectively treat children's speech sound disorders.

One of the challenges of assessing covert contrasts in clinical contexts is the fact that they can only be documented with acoustic analysis, and not by phonetic transcription alone. Ideally, there would be a perceptual method that would allow clinicians to document the fine phonetic detail needed to determine if a child is producing a covert contrast. Recent research has shown that indeed adults *can* perceive fine phonetic detail in children's speech when given a task that does not elicit a categorical response (Munson, Edwards, Schellinger, Beckman, & Meyer, 2010; Schellinger, Edwards, & Munson, 2010; Urberg-Carlson, Munson, and Kaiser, 2008). One such technique is visual analog scaling (VAS). In VAS tasks, listeners are given a visual diagram or model that represents a perceptual parameter. An individual can indicate their perception on that visual scale. One widely used application of VAS is a pain scale, where individuals rate their pain level on a scale representing a continuum ranging from the “least possible pain” on one end to the “worst possible pain” on the other end (Bijur, Sliver, and Gallagher, 2001). VAS is part of one widely-used assessment tool for voice disorders, the Consensus Auditory-Perceptual Evaluation for Voice (CAPE-V, Kempster, Gerratt, Verdolini Abbott, Barkmeier-Kraemer, & Hillman, 2009). In the Urberg-Carlson et al. (2009) and Schellinger, Edwards, and Munson (2010) studies, a VAS scale was used that consisted of a horizontal, double-headed arrow with the labels “the 's' sound” and “the 'sh' sound” (Urberg-Carlson et al.) or the 's' sound” and “the 'th' sound” (Schellinger et al.) at either end. Listeners in those studies were presented with fricative-vowel sequences and indicated where they perceived sounds to fall on that scale. These results demonstrate that even naïve adult listeners can perceive fine phonetic detail in children's speech. In both studies, individual listeners' VAS ratings were well correlated with the acoustic parameters that differentiate the endpoint sounds. Moreover, in Schellinger et al., VAS ratings differentiated between sounds that had been transcribed as correctly produced (i.e., [s] for /s/ and [θ] for /θ/) and sounds that were transcribed as substitutions (i.e., [s] for /θ/ and [θ] for /s/). This finding in particular illustrates that VAS is a promising perception technique for assessing covert contrasts in children's speech. Subsequent studies using VAS have shown that listeners can detect fine phonetic detail in a variety of other contrasts (i.e., /d/-/ɡ/ in Arbisi-Kelm, Edwards, Munson, & Kong, 2010, and /t/-/d/ in Kong, 2009). Arbisi-Kelm et al. found similar evidence that the VAS is potentially useful for the perception of covert contrasts: listeners rated [ɡ] for /ɡ/ productions as more /ɡ/-like than [ɡ] for /d/ productions, even thought both had been transcribed as [ɡ].

All of the studies cited above used listeners without specialized training in speech and language. For VAS to be clinically useful, we must next demonstrate that experienced speech-language clinicians can also use this technique to assess children's speech. The current study examines the role of clinical experience on adults' ratings of children's speech using VAS tasks. It is possible that clinicians perceive speech differently from inexperienced listeners. Clinicians' experience hearing a broader range of variation in phonetic forms might make them even more sensitive to fine phonetic detail in speech. This prediction follows from the work of Clayards, Tanenhaus, Aslin, and Jacobs (2008), who showed that listeners who are exposed to a more-uniform distribution of a novel phonetic contrast have better perception of fine detail in that contrast than do listeners exposed to a more bimodal

distribution. Alternatively, clinicians' habitual use of phonetic transcription, which forces them to parse phonetic variation into discrete categories, might lead to their being less able to perceive fine variation.

Only a small number of published research studies have examined perception differences between inexperienced listeners and clinically trained ones. Wolfe, Martin, Borton, and Youngblood (2003) showed that speech-language pathology graduate students with clinical experience were better able to identify whether a sound was closer to a canonical /r/ or /w/ than speech-language pathology graduate students without clinical experience. Conversely, Schellinger, Edwards, and Munson (2010) found no significant effect of clinical experience when they examined the differences between the responses of graduate and undergraduate students in Communicative Disorders on a forced-choice task of categorizing children's /s/ and /θ/ productions.

The purpose of the present study was to explore whether listeners with clinical experience perceive children's productions of /t/ and /k/, /s/ and /θ/, and /d/ and /ɡ/ differently from inexperienced listeners using VAS rating tasks. Specifically, this study assessed four possible differences.

First, this study examined intra-rater reliability. Here, we reasoned that clinicians would have higher intra-rater reliability than inexperienced listeners.

Second, we examined whether experienced clinicians' VAS ratings of sounds differentiated better among categories of sounds that experienced transcribers had transcribed (as in Schellinger et al. and Arbisi-Kelm et al., described earlier) than did the ratings of inexperienced listeners. We hypothesized that experienced clinicians' superior perception of fine phonetic detail would lead them to make ratings that better differentiate among different transcription categories than do inexperienced listeners, using analyses of variance.

Third, we examined whether clinicians have a closer correlation between the VAS ratings and the acoustic characteristics of the stimuli than do inexperienced listeners. We hypothesized that experienced clinicians' superior perception would result in stronger associations between ratings and the acoustic characteristics of the stimuli being rated, using regression analyses.

Finally, we examined whether experienced listeners weight the acoustic characteristics of the stimuli differently from inexperienced listeners. Two standards were used to evaluate this. First, we examined whether a greater number of acoustic parameters predict experienced listeners ratings than inexperienced listeners ratings, using regression analyses. That is, the inexperienced listeners' ratings might be predicted by only one acoustic characteristic, while the experienced listeners might take into account several acoustic parameters. Second, we examined qualitative differences in the parameters that predict experienced and inexperienced listeners' ratings.

Methods

Participants

Forty-two listeners participated in each of the three tasks. The participants were divided into two groups. The first group consisted of 21 undergraduate students (30% male) from the University of Minnesota community between the ages of 18 and 50 years. The listeners were native speakers of North American English with no reported history of speech, language, or hearing disorders. They were recruited from the University of Minnesota community through flyers distributed on campus. This group is referred to henceforth as *inexperienced*

listeners because they had no previous clinical experience with children who have speech disorders. Slightly different analyses of the data from these same listeners have been reported previously in Schellinger, Edwards, and Munson (submitted), and Arbisi-Kelm, Edwards, Munson, and Kong (2010).

The second group was 21 licensed Speech Language Pathologists (SLPs) from the Minneapolis/St. Paul, MN metropolitan area (one man, 21 women). This differed from the gender distribution in the naïve listener group, (χ^2[df=1,n=40] = 7.449, $p = 0.006$). This group is referred to henceforth as *experienced listeners*. They were between the ages of 26 and 59 and were recruited through announcements on listservs for speech-language pathologists, and through word-of-mouth. The ages of the two groups differed significantly ($M = 40$, $SD = 14$ for the experienced listeners, $M = 26$, $SD = 7$ for the inexperienced listeners, t[30.2] = −4.719, $p < 0.001$, degrees of freedom corrected for unequal variances). Given this, age was used as a covariate in the analyses. We report the results of the analyses in which age mediated differences between experienced and inexperienced listeners. The mismatch between the ages of the two groups was due to the use of convenience sampling for both groups. However, given that a great deal of existing research on speech perception uses convenience samples like our inexperienced listeners. In contrast, the community of practicing clinicians includes people of a wide range of ages. The same can be said for the asymmetry in gender: the profession of speech-language pathology is overwhelmingly female, while convenience samples from university communities are not. Thus, while these asymmetries in demographic characteristics between groups require us to make statistical adjustments, we feel that these asymmetries give our study substantial ecological validity. We return to this point in the discussion.

The experienced listeners worked full or part time in various settings with client populations composed of infants, pre-kindergarten, elementary and secondary school age children, adults, and older individuals. Years of experience ranged from two to 40 years, with an average of 13 years experience. The experienced listeners worked with a number of disorders including apraxia, dysarthria, articulation, phonological, autism, structural anomalies, hearing loss, language, aphasia, auditory processing, learning, fluency, voice, hearing, cystic fibrosis, muscular dystrophy, and traumatic brain injury.

Prior to participating in the experiment, each experienced listener completed a background questionnaire and a self-reported experience questionnaire, along with a consent form and a standard listener questionnaire completed by all of the subjects in the larger project of which this study was a part. The background questionnaire consisted of nine questions relating to years of experience, employment status, birth year, current and previous job environments, and client characteristics, including disorder and type of populations served. Please see Table 1 for the results of this questionnaire. The self-reported expertise questionnaire had eight statements about intervention practices and decisions, along with a rating scale that included the ratings of strongly agree, agree, neutral, disagree, and strongly disagree. The participants were instructed to read the statements and use the scale to rate their level of agreement or disagreement with each statement. Examples of statements from the questionnaire include: “I regularly use phonetic transcription in therapy” and “I use evidence-based research when making intervention decisions.” See Table 2 for the full questionnaire. As this Table shows, the participants varied most in their responses to questions about the regularity with which they use phonetic transcription in clinical practice, and in their use of audio recording in assessment and treatment. Responses to the other questions were relatively uniform.

One of the questions on the standard questionnaire asked how much time in a given week the listeners spend with children, from 1 (little or no interaction) to 10 (extremely frequent

interaction). Not surprisingly, the two groups differed in their ratings (M = 5.7, SD = 2.9 for the experienced listeners, M = 2.4, SD = 1.9 for the inexperienced listeners, t[30.5] = −4.348, $p < 0.001$, degrees of freedom corrected for unequal variances). There was a wide range of scores (from 1 to 10) in both groups. This wide range allowed us to examine whether experience affects the speech perception of children's speech independently from group membership. The standard questionnaire also asked about participants' speech, language, and hearing abilities. All listeners reported normal hearing in at least one ear, and no history of significant speech or language problems.

Stimuli

The stimuli consisted of children's productions of /t/, /k/, /d/, /g/, /s/, and /θ/. The stimuli were taken from the παιδολογος database of children's speech, described in Edwards and Beckman (2008a, 2008b) and Li, Edwards, and Beckman (2009). They were produced by monolingual English speaking children aged two through five years, and were elicited through picture-prompted real-word and nonword repetition tasks. These tasks involved showing children pictures of familiar objects (in the real word task) or novel objects (for the nonword task) along with audio recordings of the real word or non-word. The children were then required to repeat what they heard. The stimuli were truncated to only include a consonant-vowel syllable, beginning with the target sounds. All of the stimuli were transcribed by a native-speaker phonetician. These contrasts were chosen because they are commonly neutralized in the speech of young children. For example, Smit et al. (1990) report that [θ] for /s/, [t] for /k/, and [d] for /g/ errors are all common in normal phonological development. They were also chosen because the stimuli were readily available, as they had been collected as part of a larger study on cross-language differences in the acquisition of lingual obstruant consonants. Preliminary summaries of the larger study can be found in Beckman and Edwards (2010), Edwards and Beckman (2008a, 2008b), Li et al. (2009), and Schellinger et al. (2010).

