The Role of Clinical Experience in Listening for Covert Contrasts in Children's Speech

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Julie M. Johnson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

Advisor: Benjamin Munson, Ph.D.

June 2010

Acknowledgements

This thesis marks the end of my journey toward obtaining a Master's degree in Speech-Language Pathology. There are several people that I would like to thank for their unending support and encouragement along the way. First, I would like to thank God for giving me the strength and courage to keep moving forward.

My utmost gratitude goes to my advisor, Benjamin Munson for his invaluable guidance and support. Your enthusiasm and immense knowledge have motivated me throughout this process. Thank you for your friendship, patience, and eagerness to help. I would also like to thank my other committee members: Joe Reichle and Sue Rose, for their support, patience, and kindness to me throughout this process.

My sincere thanks goes to my parents, Bruce Holtz and Mary Hilla, and father and mother-in-law, Jack and Sharon Johnson, for their constant encouragement, love, and confidence in me. I owe a special thanks to my sisters, Becky Anderson, Kristy Farb, and Jamie Hilla, who have always believed in me and have loved and supported me in everything that I do.

Lastly, I wish to express my loving gratitude to my husband, Mark. Thank you for your constant encouragement, love, and patience. I couldn't have done this without your support. To you I dedicate this thesis.

Abstract

Children acquire speech sounds gradually. This gradual acquisition is reflected in numerous aspects of speech-sound development, from an infant's ability to distinguish between sounds that have slight variations to the production of sounds that are identifiably adult-like. Evidence of gradual acquisition is seen in acoustic studies of children's speech-sound production, many of which have shown that children develop contrasts in certain speech sounds gradually and produce intermediate stages as they progress from incorrect to correct productions. It has also been shown that adults can perceive these fine differences in young children's speech. This study examined whether experienced speech-language pathologists perceive children's consonants differently from untrained listeners. The stimuli sets consisted of /t/-/k/ (88 tokens), /s/-/θ/ (200 tokens), and /d/-/g/ (135 tokens). Forty-two participants (21 experienced speech-language clinicians and 21 non-clinician undergraduate students) heard consonant-vowel syllables truncated from words produced by children ages two through five. Listeners were asked to provide a rating of the beginning target sound using a visual-analog scale (VAS), which contained a double-headed arrow labeled with the target sound on each side. For example, one end was labeled "the 't' sound" and the other end was labeled "the 'k' sound." The rating involved clicking on the line at a location that represented the token's proximity to an ideal /t/ or /k/. The participants' click locations on the VAS line are strongly correlated with the acoustic parameters that differentiate between the endpoint categories for a variety of contrasts for the stimuli sets. Results indicated three main differences between the way clinicians and laypeople perceived the stimuli. First,

clinicians were more willing to click closer to the ends of each scale, indicating that a token was closer to a perfect representation of the target sound; second, clinicians had higher intra-rater reliability than the naïve listeners; and third, clinicians showed a tighter relationship between the acoustic properties and the VAS ratings than laypeople.

Table of Contents

List of Tables

List of Figures

Introduction

It is vitally important for clinicians and researchers that work with children to make accurate judgments of children's speech productions. 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. While still in the womb, a fetus is exposed to sound. This is plausible due to the fact that the auditory system is functional by around the 22nd week of gestation (Moore, 2002). Even though the womb is a highly sound-attenuated environment and filters out the fine details of speech, some speech properties are available, including prosody, rhythm and the timbre of individual voices (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, and Amiel-Tison, 1988).

One of the first studies conducted on infant speech perception was by Eimas, Siqueland, Jusczyk, and Vigorito (1971). This study examined infants between the ages of one and four months old and tested their ability to discriminate between the voiced and voiceless stop consonants /b/ and /p/ through the use of the high amplitude sucking paradigm. The results indicated that infants as young as one month of age could discriminate between /b/ and /p/. This finding was initially interpreted as evidence that infants are born with the ability to categorize sounds that have slight variations. Infants are able to "sort acoustic variations of adult phonemes into categories with relatively limited exposure to speech, as well as with virtually no experience in producing these same sounds" (Eimas et al., 1971). More-recent studies have suggested that early speech

perception reflects general auditory abilities (Jusczyk, Rosner, Cutting, Foard, and Smith, 1977).