All stimuli were transcribed by a trained phonetician. The 200 /s/ - /θ/ stimuli included correct /s/, [θ]-for-/s/ errors, correct /θ/, [s]-for-/θ/ errors, and two types of productions that the native-speaker phonetician transcribed as `intermediate': those that were intermediate but closer to [s] (henceforth [s]:[θ]) and that were closer to [θ] (henceforth [θ]:[s]). The use of intermediate categories is consistent with Stoel-Gammon's (2001) guidelines on the transcription of the speech of children with speech-sound disorders. The 88 /t/-/k/ stimuli similarly included correct /t/, correct /k/, [t]-for-/k/ and [k]-for-/t/ substitutions, and [t]:[k] and [k]:[t] intermediate productions.

The set of /d/-/g/ stimuli was different from the other two in that it included stimuli produced both by monolingual English-acquiring children, and monolingual Greek-acquiring children. The Greek acquiring children were recorded in Thessalonika, Greece, using a picture-repetition task with real Greek words, and nonwords based on Greek phonotactics. The decision to include Greek stimuli was made partly for the sake of convenience, and partly on theoretical grounds. The experienced listeners' performance was compared to the performance of inexperienced listeners who had already been tested in a different study examining the contribution of cross-language perception differences to cross-language asymmetries in speech-sound development (Arbisi-Kelm, Edwards, Munson, & Kong, 2010). By including the same set of stimuli for both groups instead of just using the English stimuli with the experienced listeners, we ensure that the group differences that we observe are not the result of the `set effects' that happen while listeners are presented with different acoustic-phonetic distributions in speech perception experiments (as described in Keating, Mikos, & Ganong, 1981). By including the Greek stimuli in the experiment with experienced listeners, we can examine another interesting question, namely, whether experienced listeners are more- or less-susceptible to language-specific perception effects

than are inexperienced listeners. The 135 English /d/-/ɡ/ stimuli similarly included correct /d/, correct /ɡ/, [d]-for-/ɡ/ and [ɡ]-for-/d/ substitutions, and [d]:[ɡ] and [ɡ]:[d] intermediate productions. The 114 Greek /d/-/ɡ/ stimuli included the same types of stimuli. The Greek children's productions were transcribed by a native speaker of Greek. In English, the /d/ and /ɡ/ stimuli correspond to the series of stops typically described as `voiced' and spelled with <d> and <g>. In word-initial position, these are typically realized with a zero or short-lag VOT, rather than being truly voiced. Note, however, that the /d/ and /ɡ/ sounds in Greek are regularly denoted with the letters <τ> and <κ>. Though these are cognate with the English letters <t> and <k>, Greek <τ> and <κ> are produced with short-lag VOTs, like English <d> and <g>.

The stimuli were analyzed using a set of psychoacoustic measures, as described in Arbisi-Kelm, Beckman, Kong, and Edwards (2008), Arbisi-Kelm et al. (2010), and Munson et al. (2010). Briefly, these are measures of the spectra of the stop bursts and the intervals of frication that are based on models of human hearing, rather than on linear measures. For the /s/-/θ/ stimuli, the results of this analysis are presented in Table 3. The results were obtained by analyzing a 40 ms portion taken from the middle of the fricative. The fricative's total loudness, (measured in Sones, as described in Moore, Glassburg, and Baer, 1997), peak ERB (which is determined by dividing the fricatives into equivalent rectangular bandwidths and picking the loudest ERB), and the compactness index (a measure of the proportion of the fricative's loudness contained in the three ERB sequence centered at the peak ERB) were calculated. For stops, the results of this analysis are presented in Tables 4 through 9. The results were obtained by analyzing a 10 ms portion taken from the middle of the burst. The same measurements were performed on the stops, expect the peak loudness (in sones) was used instead of total loudness. The results for stops are presented separately by front- and back-vowel contexts, as the parameters that differentiate place of articulation have been found to differ between front- and back-vowel contexts (Arbisi-Kelm, Beckman, Kong, & Edwards, 2008).

As these tables show, the psychoacoustic measures differed as a function of transcription category. However, there is clearly overlap between some of the categories. Hence, the psychoacoustic measures should be seen as a means to describe the stimuli that is parallel to the transcription categories, rather than as an independent validation of the transcription categories.

Procedures

The inexperienced listeners participated in this study in a research laboratory at the University of Minnesota. Each naïve listener wore headphones (Sennheiser HD 280) and was seated in front of a computer in a sound-treated room. Six of the experienced listeners also participated in the same speech laboratory at the University of Minnesota. The remaining 15 experienced listeners participated in this study at various locations throughout the Twin Cities, including at the subject's place of residence or place of employment, in a quiet room. They wore the same brand of headphones and were seated in front of a laptop in a quiet location. For both groups of participants in all environments, instructions were presented visually on the computer screen. Participants were instructed to listen to speech sounds that consisted of consonant-vowel syllables, beginning with the target sounds, and then provide a rating of what they heard using a VAS as described above. After each stimulus, participants were instructed to use a mouse to click on a line, where one end of the line represented a perfect representation of the target sound and the other end represented a perfect representation of the other target sound. An example response screen is shown in Figure 1. For example, for the /t/ and /k/ stimuli, listeners were instructed to click on the line closest to where it said “The `t' sound” when they thought they heard a perfect “t” sound and click on the line closest to where it said “The `k' sound” when they thought they heard a

perfect “k” sound. Next, the participants were instructed that they would not always be sure the syllable began with a “t” sound or a “k” sound. In those cases they were told to click the place on the line to show whether they thought the sound was more like a “t” or a “k.” The participants were encouraged to use the whole line when rating the sound. However, they were not given any specific instructions for what to listen for when making their ratings. Participants were instructed to go with their `gut' feeling about what they heard at the beginning of the syllable. Before the participants started the experiment, they were given practice items to better familiarize themselves with the way the experiment would be conducted. These instructions were repeated for each of the three listening conditions, which consisted of children's productions of /t/ and /k/, /s/ and /θ/, and /d/ and /g/. In the /s/-/θ/ task, the example <th> words were chosen to emphasize that participants should listen for the voiceless interdental fricative. For the /d/-/g/ task, the /g/ end of the VAS line was anchored with the text “the `gh' sound”, to emphasize that participants should listen for the stop in words like <ghost> rather than the voiced affricate in words like <gem> or <giant>. Moreover, the prescriptive grammar term “hard `g'” was used in the instructions, and all example words in the instructions contained /g/. A group of naïve participants was tested on the clarity of the instructions. This group indicated that “gh” was readily and unambiguously interpretable as /g/. A subset of 10% the productions was repeated twice to assess inter-rater reliability.

Analysis

For each condition, the click location was recorded and then averaged for the different transcription categories, removing the second repetition of the reliability item. These were used as the dependent measures in a series of mixed-model analyses of variance (ANOVA). Reliability was calculated by examining the correlation (Pearson's r) between the first and second rating of each of the subset of items that were repeated to measure reliability. These were the dependent measures in a series of non-parametric Mann-Whitney U tests examining group differences in reliability.

Results

This section presents the results of statistical analyses examining the primary research questions. Prior to conducting these tests, we examined individuals' raw data to assess whether the two groups differed qualitatively in their patterns of responses on the VAS rating tasks. This was motivated in part by the findings of Munson, Kaiser, and Urberg-Carlson (2008) and Kaiser, Munson, Li, Holliday, Beckman, Edwards, and Schellinger (2009) that listeners vary in the extent to which ratings use the entire VAS scale, versus just discrete parts of the scale. To examine this, we plotted probability-density distributions of click locations on the VAS scale, irrespective of target. Separate plots were made for the two groups and the four stimulus sets. These are shown in Figure 2. Three features of these are notable. First, the experienced listeners had more clicks toward the /θ/, /k/, and /g/ ends of the VAS scales. Second, the ratings for the experienced listeners are more clearly bimodal than those of the inexperienced listeners. While the inexperienced listeners' ratings had modes at the /k/, English /g/, and Greek /g/ endpoints, they were not as strong as those for the experienced listeners. Moreover, they appeared to have no clear mode at the /θ/ endpoint of the /s/-/θ/ scale. Third, the distributions for both groups /t/-/k/, English /d/-/g/, and Greek /d/-/g/ responses showed a trimodal pattern, with clear modes at both ends of the distribution, and a weaker mode at the middle of the scale.

Nonparametric Tests of Reliability Measures

The first set of analyses examined our first research question, namely, whether the two groups differed in intra-rater reliability. A series of nonparametric Mann-Whitney U test was

used to examine whether individual subjects' reliability measures (i.e., the Pearson product-moment correlations for the subset of tokens repeated to assess reliability) differed between groups. The nonparametric test was used because the Pearson's product-moment correlations were not expected to be distributed normally. Significant differences were found for all four stimulus sets when a step-down/Holm procedure was used to correct for multiple comparisons (for /s/-/θ/ Mann-Whitney U = 104, Wilcoxon W =335, z = −2.931, p = 0.003; for /t/-/k/: Mann-Whitney U = 117, Wilcoxon W =327, z = −2.426, p = 0.015; for English /d/-/g/ Mann-Whitney U = 83, Wilcoxon W =293, z = −3.312, p = 0.001; and for Greek /d/-/g/, Mann-Whitney U = 94, Wilcoxon W =304, z = −3.026, p = 0.002). The boxplots in Figure 3 illustrate these results. As this Figure shows, the experienced listeners were more reliable than the inexperienced listeners for all four stimulus sets. The difference was somewhat greater for the two sets of /d/-/g/ stimuli than for the other two, owing largely to the especially low reliability of some of the inexperienced listeners for those continua.

A second set of analyses examined whether reliability measures were correlated with self-reported experience perceiving children's speech. Partial correlations controlling for age were used, as this differed between the two groups. None of the correlations reached or approached significance at the $\alpha < 0.05$ level, using a step-down/Holm procedure to correct for multiple comparisons. There were weak, marginally significant[1] partial correlations between English /d/-/g/ reliability and self-reported experience ($r_{partial} = 0.274$, $p = 0.091$). The relationship was in the expected direction: those with greater experience listening to children's speech were more reliable than those with less experience.

Analyses of Variance

The second analysis examined our second research question, namely, whether the two groups differed in the extent to which their ratings differentiated among the different transcription categories. Each participant's ratings were averaged across the six transcription categories for each of the four stimulus sets (/s/-/θ/, /t/-/k/, English /d/-/g/, and Greek /d/-/g/). For the /s/-/θ/ and /t/-/k/ stimuli, the same two-factor, mixed-model ANOVA was used, with one between-subjects factor, group (experienced versus inexperienced), and one within-subjects factor, transcription category (which had six different levels for each ANOVA i.e., [s] for /s/, [s] for /θ/, [s]:[g], etc.). For the /s/ and /θ/ stimuli there was a significant main effect of transcription category, $F[5,200] = 324.9$, $p < 0.001$, $\eta^2_{partial} = 0.89$. This interacted significantly with group, $F[5,200] = 8.8$, $p < 0.001$, $\eta^2_{partial} = 0.18$. Again, there was no significant main effect of group, $F[1,40] = 1.139$, $p > 0.05$. For the /t/ and /k/ stimuli there was a significant main effect of transcription category, $F[5,195] = 156.9$, $p < 0.001$, $\eta^2_{partial} = 0.80$. This interacted significantly with group, $F[5,195] = 4.7$, $p < 0.001$, $\eta^2_{partial} = 0.11$. There was no significant main effect of group, $F[1,39] = 3.302$, $p > 0.05$. For the /d/ and /g/ stimuli a different ANOVA model was used. This was a three-factor mixed-model ANOVA, with one between-subjects factor, group, and two within-subjects factor, talker language and transcription category. In this ANOVA, there was no significant main effect of group, $F[1,39] < 1.0$, $p > 0.05$. There were significant main effects of transcription category, $F[5,195] = 324.6$, $p < 0.001$, $\eta^2_{partial} = 0.89$, and talker language, $F[1,39] = 39.4$, $p < 0.001$, $\eta^2_{partial} = 0.50$ All three two-way interactions were significant: language by transcription category, $F[1,39] = 7.5$, $p = 0.009$, $\eta^2_{partial} = 0.16$, transcription category by group, $F[5,195] = 8.84$, $p < 0.001$, $\eta^2_{partial} = 0.19$, and language by transcription category, $F[5,195] = 8.84$, $p < 0.001$, $\eta^2_{partial} = 0.19$. The three-way interaction among language, transcription category, and group did not achieve statistical significance at the conventional $\alpha < 0.05$ level, but did approach this, $F[5,195] = 2.1$, $p = 0.07$, $\eta^2_{partial} = 0.05$.

[1]The term “marginally significant” is used as a proxy for the phrase “not statistically significant at the conventional $\alpha < 0.05$ level, but at the less conservative $\alpha < 0.10$ level.”

Figure 4 shows the mean VAS ratings for each of the transcription categories by group. The numbers on the y-axis represent the pixel location on the screen where the participants clicked on the VAS scale. They ranged from 90 to 535. The leftmost panel shows the mean VAS ratings for /s/-/θ/. As this figure shows, the mean click locations for both the experienced listeners and inexperienced listeners followed the transcription categories very closely. Post-hoc Bonferroni-corrected paired comparisons were significantly different for each comparison for both the experienced listeners and the inexperienced listeners. A second set of paired comparisons examined group differences for the six transcription categories using a step-down/Holm procedure. Using this method, the experienced listeners rated the [θ] for /s/ and [θ] for /θ/ stimuli as significantly more /θ/-like, and the [s] for /s/ stimuli as significantly more /s/-like than the inexperienced listeners. The second-to-left panel shows the /t/-/k/ results. In this task, post-hoc Bonferroni-corrected paired comparisons were significantly different for each comparison for both the experienced listeners and the inexperienced listeners for all but two contrasts, that between [t] for /t/ and [t] for /k/, and that between [k]:[t] and [k] for /t/. Paired comparisons between groups showed significant differences for the [k]:[t], [k] for /t/, and [k] for /k/ categories. In all three cases, experienced listeners rated the sounds as being more /k/-like.

The rightmost two panels in Figure 2 show the perception of the English and Greek /d/-/g/ stimuli, respectively. Three sets of post-hoc tests were conducted to examine the effects of transcription category, group, and talker language on these data. The first set examined differences in ratings among the six transcription categories. Separate analyses were conducted by group and by listener language. For the experienced listeners' perception of English stimuli, all pairwise comparisons were significantly different from one another except [d] for /d/ and [d] for /g/. For the inexperienced listeners, all differences were significant except that between [d] for /d/ and [d] for /g/, and between [g] for /d/ and [g] for /g/. The perception of Greek stimuli was somewhat more complex, in that both listener groups' perception did not follow the transcription categories as it had for the English /d/-/g/. For both groups, all pairwise comparisons were significant except the following: [g] for /g/ and [g] for /d/, [d] for /d/ [d]:[g], /d/ for [d] and [g]:[d], [d] for /g/ and [d]:[g], and [d] for /g/ and [g]:[d]. Moreover, the [d] for /g/ stimuli were rated as significantly more /d/-like than the [d] for /d/ stimuli were. Next, the effect of listener group was compared with a series of Bonferroni-corrected t-tests for the twelve stimuli (i.e., six transcription categories in each of the two languages). The experienced listeners rated the [g] for /g/ stimuli as significantly more /g/-like in both English and Greek. No other differences were significant. Finally, the effect of language was examined by comparing the perception of each transcription category in the two languages. These were conducted separately for the two listener groups. For the clinical listeners, a significant language effect was found for all categories except [d] for /d/ and [g] for /d/. For the inexperienced listeners, significant differences were found for all categories except [d]:[g] and [g] for /d/.

The final analysis of these data examined correlations between individual listeners' ratings and the same four self-reported measures of expertise as were described in the previous section. For the /s/-/θ/ ratings, self-reported experience was significantly correlated with ratings for the [s] for /g/ stimuli ($r_{partial} = -0.643$, $p < 0.001$), [s]:[θ] ($r_{partial} = 0.396$, $p = 0.041$), and [θ] for /θ/ stimuli ($r_{partial} = 0.486$, $p = 0.010$). Listeners with more experience rated stimuli in the latter two categories as more /θ/-like, and the stimuli in the first category as more /s/-like, than those with less experience. Ratings of [s] for /θ/ stimuli were also correlated with self-reported accuracy of phonetic transcription, and self-reported frequency of audio recording of assessments ($r_{partial} = 0.486$, $p = 0.035$ for both correlations). Listeners with greater self-reported accuracy and more-frequent use of audio recording rated the stimuli as more /s/-like. When examining ratings of the /t/-/k/ stimuli, a significant correlation was found between [k]:[t] ratings and self-reported experience ($r_{partial} = 0.350$, p

= 0.029). Additionally, there was a marginally significant correlation between [k] for /t/ ratings and experience ($r_{partial}$ = 0.275, p = 0.091). Both of these correlations suggested that greater experience perceiving children's speech is associated with greater willingness to rate sounds as /k/-like. When examining the English /d/-/g/ ratings, experience was found to correlate with ratings of [d] for /d/ stimuli ($r_{partial}$ = −0.346, p = 0.023), [d] for /g/ stimuli ($r_{partial}$ = −0.427, p = 0.007), [g] for /d/ stimuli ($r_{partial}$ = 0.356, p = 0.026), and [g] for /g/ stimuli ($r_{partial}$ = 0.428, p = 0.007). Those with more experience rated the former two categories as more /d/-like, and the latter two as more /g/-like. When examining the Greek /d/-/g/ stimuli, experience correlated significantly with ratings for [d] for /d/ stimuli ($r_{partial}$ = −0.331, p = 0.040), [d]:[g] stimuli ($r_{partial}$ = −0.400, p = 0.012), and [g] for /g/ stimuli ($r_{partial}$ = 0.402, p = 0.011). More-experienced listeners rated the former three categories as more /d/-like, and the latter as more /g/-like than listeners with less experience. The effect of experience on the perception of English [d] for /d/ and [g] for /g/ is illustrated graphically in Figure 5. This Figure plots the average ratings for those two categories by the self-reported measures of experience. As this Figure shows, there is a generally continuous relationship between self-reported experience and the degree of differentiation between those two categories.

Regression Analyses

The final set of analyses examined relationships between the psychoacoustic properties of the stimuli and listeners' responses. These analyses allowed us to examine our third and fourth research questions, whether the experienced listeners' perception was more closely related to the acoustic characteristics of the stimuli than the inexperienced listeners' perception, and whether the same acoustic parameters predicted the two groups' performance. To examine this, a series of multiple regression analyses were conducted. For each contrast, regressions were run for the average ratings for the two groups, and then separately for each of the participants, an analysis approach advocated by Lorch and Meyers (1990). The β coefficients and R^2 values from these analyses were used as the dependent measures in a series of statistical tests examining group differences, and examining relationships with measures of experience. The R^2 values allow us to answer our third research question, and the β coefficients allow us to examine our fourth research question.

***/s/-/θ/* Stimuli—**The first analysis examined the /s/-/g/ ratings. Multiple regressions predicting the average ratings for the inexperienced and experienced listeners' ratings from the peak ERB, compactness, and loudness of the fricatives were strong and significant (R^2 = 51.4%, F[1,196] = 69.201, p < 0.001; R^2 = 54.8%, F[1,196] = 79.064, p < 0.001, respectively). The β coefficients were significant for all three independent measures (for peak ERB: 2.415 for inexperienced listeners, −3.266 for experienced listeners; for compactness: −593 for inexperienced listeners, −835 for experienced listeners; for loudness: −138 for experienced listeners, −107 for experienced listeners; all t's < −3.8, all p's < 0.001). They were also all negative indicating, as expected, that stimuli were more likely to be rated as /s/-like if they had spectra with higher peak frequencies, more compact spectra, and were louder. The absolute values for the coefficients were also larger for the experienced listeners than for the experienced ones. The biggest difference was noted for the slopes for the compactness index. These are shown in Figure 6. This figure suggests that the steeper slope for the experienced listeners is due, in part, to their greater willingness to make ratings at the extreme /s/ and /θ/ ends of the VAS scale.