At the same time that speech perception abilities are developing rapidly, infants' vocal productions are also changing rapidly, especially during early development through the first few years of life. During the first six months of life, children's vocal productions evolve from reflexive vocal behaviors, such as crying, to sustained vocalizations (Oller, 1980). These progressions are most likely due to anatomical and physiological changes to the speech mechanism. Early canonical babbling, which is repetitive consonant-vowel combinations such as sequences of /ba/, begins to emerge at around six months of age. Near 12 months of age, children start altering their utterances to be more like the sounds of their native language. This is evident in the distribution of transcribed consonants and vowels in children's babble (de Boysson-Bardies and Vihman, 1991).Children's first words resemble adults' speech only coarsely. Examination of toddlers' speech shows widespread errors, including deletion and substitution errors. Preschoolers' speech transcriptions commonly reveal speech-sound production patterns that are constantly changing until they attain adult-like levels, which takes place at around six years of age (Smit, Freilinger, Bernthal, Hand, and Bird, 1990). Speech development is a protracted process. It does not just end when sounds are transcribed as correct. The following discussion of transcription explains in greater detail why transcriptions do not necessarily correspond with the actual sounds that were produced.

While speech sound development is a gradual process, transcription is not. Phonetic transcription involves taking speech, which is a continuous signal, and denoting

it with a discrete (and relatively small) set of symbols. For example, when transcribing the word "bend," the transcriber has to separate each sound and assign it a phonetic symbol. In the word "bend," the transcriber would use the phonemic symbol /b/ for the "b" sound, /ɛ/ for the vowel, /n/ for the "n" sounds, and /d/ for the "d" sound. The transcriber would then combine the phonetic symbols together like /bɛnd/ to phonetically transcribe the word "bend." Since transcription parses all speech sounds into discrete categories, the transcriber is often required to "round off" some sounds to their closest phonetic symbol. This sometimes leads the transcriber to categorize a sound by relying on the context it is in. This results in missing in-between sounds that do not have a phonetic symbol.

It could also be difficult to put speech sounds into discrete categories for individuals who speak other dialects. For example, it would be difficult to clearly label the vowel in the word "bend" for someone who speaks a dialect where the vowel /ɛ/ is moving toward /æ/. In this case the vowel could be intermediate between /ɛ/ and /æ/, which means the transcriber would have to "round off" their transcription of the vowel to either /ɛ/ or /æ/.

Even though limitations exist, phonetic transcription has several benefits that make it a useful analysis tool for clinicians and researchers. Transcription is a way to record speech sounds and communicate them among professionals. Without a generally accepted format for documenting speech sounds, clinicians and researchers would be forced to use either *ad hoc* descriptions or acoustic analysis of audio recordings. Audio recordings and acoustic analysis can capture the most accurate data of any method of

recording speech, however, it can be time consuming, costly, or difficult to standardize. On the other hand, phonetic transcription only requires training and a pencil and paper. Because phonetic transcription is widely taught in speech-language pathology training programs and incorporated in standard assessment instruments, clinicians can use it to communicate information about clients to other clinicians as well as track progress and generate reports. Researchers use phonetic transcription to record data from varying populations and conduct statistical analysis on large amounts of data. It would be difficult to perform large-scale studies using recorded audio or acoustic analysis.

Although there are notable benefits to using phonetic transcription, there are also significant limitations that need to be considered. As mentioned above, speech sound development is a gradual process. Children do not move directly from incorrect to correct productions of speech. Because speech development is gradual, children start out with incorrect productions and then transition into correct productions. During this time transcribers have to make judgments on the sounds they hear and label them with the closest phonetic symbol. Transcription data is used assuming everyone judges things the same way, but in actuality they don't, especially when it comes to these intermediate forms. Phonetic transcription is often taught by associating sounds from adult speech, in the language where the course is being taught, with the symbols of the phonetic alphabet. There is less emphasis on transcribing children's speech, disordered speech, and speech from other languages. This becomes problematic when clinicians are called upon to transcribe speech that does not clearly fit into one particular phonetic symbol, such as that of young children with speech disorders. Clinicians must resort to guessing, and will

undoubtedly form their own ways of transcribing children's speech that may vary greatly from clinician to clinician. This problem may be exacerbated when working with children and adults who have speech disorders or individuals with other native languages that have speech sounds that do not clearly fit into the phonetic alphabet. An individual's experience affects how they map acoustic events into phonological categories. This may vary from individual to individual based on the language(s) and dialect(s) to which they are exposed. It may also vary substantially from clinician to clinician, given the types of clients they encounter. If their training did not incorporate intermediate sounds, such as those produced by children, people with disabilities, or individuals with other native languages, then their model of the phonetic alphabet is likely based solely on their perception of adult speech in their own language.

Large-scale studies of phonetic transcriptions of children's speech have heavily influenced current understanding of speech-sound development. A relatively smaller number of studies have performed acoustic analysis on children's speech production. Acoustic analysis makes it possible to examine fine phonetic detail in speech production. As a tool, it provides a different, and arguably finer, level of detail than is provided by transcription. For example, with the use of acoustic analysis, measurements of formant frequencies can be attained and represented graphically. Spectrographic representation could help describe a sound that is intermediate between /ɛ/ and /æ/ in the previous "bend" example. Given that F1 indexes vowel height, and F2 indexes vowel backness, a researcher can use a spectrogram to determine whether a vowel was closer to a canonical /ɛ/ or to a canonical /æ/.

Studies involving acoustic analysis of children's speech production suggest that children develop contrasts in certain speech sounds gradually, and there are intermediate stages as a child progresses from incorrect to correct production of some sounds. (Macken and Barton, 1980; Edwards, Gibbon, and Fourakis, 1997; Li, Edwards, and Beckman, 2009). These intermediate stages can include covert contrasts, or a "subphonemic difference that is typically not large enough to warrant being transcribed by a different phonemic symbol, but which can be measured acoustically" (Munson, Edwards, Schellinger, Beckman, and Meyer, 2010). Studies involving acoustic analysis of children's speech productions show that sound substitutions are frequently in-between the target sound and the replacement sound. Scobbie, Gibbon, Hardcastle, and Fletcher (2000) demonstrate this with their case study on children's acquisition of word initial /s/ stop clusters. They discovered that productions of the target /st/ cluster sounds, which were transcribed as either /t/ or /d/, were actually acoustically different compared to correct /t/ and /d/ productions. Li et al. (2009) conducted a recent study that examined the acquisition of contrasts between voiceless-sibilant fricatives in two and three year old English and Japanese speaking children. They found that covert contrasts are present in the productions of some English and Japanese speaking children.

It is critical for clinicians and researchers to take into account these in-between sound stages in order to effectively treat and conduct research on children's speech production. New models need to be created that allow clinicians to effectively identify speech problems, identify the goal, and effectively move from the problem to the goal. Not only do they need to recognize these intermediate stages of sounds, they need to be

more aware of fine details in errored productions. Clinicians also need to be provided with a set of techniques that will help their clients move their articulators to the right place. This whole process begins with accurately identifying the problem. If clinicians are not provided with the necessary training to identify these intermediate sounds, they will have to rely on their best guess and learn through trial and error. Also, assuming that a particular clinician gets it right, there is no current mechanism for them to communicate their progress or findings to the next clinician that may work with their client. In the end, it is the child who is most affected by these inconsistencies. As a result, some children may go through years of speech therapy with minimal improvements in certain areas. However, this does not have to be the case. If clinicians had in their arsenal techniques to accurately identify and communicate developing and disordered speech sounds, including covert contrasts, they could more effectively prevent the above inconsistencies.

Currently, one of the challenges of incorporating covert contrasts into therapy is the fact that documenting covert contrasts requires acoustic analysis and cannot be accomplished through phonetic transcription alone. This is why techniques need to be developed to account for these covert contrasts and make them accessible to clinicians and researchers. The broad goal of this research is to develop improved tools and techniques for assessing children's speech-production accuracy. Incorporating covert contrasts into therapy requires the listener to be able to perceive graduations of fine phonetic detail in children's speech. Recently, research has shown that even naïve adult listeners can perceive these gradual changes in children's speech when they use the right

techniques (Urberg-Carlson, Munson, and Kaiser, 2009; Kaiser, Munson, Li, Holliday, Beckman, Edwards, and Schellinger, 2009).

One such technique is the use of visual analog scaling (VAS). VAS is a visual diagram or model that represents a particular perceptual element. An individual can indicate a rating on the diagram or model that they feel best correlates with their perception of the element. A common example 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). In the Urberg-Carlson et al. (2009) and Schellinger, Edwards, Munson, and Beckman (2008) studies, a VAS scale was used, which consisted of a horizontal line with written speech symbols on each end. Using this technique, the studies demonstrated that even naïve adult listeners could perceive gradual changes in children's speech by correlating listeners VAS ratings with acoustic characteristics.

This study examines the role of clinical experience on adults' ratings of children's speech in VAS tasks. It may be that clinicians perceive speech differently from naïve listeners. Some research has already examined perception differences between naïve listeners and clinically trained adults. Results from a study conducted by Wolfe, Martin, Borton, and Youngblood (2003) revealed that speech-language pathology graduate students with clinical experience were better able to identify whether a sound was closer to /r/ or /w/ than speech-language pathology graduate students without clinical experience. However, Schellinger et al. (2008) found no significant effect of clinical experience when they examined how well graduate students in communicative disorders,

compared with undergraduate students in communicative disorders, could perceive correct productions of /s/ and /θ/, clear substitutions of the two sounds, and intermediate productions. Sharf, Ohde, and Lehman (1988) conducted a study to determine if listener training could improve a participant's ability to identify subtle acoustic cues between sounds. They examined whether group-response feedback or training would improve a participant's ability to identify distorted /r/ in synthesized acoustic tokens of child-like speech in which the second and third formant onset frequencies were varied for the speech sounds /w/, /r/, and distorted /r/. Sharf et al. (1988) found some beneficial effects of the group-response feedback, however training was not effective in improving a participant's ability to identify subtle acoustic cues in /w/, /r/, and distorted /r/.