A series of individual regression analyses was run predicting the 42 individuals' ratings from the peak ERB, compactness, and loudness of the stimuli. Each of the regressions was significant. The individual R^2 values were lower than those for the average stimuli. The average R^2 for the experienced listeners was 40.9% (SD = 4.7%); the R^2 for the

inexperienced listeners was 30.6% (SD = 8.9%). This difference was significant in an independent-samples t-test (t[30.3] = 4.696, p < 0.001, degrees of freedom corrected for unequal variances). The β coefficients for loudness were significant for all 42 listeners. The coefficients for the compactness index were significant for all listeners except one, an inexperienced listener. The coefficients for peak ERB were significant for 19 of the experienced listeners, and 15 of the inexperienced listeners. This asymmetry did not achieve statistical significance in a chi-squared contingency test. The group differences in the β coefficients for loudness, compactness, and peak ERB were significant in independent-samples t-tests (for loudness: M = −137, SD = 37 for the experienced listeners, M = −107, SD = 51 for the inexperienced listeners, t[36.5] = −2.309, p = 0.027, degrees of freedom corrected for unequal variances; for compactness: M = −836, SD = 282 for the experienced listeners, M = −593, SD = 301 for the inexperienced listeners, t[38.8] = −2.838, p = 0.007, degrees of freedom corrected for unequal variances; for peak ERB: M = −3.29, SD = 1.10 for the experienced listeners, M = −2.41, SD = 1.49 for the inexperienced listeners, t[40] = −2.157, p = 0.037). The Cohen's d values for the differences in loudness and peak ERB were 0.65, indicating a medium-sized effect. The Cohen's d for compactness was 0.97, indicating a large effect. The differences in coefficients for compactness were significant when the single nonsignificant value was removed. Those for peak ERB did not retain significance when the seven nonsignificant coefficients were removed. The absolute values for the coefficients were consistently higher for experienced listeners than for inexperienced ones, showing that experienced listeners weighted all three acoustic characteristics more strongly than inexperienced ones. The differences were particularly marked for the weighting of compactness.

A series of partial correlations examined the extent to which measures of self-reported experience perceiving children's speech were correlated with R^2 values and which regression coefficients. Partial correlations controlling for age were used, as this differed between the two groups. None of the correlations reached or approached significance at the α < 0.05 level, using a step-down/Holm procedure to correct for multiple comparisons. However, three correlations were significant when no correction was made. The coefficient for peak ERB was significantly correlated with self-reported experience ($r_{partial}$ = −0.441, p = 0.021). This indicates that people with more experience weighted peak ERB more strongly than people with less experience and less consistent use of phonetic transcription. The R^2 values of /s/-/θ/ were also correlated with self-reported experience ($r_{partial}$ = 0.415, p = 0.031). Greater experience perceiving children's speech was associated with ratings that are more strongly correlated with the acoustics characteristics of the stimuli.

***/k/-/t/* Stimuli—**Multiple regressions predicting the average ratings for the inexperienced and experienced listeners' ratings from the peak ERB, compactness, and peak loudness of the stop bursts were conducted separately for front- and back-vowel contexts, as work by Arbisi-Kelm et al. (2008) showed that different sets of psychoacoustic measures discriminated between /t/ and /k/ in different vowel contexts. The correlations between the acoustic measures and responses was so low that the analyses were not meaningful and are hence not reported here.

***/d/-/g/* Stimuli—**Multiple regressions predicting the average ratings for the inexperienced and experienced listeners' ratings from the peak ERB, compactness, and peak loudness of the stop bursts were conducted separately for front- and back-vowel contexts, and separately by talker language. For the regression predicting English back-vowel stimuli, the addition of peak ERB resulted in a significant increase in R^2 for both the inexperienced and experienced listeners (ΔR^2= 11.9%, F[1,78] = 10.516, p = 0.002, ΔR^2= 5.6%, F[1,78] = 4.65, p = 0.034, respectively). The addition of peak sound level also resulted in a significant increase in R2 for both groups (ΔR^2= 6.9%, F[1,76] = 6.414, p = 0.013, ΔR^2= 14.9%, F[1,78] = 14.989, p

< 0.001, respectively). Finally, there was a marginally significant increase in R^2 when compactness was entered into the model predicting the experienced listeners' average ratings (ΔR^2= 3.7%, F[1,77] = 3.162, p = 0.078). Both models were significant, (Total R^2 = 19.7%, F[3,76] = 6.220, p = 0.001 for the inexperienced listeners, Total R^2 = 24.3%, F[3,76] = 8.124, p < 0.001 for the inexperienced listeners). In both groups' full models, β coefficients were significant for peak ERB (for inexperienced listeners: β = −9.052, p < 0.001; for experienced listeners: β = −9.650, p = 0.002) and peak loudness (for inexperienced listeners: β = 2.669, p = 0.013, for experienced listeners: β = 5.160, p < 0.001). For both groups, sounds were more likely to be rated as /g/-like if they had lower peak ERBs and higher peak loudness. Both of these follow the observed characteristics of the stimuli.

For the regression predicting English front-vowel stimuli, the addition of peak ERB resulted in a marginally significant increase in R^2 for the inexperienced and fully significant increase for the experienced listeners (ΔR^2= 6.2%, F[1,53] = 3.512, p = 0.066, ΔR^2= 9.1%, F[1,53] = 5.279, p = 0.026, respectively). The addition of peak sound level also resulted in a significant increase in R^2 for both groups (ΔR^2= 7.9%, F[1,51] = 4.763, p = 0.034, ΔR^2= 11.4%, F[1,51] = 7.549, p = 0.008, respectively). Both models were significant, (Total R^2 = 15.2%, F[3,51] = 3.040, p = 0.037 for the inexperienced listeners, Total R^2 = 22.7%, F[3,51] = 4.984, p = 0.004 for the experienced listeners). In both groups' full models, β coefficients were significant for peak ERB (albeit only marginally so for inexperienced listeners: β = 5.866, p = 0.087; for experienced listeners: β = 8.945, p = 0.036) and peak loudness (for inexperienced listeners: β = 2.539, p = 0.034, for experienced listeners: β = 3.961, p = 0.008). For both groups, sounds were more likely to be rated as /g/-like if they had higher peak ERBs and higher peak loudness. Both of these follow the observed characteristics of the stimuli.

The next analysis examined the coefficients for the individual subjects' regression analyses. For the regressions on the English back-vowel stimuli, 90% of the experienced listeners and 75% of the inexperienced listeners' β coefficients for peak ERB were significant. One of the experienced listener's and two of the inexperienced listeners' coefficients for compactness were significant. Neither of these asymmetries was significant in a chi-squared contingency test. In contrast, the group differences in the significance of the β coefficients for peak loudness and R^2 were significant (for peak loudness: 95% of experienced listeners, 50% of inexperienced listeners, $\chi^2_{[df=1,n=42]}$= 10.157, p = 0.001; for R^2: 100% of experienced listeners, 70% of inexperienced listeners, $\chi^2_{[df=1,n=42]}$ = 7.059, p = 0.008). For the English front-vowel stimuli, 60% of the experienced listeners and 35% of the inexperienced listeners' b coefficients were significant. One of the experienced listeners' coefficients and four of the inexperienced listeners' coefficients were significant for peak loudness. Overall, 90% of the experienced listeners' regressions were significant, and 70% of the inexperienced listeners' were. None of these asymmetries were significant. In contrast, there were group differences in the significance of the b coefficient for compactness: 85% of the experienced listeners but only 50% of the inexperienced listeners had significant coefficients ($\chi^2_{[df=1,n=42]}$ = 5.584, p = 0.018). For the regressions predicting ratings of the Greek back-vowel stimuli, β coefficients for peak ERB were significant for 60% of the experienced listeners and 30% of the inexperienced listeners. Coefficients for peak loudness were significant for 65% of the experienced listeners and 40% of the inexperienced listeners. Coefficients for compactness were significant for one listener in each group. The R^2 for these regressions was significant for 95% of experienced listeners and 75% of inexperienced listeners. None of these asymmetries reached significant in a chi-squared contingency test. For the Greek front-vowel regressions, two experienced listeners and one inexperienced listener had significant β coefficients for peak ERB. This asymmetry was not significant. There were significant asymmetries between the groups in the significance of the β coefficient for peak loudness (30% of experienced listeners, 0% of inexperienced listeners,

$\chi^2_{[df=1,n=42]} = 7.059$, $p = 0.008$), the β coefficient for compactness (55% of experienced listeners, 20% of inexperienced listeners, $\chi^2_{[df=1,n=42]} = 5.227$, $p = 0.022$), and the R^2 (75% of experienced listeners, 35% of inexperienced listeners, $\chi^2_{[df=1,n=42]} = 6.465$, $p = 0.022$).

A series of independent-samples t-tests examined group differences in the β coefficients and R^2 values from the individual subjects' regressions. There were significant group differences for the following values for the English back-vowel regression when the entire set of values was considered: peak loudness β coefficients ($M = 5.20$, $SD = 1.65$ for experienced listeners, $M = 2.66$, $SD = 1.66$ for inexperienced listeners, t[38] = −4.446, $p < 0.001$, Cohen's $d = 0.80$), compactness β coefficients ($M = 516$, $SD = 497$ for experienced listeners, $M = -83$, $SD = 586$ for inexperienced listeners, t[38] = −3.488, $p = 0.001$, Cohen's $d = 0.97$), and R2 ($M = 15.9\%$, $SD = 4.9\%$ for experienced listeners, M = 10.3%, SD = 5.3% for inexperienced listeners, t[38] = −3.434, $p = 0.001$). When only the significant values were considered, the group difference in peak sound level was significant, while the difference in R^2 was significant only at the $\alpha < 0.07$ level.

The following values for the English front-vowel regression differed between the two groups: β coefficients for peak ERB ($M = 9.21$, $SD = 4.74$ for experienced listeners, $M = 5.25$, $SD = 7.19$ for inexperienced listeners, t[38] = −2.056, $p = 0.047$, Cohen's d = 0.63), β coefficients for peak loudness ($M = 4.03$, $SD = 1.41$ for experienced listeners, $M = 2.26$, $SD = 1.77$ for inexperienced listeners, t[38] = −3.495, $p = 0.001$, Cohen's $d = 0.97$), and R2 ($M = 15.0\%$, $SD = 6.7\%$ for experienced listeners, $M = 10.0\%$, $SD = 5.2\%$ for inexperienced listeners, t[38] = −2.646, $p = 0.012$). The R^2 difference retained its significance even when the nonsignificant values were removed.

The next analysis of the /d/-/g/ perception data examined partial correlations between the self-reported experience perceiving children's speech and the R^2 and β coefficients. None of the correlations was significant when a step-down/Holm procedure was used to correct for multiple tests, but seven of the 16 correlations were significant when the uncorrected a < 0.05 level was used: peak sound level β coefficients for English back-vowel regressions ($r_{partial} = 0.376$, $p = 0.018$), compactness β coefficients for English back-vowel regressions ($r_{partial} = 0.419$, $p = 0.008$), peak ERB β coefficients for English front-vowel regressions ($r_{partial} = 0.408$, $p = 0.010$), The positive correlation coefficients show that people with more experience weighted acoustic measures more strongly or had ratings that were more strongly correlated with the acoustic measures than those with relatively less self-reported experience[2].

Discussion

Summary

This section considers the findings in light of our research questions. Consider the first research question: As predicted, experienced listeners had higher intra-rater reliability than the inexperienced listeners. The group difference is strongest for the /s/-/θ/ stimuli, weakest for the /t/-/k/ stimuli, and intermediate for the English and Greek /d/-/g/ stimuli. This finding is both expected and important. It shows that clinical experience leads listeners to be more systematic in their judgments of speech than are inexperienced listeners. It is important because reliability is an important characteristic for experienced listeners to have. Consistent rating would presumably indicate consistent feedback to clients, which would arguably help them learn target behavior more readily.

[2]A parallel set of analyses was conducted for the Greek stimuli. These were extremely similar to those examining the English stimuli, and are not reported here for the sake of space.

Consider next the second research question. By and large, both groups' ratings differentiated well between the transcription categories for all stimulus sets except the Greek /d/-/ɡ/ productions. A more striking finding from this analysis concerns group differences in apparent perceptual biases. The data in Figure 3 show that experienced listeners are more willing than inexperienced listeners to rate a sound as being /θ/-like, /k/-like, and /ɡ/-like, for both the English and Greek stimuli. In all four conditions, the primary group differences were in the perception of sounds at that end of the continuum, rather than on the /s/, /t/, or /d/ end. In only one case did the experienced listeners differ significantly from the inexperienced listeners in the opposite direction: they also perceived [s] for /s/ stimuli as more /s/-like than did the inexperienced listeners. Put differently, the inexperienced listeners were more likely to label a child's production as a sound that occurs more frequently in real words (i.e., /t/ instead of /k/, /d/ instead of /ɡ/, and /s/ instead of /θ/). One possible explanation for why the experienced listeners were more likely to rate closer to the /θ/, /k/, and /ɡ/ ends of the spectrum than inexperienced listeners is that they have more experience of working with clients on the less-commonly occurring /θ/, /k/, and /ɡ/ sounds. Working with those sounds in therapy built up their knowledge of those sounds' acoustic properties, and changed the nature of their perceptual categories for the six sounds in the study. Another possibility is simply that experienced listeners have overt awareness of children's likely substitution patterns for /k/, /ɡ/, and /θ/, and that their responses reflected perceptual compensation. We consider this latter explanation unlikely, as other studies we have done have shown that neither naïve nor experienced listeners consistently compensate for overall developmental level when perceiving children's speech (e.g., Andrzejewski, Edwards, & Kong, 2010; Munson et al., 2010; Schellinger et al., 2010). A final possibility is that experienced listeners are simply more sensitive to the acoustic properties of sounds in general and thus are less biased by the frequency of occurrence of sounds in words.

Considering the third research question, we found consistent evidence that experienced listeners' ratings were more closely related to the acoustic characteristics of the stimuli than were inexperienced listeners'. The R^2 for the experienced listeners' average ratings regressions was higher than the R^2 for inexperienced listeners for all stimulus sets, excluding the /t/-/k/ dataset, where the acoustic variables predicted very little variance in ratings. The R^2 value for individual subjects' regressions were significantly different for all stimulus sets, again excluding the /t/-/k/ stimuli.

Finally, we found some evidence that experienced listeners and inexperienced listeners weight acoustic parameters differently when rating children's speech. In general, when there were differences between the β coefficients for the experienced listeners and the inexperienced listeners, the β coefficients for the experienced listeners were larger, indicating that changes in acoustic parameters had a larger influence on experienced listeners' ratings than inexperienced ones. However, within continua, some differences in β coefficients were larger than others, indicating qualitatively different weighting of acoustic parameters. For example, in perceiving the /s/-/θ/ continuum, the differences in β coefficients between experienced listeners and inexperienced ones was disproportionately larger than the differences in the coefficients for loudness and peak ERB.

One interesting negative finding of this study is that the experienced listeners were not less susceptible to native-language effects than the inexperienced listeners were. Both groups perceived Greek- and English-acquiring children's productions of /d/ and /ɡ/ quite differently. The existence of language-specific perception of children's errors was also found by Li, Munson, Edwards, Yoneyama, and Hall (2010, in press). Li et al. examined Japanese- and English-speaking adults' perception of Japanese- and English-acquiring children's productions of /s/ and /ʃ/. The finding that experienced listeners behave similarly to

inexperienced ones suggests that clinical experience in one language is not sufficient to override language-specificity in perception.

Implications

These results are relevant for practicing clinicians for several reasons. First, they provide empirical evidence that clinical experience leads to judgments of speech sounds that are both more reliable (in that they are more consistent) and more valid (in that they are more closely related to the acoustics). They also suggest how clinical training programs might be improved. If we were to train clinicians both to hear intermediate sounds and to understand the physiological processes that generate those sounds, clinicians would be able to more accurately determine what is happening in the speech mechanism of clients. This would allow clinicians to better identify problems and to develop more effective strategies to treat these problems.

This research is also important because it demonstrates that listeners abilities to make gradient judgments of speech sounds can be learned. Previous studies of ours had shown that listeners can detect fine-grained phonetic detail given the right perception task, but these studies did not indicate to what extent these skills can be improved upon. This study shows clearly that clinicians are better able to perceive fine phonetic details of children's speech than are inexperienced listeners. This demonstrates that people can learn or improve this ability, at the very least through clinical experience, and possibly through other shorter (but more intensive) forms of training as well.

Limitations and Future Research

We see two limitations to this study. The first concerns the asymmetry in ages and sexes between the two listener groups. There are at least two reasons to believe that this did not drive the group differences that we observed. The first is that significant associations between experience and selected outcome measures were still present even when age was controlled statistically. The second (and arguably more compelling) reason is that the performance of the older, experienced listeners was consistently superior to that of the younger, inexperienced listeners, in that the experienced listeners' ratings were more reliable, better differentiated among different transcription categories, and more closely related to the acoustic characteristics of the stimuli than inexperienced listeners. If the effects we observed were due to the age-related changes in hearing sensitivity we might have expected the experienced listeners to perform more poorly. It is particularly impressive that the experienced listeners' outperformed the inexperienced listeners in their perception of the /s/-/θ/ stimuli, as these sounds are often difficult for older adults with hearing loss to perceive (Pittman & Stalmachowicz, 2000). It is notable that ours is not the only study to find that older listeners have superior performance on some speech-perception measures. A recent study by Harrington, Kleber and Reubold (2008) found that older speakers in the UK were better able to compensate for coarticulation in their perception of consonant+/u/ sequences than were younger listeners. We concede, however, that the age and gender asymmetries limit the conclusions that we can make about the actual mechanisms that lead to the superior performance of the experienced listeners.

While the decision to use unmatched groups does introduce some limitations, there were compelling *a priori* reasons to do so. As discussed in the introduction, the data in this paper are part of a larger project developing semisupervised learning models of articulatory-acoustic learning in children (see Plummer, Beckman, Belkin, Fosler-Lussier, & Munson, 2010, for an example of this modeling work). In our computational models, the error-correcting feedback is taken directly from experiments like those presented herein. Hence, it was critical that the listeners in this study exemplified the kind of variability that is present

in populations that interact with children. Subsequent models of learning that we build using the data in this presentation will have the ecological validity of being based on a realistic set of speech-language pathologists: primarily female, and ranging in age and, presumably, hearing, language, speech, and cognitive ability.

The second limitation to this study relates to the stimuli. Namely, stimuli used consisted solely of segments of words. The decision to include word fragments was made to focus the listeners on the phonetic characteristics of sounds, rather than allowing them to use contextual support (though see Julien, Munson, Beckman, Edwards, & Holliday, 2010, for evidence that knowing the intended target word does not affect inexperienced listeners' VAS ratings of the sounds /s/ and /ʃ/ in CV syllables). One possible reason why inexperienced listeners and clinician perceptions of children's speech were different could be due to the fact that clinicians are more accustomed to hearing fragments of words than inexperienced listeners. Results could have been different if the sounds used were part of words or phrases. Using whole words or phrases as stimuli could make it more difficult to identify covert contrasts, as the lexical and semantic support provided by this context might lead to a more automatic categorization of sounds based on context. Having words or phrases as stimuli could bias participants to label a speech sound based on its context rather than the actual sound that was produced. Future research could be conducted on differences in how clinicians and inexperienced listeners rate children's speech when whole words are used as stimuli. Because the results of this study indicate that clinicians are more sensitive to the psychoacoustic properties of sounds, it is plausible that they would be less susceptible to lexical bias than inexperienced listeners. Finally, our ongoing research on this topic should take into account recent findings by Yu (2010) and Stewart and Ota (2008) on sources of inter-subject differences in speech perception. Those authors found that individuals' self-ratings on a questionnaire of cognitive styles (the *Autism Spectrum Quotient*, Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001) predicted measures of attention to fine spectral detail in speech. The findings in this paper may reflect the possibility that speech-language pathologists are a self-selected group of people who happen to have a cognitive-processing style that predisposes them to attend to fine spectral detail. Future research should include measures of cognitive-processing ability like the ASQ, as well as incorporate longitudinal designs, to better understand the mechanisms that give experienced listeners the advantages we observed in this study.

Future research should also include the development of a system that could assist individuals to accurately identify intermediate sounds. This study has shown that people can learn to more accurately identify covert contrasts through experience, but it has not shown if a training system could be created to explicitly teach these skills without requiring years of clinical experience. Even though there is sound evidence that covert contrasts are pervasive in children's, we cannot mere expect clinicians and researchers to automatically hear, identify, and record them. As stated in the introduction, clinicians learn the IPA and transcription primarily based on their perception of normal adult speech. They need to be provided with new tools and training in order to effectively understand and treat clients who produce covert contrasts. This would allow clinicians to move away from treatment techniques for phonological and articulation disorders that are heavily influenced by trial and error. We hope that research like this inspires researchers and clinicians to build upon the limited symbols of the International Phonetic Alphabet to develop new ways of transcribing covert contrast, as well as new training to help speech language professionals learn to identify intermediate speech sounds and understand the physiology behind them. By following these two steps we can continue to increase the effectiveness of treatments for speech sound disorders.