The amount of experience or training the participants had in all three of these studies was limited. The graduate students had limited clinical experience, which may or may not have even included working with children on the speech sounds used in the studies. One of the aims of this study is to test licensed clinicians with more experience. The greater the length of experience, the more likely the clinicians have been exposed to children exhibiting these "covert contrasts," and this may affect their responses.

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 /g/ differently than naïve listeners. Specifically, this study assessed three possible differences. First, it explored whether experienced clinicians have a better perception of fine phonetic detail compared to naïve listeners. This could be determined by examining whether clinicians have a closer correlation between the VAS ratings and the acoustic

characteristics of the stimuli than the naïve listeners. It was hypothesized that experienced clinicians would perceive fine phonetic detail better than naïve listeners. Next, this study assessed if clinicians weigh the acoustic characteristics of the stimuli differently. For example, the naïve listeners might just hone in on one or two acoustic characteristics, while the clinicians might take into account several acoustic differences. It was hypothesized that clinicians would take into account several more acoustic differences compared to the naïve listeners. Finally, this study examined differences in how the clinicians and naïve listeners categorized the speech sounds. It was hypothesized that clinicians would use more of the scale than the naïve listeners.

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 from the University of Minnesota 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 was classified as *naïve listeners* because they had no previous clinical experience with children who have speech disorders. The second group had 21 licensed Speech Language Pathologists (SLPs) from the Twin Cities area. This group was classified 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 SLPs worked full or part time in various settings with client populations composed of infants, pre-kindergarten, elementary and secondary school age children, adults, and elderly individuals. Years of experience ranged from 2 to 40 years with an average of 13 years experience. The clinicians 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 contained 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. Table 1 provides the clinicians' background information. 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 and Table 3 for clinician ratings of self-reported expertise.

Stimuli:

The stimuli were taken from the παιδολογος database of children's speech (Edwards and Beckman, 2008). They were produced by monolingual English speaking children ages two through five and were elicited through picture-prompted real-word and non-word repetition tasks. These tasks involved showing children pictures of familiar objects (in the real word task) or a novel object (for the non-word 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. 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 135 /d/-/g/ stimuli similarly included correct /d/, correct /g/, [d]-for-/g/ and [g]-for-/d/ substitutions, and [d]:[g] and [g]:[d] intermediate productions. These sounds were chosen because they are commonly produced in error by 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 (Edwards & Beckman, 2008). .

The stimuli were analyzed acoustically using a set of psychoacoustic measures to characterize the consonants. For fricatives, the results of this analysis are presented in Table 4a. 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 the compactness index (a measure of the distribution of energy around the peak) were calculated. For stops, the results of this analysis are presented in Table 4b through 4e. 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 as on the fricatives, except that voice onset time was used in place of vowel onset time (VOT). VOT is the amount of time between when a stop consonant is released and when voicing starts. The above psychophysical measures are described in Arbisi-Kelm, Beckman, Kong, and Edwards (2008). As these tables show, the psychoacoustic measures differed as a function of transcription category.

Procedures:

The naïve listeners participated in this study in a speech 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 isolated room. Six of the experienced

listeners also participated in the same speech laboratory at the University of Minnesota. Their listening environment was identical to the naïve listeners. 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. They also wore headphones (Sennheiser HD 280) 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. See an example VAS 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/. A subset of the productions were 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.

A linear mixed-effects model was used to analyze whether the relationship between the acoustic predictors and the VAS ratings were equivalent across listeners. In this model, items and subjects were treated as random effects. For the /s/-/θ/ analysis, the five acoustic predictors (peak ERB, compactness, total loudness, onset F2 frequency, and duration) were treated as fixed effects. For the /t/-/k/ and /d/-/g/ stimuli, the three acoustic predictors were peak sound level in the burst, the compactness of the burst, and the peak ERB in the burst. There was an additional fixed factor for the contrast between the two groups. Interaction terms were also included to assess whether there was an interaction between the acoustic variables and the factor coding the group difference.