Acknowledgments

This research was supported by NSF grant BCS0729277 to Benjamin Munson, and by NIH grant R01 DC02932 and NSF grant BCS0729140 to Jan Edwards. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Portions of this study were completed as part of the second author's M.A. thesis from the University of Minnesota. We generously thank Joe Reichle for extensive and very useful comments on that document and to Susan Rose for additional comments. We thank Eden Kaiser, Marie Meyer, Renata Solum, and Kari Urberg-Carlson for testing subjects and developing experimental protocols, and Eun Jong Kong and Edward Carney for assistance with data analysis.

References

Arbisi-Kelm, T.; Beckman, ME.; Kong, E.; Edwards, J. Psychoacoustic measures of stop production in Cantonese, Greek, English, Japanese, and Korean. Paper presented at the 156th Meeting of the Acoustical Society of America; Miami. 2008. p. 10-14.November 2008

Arbisi-Kelm T, Edwards J, Munson B, Kong E-J. Cross-linguistic perception of velar and alveolar obstruents: A perceptual and psychoacoustic study. Poster presentation at the Acoustical Society of America, also in *Journal of the Acoustical Society of America*. 2010; 127:1957.

Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): Evidence from asperger syndrome/highfunctioning autism, males, females, scientists and mathematicians. Journal of Autism & Developmental Disorders. 2001; 31:5–17. [PubMed: 11439754]

Baum SR, McNutt JC. An acoustic analysis of frontal misarticulation on /s/ in children. Journal of Phonetics. 1990; 18:51–63.

Beckman ME, Edwards J. Generalizing over lexicons to predict consonant mastery. Laboratory Phonology. 2010; 1:319–343. [PubMed: 21113388]

Bernhardt B, Bacsfalvi P, Gick B, Radanov B, Willains R. Exploring the use of ultrasound and electropalatography in speech habilitation. Journal of Speech, Language and Audiology. 2005; 29:169–182.

Bijur PE, Sliver W, Gallagher EJ. Reliability of the visual analog scale for measurement of acute pain. Journal of Academic Emergency Medicine. 2001; 8(12):1153–1157.

de Boysson-Bardies B, Vihman MM. Adaptation to language: Babbling and first words in four languages. Language. 1991; 67:297–319.

Clayards M, Tanenhaus MK, Aslin RN, Jacobs RA. Perception of speech reflects optimal use of probabilistic speech cues. Cognition. 2008; 108:804–809. [PubMed: 18582855]

Edwards J, Beckman ME. Methodological questions in studying phonological acquisition. Clinical Linguistics and Phonetics. 2008a; 22:939–958.

Edwards J, Beckman ME. Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in the acquisition of consonant phonemes. Language Learning & Development. 2008b; 4:122–156. [PubMed: 19890438]

Forrest K, Weismer G, Milenkovic P, Dougall RN. Statistical analysis of word-initial voiceless obstruents: Preliminary data. The Journal of the Acoustical Society of America. 1988; 84:115–123. [PubMed: 3411039]

Goldman, R.; Fristoe, M. Goldman Fristoe Test of Articulation-2. American Guidance Service; Circle Pines, MN: 2000.

Harrington J, Kleber F, Reubold U. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. Journal of the Acoustical Society of America. 2008; 123:2825–2835. [PubMed: 18529198]

Julien H, Munson B, Edwards J, Beckman M, Holliday J. Modifying speech to children based on perceived developmental level: An acoustic study of adults' fricatives. Poster presentation at the Acoustical Society of America, also in *Journal of the Acoustical Society of America*. 2010; 127:1852.

Kaiser E, Munson B, Li F, Holliday J, Beckman M, Edwards J, Schellinger S. Why do adults vary in how categorically they rate the accuracy of children's speech? Journal of the Acoustical Society of America. 2009; 125:27–53. [PubMed: 19173391]

Keating P, Mikos M, Ganong W. A cross-language study of range of voice onset time in the perception of initial stop voicing. Journal of the Acoustical Society of America. 1981; 70:1261–1271.

Kent RD. Hearing and believing: some limits to the auditory-perceptual assessment of speech and voice disorders. American Journal of Speech-Language Pathology. 1996; 5:7–23.

Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. American Journal of Speech-Language Pathology. 2009; 18:124–132. [PubMed: 18930908]

Kong, E. Unpublished Ph.D. Dissertation. Department of Linguistics, Ohio State University; Columbus, OH: 2009. The development of phonation-type contrasts in plosives: Cross-linguistic perspectives.

Li F, Edwards J, Beckman ME. Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. Journal of Phonetics. 2009; 37:111–124. [PubMed: 19672472]

Li F, Munson B, Edwards J, Yoneyama K, Hall K. Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development. Journal of the Acoustical Society of America. 2010 in press.

Lorch RF, Myers JL. Regression analyses of repeated measures data in cognitive research. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1990; 16:149–157.

Macken M, Barton D. The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. Journal of Child Language. 1980; 7:41–74. [PubMed: 7372738]

Moore BC, Glasberg BR, Baer T. A Model for the prediction of thresholds loudness, and partial loudness. Journal of Audio Engineering Society. 1997; 45:224–240.

Munson B, Edwards J, Schellinger SK, Beckman ME, Meyer MK. Deconstructing Phonetic Transcription: Covert Contrast, Perceptual Bias, and an Extraterrestrial View of Vox Humana. Clinical Linguistics and Phonetics. 2010; 24:245–260. [PubMed: 20345255]

Munson, B.; Kaiser, E.; Urberg Carlson, K. Assessment of children's speech production 3: Fidelity of responses under different levels of task delay. Poster presented at the 2008 ASHA Convention; Chicago. 2008. p. 20-22.

Oller, RK. The emergence of the sounds of speech in infancy. In: Yeni-Komshian, G.; Kavanagh, J.; Ferguson, C., editors. Child Phonology I: Production. Academic Press; New York: 1980.

Pittman A, Stalmachowicz P. Perception of voiceless fricatives by normal-hearing and hearing-impaired children and adults. Journal of Speech, Language, and Hearing Research. 2000; 43:1389–1401.

Plummer, A.; Beckman, ME.; Belkin, M.; Fosler-Lussier, E.; Munson, B. Learning speaker normalization using semisupervised manifold alignment. the Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH 2010); Makuhari, Japan. 2010. p. 2918-2921.ISSN 1990-9772

Sharf D, Ohde R, Lejman M. Relationship between the discrimination of /w-r/ and /t-d/ continua and the identification of distorted /r/. Journal of Speech and Hearing Research. 1988; 31:193–206. [PubMed: 3398493]

Schellinger S, Edwards J, Munson B. The Role of Intermediate Productions and Listener Expectations on the Perception of Children's Speech. 2010 submitted. Manuscript submitted for publication.

Scobbie, J.; Gibbon, F.; Hardcastle, W.; Fletcher, P. Covert contrast as a stage in the acquisition of phonetics and phonology. In: Broe, M.; Pierrehumbert, J., editors. Papers in Laboratory Phonology V. Cambridge University Press; Cambridge, MA: 2000. p. 194-207.

Smit AB, Freilinger JJ, Bernthal JE, Hand L, Bird A. The Iowa articulation norms project and its Nebraska replication. Journal of Speech and Hearing Disorders. 1990; 55:779–798. [PubMed: 2232757]

Stoel-Gammon C. Transcribing the speech of young children. Topics in Language Disorders. 2001; 21:12–21.

Stewart ME, Ota M. Lexical effects on speech perception in individuals with `autistic' traits. Cognition. 2008; 109:157–162. [PubMed: 18834977]

Urberg-Carlson K, Munson B, Kaiser E. Gradient measures of children's speech production: Visual analog scale and equal appearing interval scale measures of fricative goodness. Journal of the Acoustical Society of America. 2009; 125:25–29.

Walsh B, Smith A. Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. Journal of Speech, Language, and Hearing Research. 2002; 45:1119–1133.

Wolfe V, Martin D, Borton T, Youngblood HC. The effect of clinical experience on cue trading for the /r-w/ contrast. American Journal of Speech-Language Pathology. 2003; 12:221–228. [PubMed: 12828535]

Yu ACL. Perceptual compensation is correlated with individuals' “autistic” traits: Implications for models of sound change. PLoS One. 2010; 5:e11950. [PubMed: 20808859]

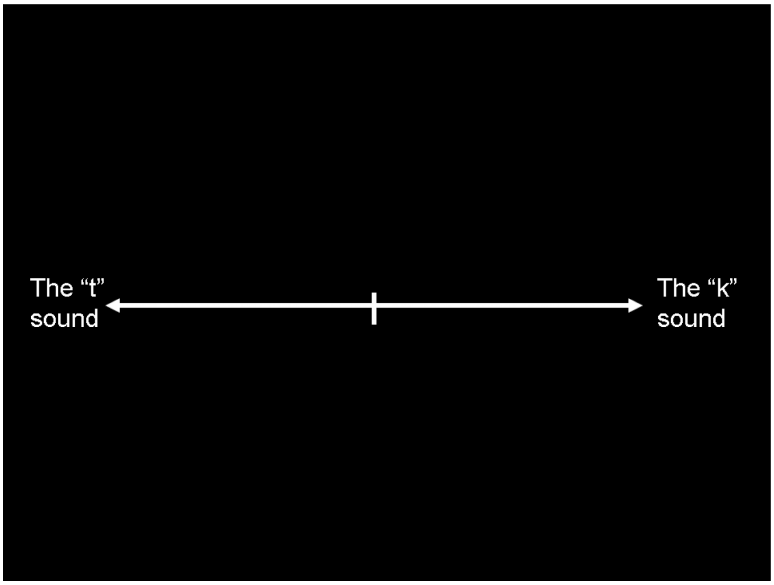

Figure 1.
Example Visual Analog Scale for the /t/-/k/ rating task.

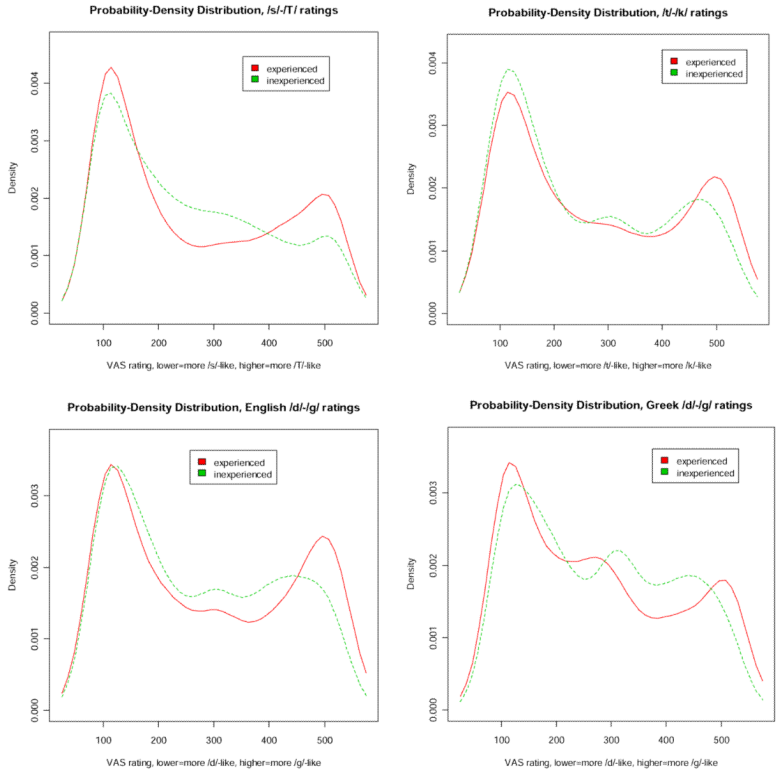

Figure 2.
Probability density distributions of ratings for the four VAS scales, separated by group. (Note: /T/ indicates /θ/)

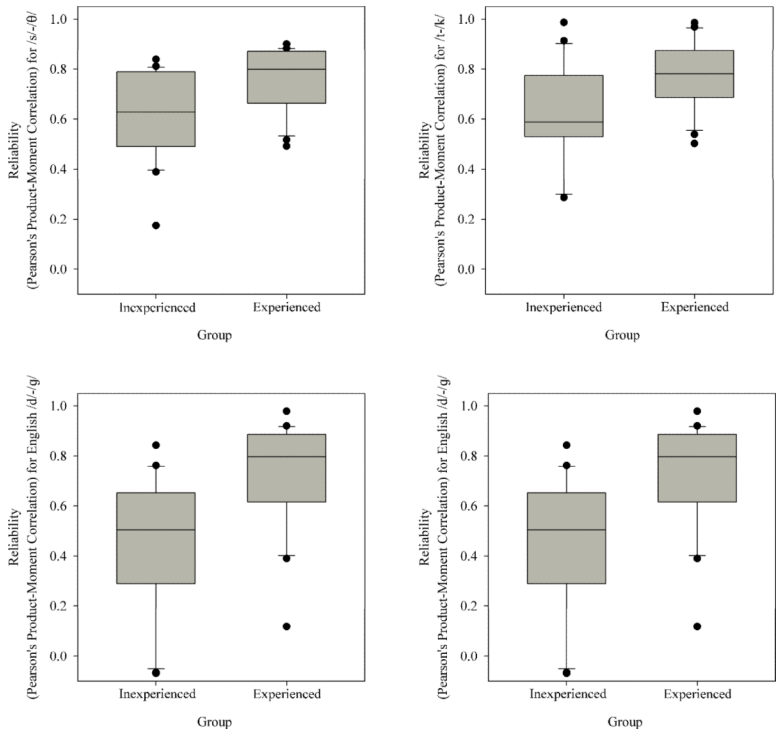

Figure 3.

Boxplots of Reliability Measurements for the four VAS scales, separated by group. Boxes extend above and below the medians to indicate the interquartile ranges and whiskers extend to indicate the extreme values.

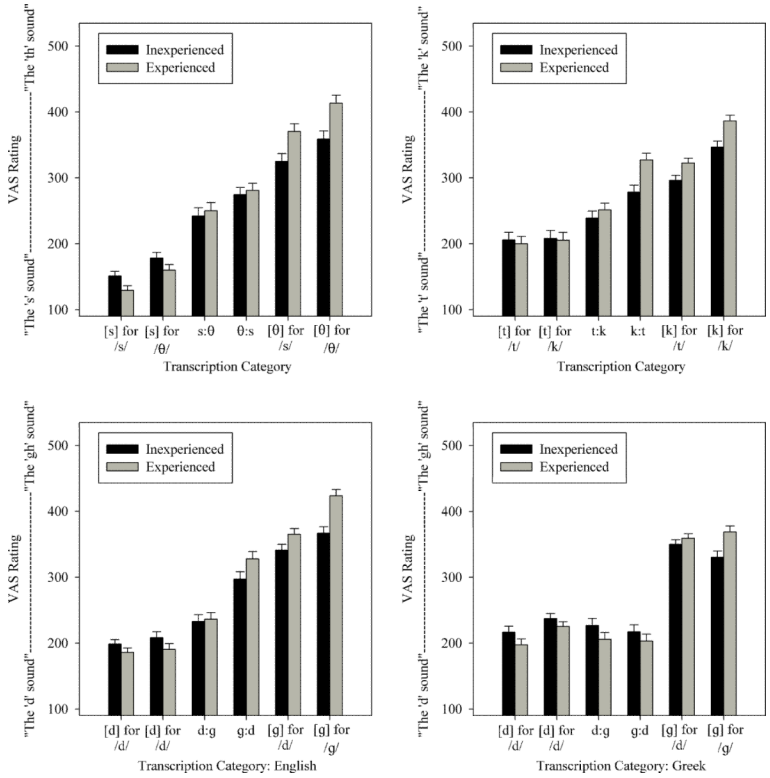

Figure 4.
Mean VAS Ratings for Each Transcription Category for each of the four contrast types, separated by group.

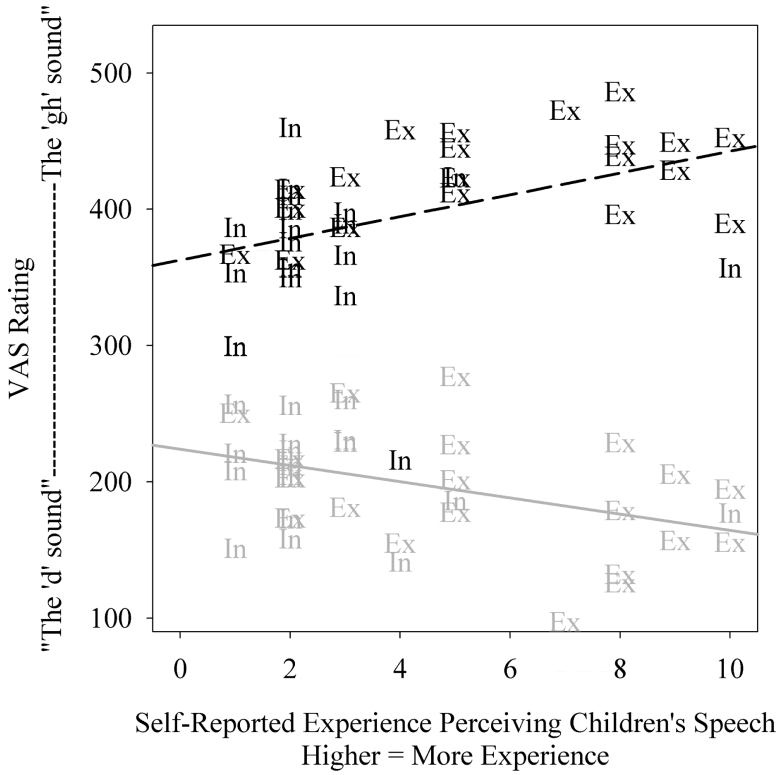

Figure 5.
Average ratings for the English [d] for /d/ (gray symbols) and [g] for /g/ (black symbols) as predicted by self-ratings of experience by the experienced listeners (shown with the “Ex” symbol) and inexperienced listeners (“In”). The lines are the outcome of regression analyses predicting ratings from experience.

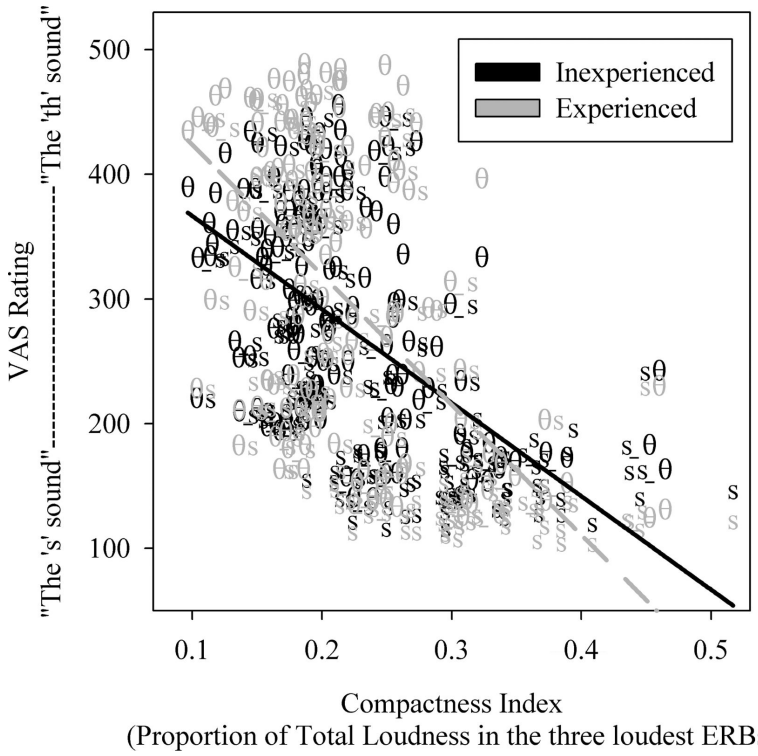

Figure 6.
Regressions predicting average ratings of the /s/-/θ/ stimuli for inexperienced listeners (black symbols) and experienced listeners (gray symbols) from the compactness index of individual tokens. Symbols indicate the transcription category (θ is [θ] for /θ/, θ_s is [θ] for /s/, θs is [θ]:[s], sθ is [s]:[θ], s_θ is [s] for /θ/, and s is [s] for /s/).

Am J Speech Lang Pathol. Author manuscript; available in PMC 2013 August 05.

Table 1

Clinician Background Information

Years of Experience	Work Status	Current Work Environment	Current Clientele	Years at Current Job	Clinical Population Disorders	Previous Work Environments & Years in Each Environment
9	FT	Elementary school	Elementary	9	Apraxia, Articulation, Phonological, Autism, Structural amoralities	
3	FT	Elementary school	Elementary	3	Did not provide	
20	PT	Elementary school	Pre-Kindergarten	10	Did not provide	Elementary school (5), Middle school (1), High school (1), Easter Seals Foundation (3)
12.5	PT	Private practice	Pre-Kindergarten, Elementary, Secondary, Adults	10	Apraxia, Dysarthria, Articulation, Phonological, Autism, Structural amoralities	High school (2)
3.5	FT	Hospital	Infants, Pre-Kindergarten, Elementary	2	Apraxia, Dysarthria, Articulation, Phonological, Autism, Structural amoralities, Hearing loss	Private practice (1.5)
7	FT	Elementary school	Elementary	7	Apraxia, Articulation, Phonological, Autism, Language, Fluency, Voice	
7.5	FT	Elementary school	Elementary	4.5	Articulation, Phonological, Autism	Middle school (1), High school (1), Early education center (2)
31	PT	Elementary school & Early education center	Pre-Kindergarten, Kindergarten	31	Apraxia, Articulation, Phonological, Autism, Structural amoralities	Elementary school (9), Middle school (1), Early education center (22)
27	FT	Elementary school	Elementary	2	Apraxia, Dysarthria, Articulation, Phonological, Autism, Structural amoralities	High school (9)
20	FT	Hospital	Infants, Pre-Kindergarten, Elementary, Secondary	3	Apraxia, Dysarthria, Articulation, Phonological, Autism, Structural amoralities, Aphasia	Elementary school (2), Hospital (6), Private practice (10)
40	PT	Outpatient Rehab	Infants, Pre-Kindergarten, Elementary, Secondary	15	Apraxia, Dysarthria, Articulation, Phonological, Autism, Structural amoralities, Auditory processing, Learning, SLI	Elementary school (25)
27	FT	Elementary school	Elementary	17	Articulation, Stuttering, Voice, Hearing impaired, Language disorder	Middle school, High school, Early education center
2	FT	Private practice	Pre-Kindergarten, Elementary, Secondary	2	Apraxia, Dysarthria, Articulation,Phonological, Autism, Structural amoralities, CP, Muscular Dys, Cocolear Implant, Auditory processing	

Years of Experience	Work Status	Current Work Environment	Current Clientele	Years at Current Job	Clinical Population Disorders	Previous Work Environments & Years in Each Environment
22	FT	Elementary school	Elementary	22	Apraxia, Articulation, Phonological, Autism	
6.5	PT	Elementary school & High school	Elementary, Adults	6.5	Articulation, Phonological, Autism, Structural amoralities, Developmental cognitive delays	Elementary school (6.5), Middle school (2), High school (1.5)
9	FT	Private practice	Pre-Kindergarten, Elementary	4	Apraxia, Dysarthria, Articulation, Phonological, Autism, Structural amoralities, Hearing loss	
3	PT	Hospital	Adults, Elderly	2	Apraxia, Dysarthria	
8	PT	Middle school	6th–8th Grade	0.5	Did not provide	Elementary school (8), Middle school (.5), High school (1), Early education center (8)
6	FT	Hospital	Infants, Pre-Kindergarten, Elementary, Secondary	0.5	Apraxia, Articulation, Phonological, Autism, Feeding, AAC	Hospital (5.5)
8	FT	Early childhood special education	Pre-Kindergarten	4	Apraxia, Dysarthria, Articulation, Phonological, Autism, Aphasia, TBI	Elementary school (1), Middle school (.5), Hospital (5), Private practice (1.5)
4	NA	Consultant	Infants	2	Articulation, Phonological, Autism, Structural amoralities, Language	Elementary school (1), High school (1), Early education center (2)

Am J Speech Lang Pathol. Author manuscript; available in PMC 2013 August 05.

Table 2

Clinician Self-reported Expertise Questionnaire

	Number of Listeners Rating the Following				
Question	**1. Strongly Agree**	**2. Agree**	**3. Neutral**	**4. Disagree**	**5. Strongly Disagree**
I can phonetically transcribe children's speech accurately.	6	13	2	0	0
I feel confident that I can accurately differentiate between a phonological disorder and a diagnosis of childhood apraxia.	4	12	3	1	0
I incorporate literacy education in my intervention methods.	11	9	1	0	0
I use evidence based research when making intervention decisions.	5	15	1	0	0
I rely on the opinions of colleagues when making clinical decisions.	5	13	3	0	0
I consider myself skilled at administering and interpreting standardized speech tests (e.g. GFTA, PAT-3, etc.).	15	6	0	0	0
I regularly use phonetic transcription in therapy.	2	9	6	4	0
I regularly audio record and review my clients' speech as part of my practice.	4	4	6	7	0

Table 3

Acoustic Characteristics of /s/-/θ/ Stimuli.

Measure	[s] for /s/		[s] for /θ/		s:θ		θ:s		[θ] for /s/		[θ] for /θ/	
	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD
N	50		24		26		30		24		46	
Peak ERB[a]	34.6	1.1	34.2	1.6	34.4	1.5	32.9	1.4	26.9	1.6	25.5	1.1
Compactness Index[a]	0.32	0.01	0.30	0.01	0.23	0.01	0.23	0.01	0.20	0.01	0.20	0.01
Total Loudness (sones)[a]	0.81	0.04	0.86	0.05	0.82	0.05	0.83	0.05	0.69	0.05	0.55	0.04

Table 4

Acoustic Characteristics of /t/-/k/ Stimuli, Front-vowel Context

Measure	[t] for /t/		[t] for /k/		t:k		k:t		[k] for /t/		[k] for /k/	
	Avg.	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**
N	5		9		8		10		12		3	
Peak ERB[a]	24.17	5.71	24.17	5.71	25.44	4.88	25.88	3.31	26.40	1.52	25.17	3.25
Compactness Index[a]	0.18	0.02	0.18	0.02	0.020	0.03	0.22	0.04	0.19	0.02	0.22	0.05
Peak Loudness (sones)[a]	45.56	8.13	45.56	8.13	53.45	10.72	47.28	9.86	49.70	10.65	51.70	7.26

Table 5

Acoustic Characteristics of /t/-/k/ Stimuli, Back-vowel Context

Measure	[t] for /t/		[t] for /k/		t:k		k:t		[k] for /t/		[k] for /k/	
	Avg.	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**
N	5		6		10		8		6		6	
Peak ERB[a]	25.60	2.41	25.89	1.83	24.50	2.62	24.40	2.07	21.64	5.41	27.33	0.58
Compactness Index[a]	0.19	0.03	0.020	0.03	0.020	0.02	0.19	0.02	0.19	0.02	0.20	0.02
Peak Loudness (sones)[a]	43.28	9.19	51.24	5.92	48.69	8.57	50.76	14.15	51.96	10.46	49.18	6.74

Note: the distribution of transcription categories in front and back-vowel contexts did not differ significantly, $\chi^2_{[df=5,n=88]} = 3.652$, $p = 0.600$

Table 6

Acoustic Characteristics of English /d/-/g/ Stimuli, Front-vowel Context

Measure	[d] for /d/		[d] for /g/		d:g		g:d		[g] for /d/		[g] for /g/	
	Avg.	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**
N	12		6		7		6		12		12	
Peak ERB[a]	25.08	4.36	23.67	5.85	25.86	1.46	26.33	1.75	25.58	3.50	26.75	1.66
Compactness Index[a]	0.20	0.03	0.19	0.03	0.20	0.04	0.20	0.05	0.20	0.04	0.22	0.05
Peak Loudness (sones)[a]	41.98	6.45	46.22	6.84	48.45	10.56	49.24	5.49	54.38	10.13	47.21	10.43

Table 7

Acoustic Characteristics of English /d/-/g/ Stimuli, Back-vowel Context

Measure	[d] for /d/		[d] for /g/		d:g		g:d		[g] for /d/		[g] for /g/	
	Avg.	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**
N	17		17		7		14		9		16	
Peak ERB[a]	25.71	2.93	23.35	5.68	23.43	5.09	24.64	1.95	21.11	5.28	23.13	2.80
Compactness Index[a]	0.18	0.02	0.19	0.02	0.19	0.03	0.020	0.03	0.18	0.02	0.20	0.02
Peak Loudness (sones)[a]	45.07	5.80	45.98	8.41	46.69	11.15	51.09	11.90	46.84	10.87	52.98	8.73

Note: the distribution of transcription categories in front and back-vowel contexts did not differ significantly, $\chi^2_{[df=5,n=135]} = 5.895$, $p = 0.307$

Table 8

Acoustic Characteristics of English /d/-/g/ Stimuli, Front-vowel Context

Measure	[d] for /d/		[d] for /g/		d:g		g:d		[g] for /d/		[g] for /g/	
	Avg.	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**
N	14		10		6		13		11		14	
Peak ERB[a]	26.14	1.66	26.30	2.21	26.00	1.55	25.85	4.58	26.91	1.81	26.50	1.83
Compactness Index[a]	0.20	0.03	0.18	0.02	0.19	0.02	0.20	0.04	0.22	0.04	0.24	0.05
Peak Loudness (sones)[a]	43.98	9.55	47.58	10.53	48.13	4.09	45.74	4.90	48.05	9.96	49.62	7.18

Am J Speech Lang Pathol. Author manuscript; available in PMC 2013 August 05.

Table 9

Acoustic Characteristics of Greek /d/-/g/ Stimuli, Back-vowel Context

Measure	[d] for /d/		[d] for /g/		d:g		g:d		[g] for /d/		[g] for /g/	
	Avg.	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**	**Avg.**	**SD**
N	11		11		2		1		10		11	
Peak ERB[a]	24.73	2.53	23.45	5.87	25.5	2.12	24	N/A	22.1	2.85	21.27	3.52
Compactness Index[a]	0.17	0.01	0.18	0.03	0.22	0.01	0.21	N/A	0.19	0.02	0.20	0.02
Peak Loudness (sones)[a]	45.23	6.35	51.17	4.84	54.93	12.14	42.00	N/A	46.10	9.90	50.24	3.52

Note: the distribution of transcription categories in front and back-vowel contexts did not differ significantly at the conventional $\alpha < 0.05$ level, though it did approach significance, $\chi^2[df=5,n=114] = 9.198$, $p = 0.101$.