Results

ANOVAs Examining the Influence of Group and Transcription Category on Ratings.

A two-factor, mixed-model ANOVA was used, with one between-subjects factor, (group), and one within-subjects factor, (transcription category, i.e., [s]-for-/s/, [s]-for-/θ/, [s]: [θ], etc.). 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 /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 /d/ and /g/ stimuli there was a significant main effect of transcription category, $F[5,195] = 252.3$, $p < 0.001$, $\eta^2_{partial} = 0.87$. This interacted significantly with group, $F[5,195] = 7.1$, $p < 0.001$, $\eta^2_{partial} = 0.15$. There was no significant main effect of group, $F[1,39] = 2.492$, $p > 0.05$.

Figures 2a, 2b, and 2c show 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. For figure 2b, which shows the mean VAS ratings for /s/-/θ/, there is a visible trend where by the participants clicked closest to the "s" side for the transcription category correct /s/,

closest to the "th" side for the correct /θ/, and correspondingly between the extremes for the substitutions and intermediate sounds. While the trend is visible for both clinicians and naïve listeners, the clinicians clicked closer to the extremes on both sides of the VAS line. For example, clinicians VAS ratings for correct /θ/ were closer to 500 than the naïve listeners. Figure 2c, which shows the mean VAS ratings for /d/-/g/, reveals a similar trend with the exception that the correct /d/ and [d]-for-/g/ mean VAS ratings were fairly close to each other and were not as close to the far "d" end of the VAS scale. Figure 2a, which shows the means VAS ratings for /t/-/k/, is similar to figure 2c with the exception that the clinician mean rating for the [k]:[t] intermediate sound does not fall within the trend, because they rated that category as closer to the /k/ sound than they did the [k]-for-/t/ substitution.

Figures 3a, 3b, and 3c show scatterplots mapping the average naïve listeners' VAS rating for each stimulus to the average clinicians' VAS ratings. Each point represents the average click location on the VAS, which was labeled 100-500. A point that falls right on the line shows that the average clicks for the clinicians and naïve listeners were very similar. The farther a point is from the line indicates the greater the difference between the average ratings for the clinicians and naïve listeners. All three figures show that the clinicians were more willing to click farther to the ends of the spectrums than the naïve listeners were. Note the concentration of clicks near the "s" side of the VAS on figure 3b. This is the strongest anchor visible in the three graphs. There are very slight concentrations on the "t" side of figure 3a and the "d" side of figure 3c, but overall none of the other sounds display a clear anchor like the one in figure 3b.

The lack of visible anchors in the other two figures could partially be because there were fewer stimuli in those sets.

Nonparametric Tests Examining Reliability Measures

A 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. The Pearson's product-moment correlations for reliability for /t/-/k/ differed significantly (Mann-Whitney U = 117.000, Wilcoxon W = 327.000, z = -2.426, p = 0.015). The Pearson's product-moment correlations for reliability for /s/-/θ/ differed significantly (Mann-Whitney U = 104.000, Wilcoxon W = 335.000, z = -2.931, p = 0.003). The Pearson's product-moment correlations for reliability for /d/-/g/ differed significantly (Mann-Whitney U = 83.000, Wilcoxon W = 293.000, z = -3.312, p = 0.001). The boxplots in Figures 4a, 4b, and 4c illustrate these results. They show that the clinicians' levels of reliability were about the same across the three tasks, but there was more variability within the group for the /d/-/g/ task than for the other two, as well as a greater difference between the clinicians and naïve listeners.

Linear Mixed-Effects Models Examining the Influence of Acoustic Predictors on Ratings

To examine the relationship between the acoustic characteristics of the stimuli and perception by listeners in the two groups, a linear mixed-effects model with crossed random effects for subjects and items (as described in Baayen, Davidson, and Bates, 2008) was performed. Five of these models were constructed. For the /s/-/θ/ stimuli, a single model was constructed. For both the /t/-/k/ and /d/-/g/ data, two models were constructed, one for ratings in front-vowel contexts and one for ratings in back-vowel contexts. This decision was motivated by Arbisi-Kelm et al.'s (2008) finding that the psychoacoustic measures that discriminated between velar and alveolar stops are drastically different in front-vowel and back-vowel contexts. Each model included a dummy-coded factor coding group (naïve listeners versus clinician, where clinicians were assigned a 1 and naïve listeners a 0), three factors for the three principle psychoacoustic measures needed to characterize the continuum as described in the methods, and three two-way interaction terms between each of the psychoacoustic measures and the dummy-coded variable for group.

The model for front-vowel context of /t/-/k/ is shown in Table 5a. As this table shows, the compactness index is the only significant main effect and the total loudness to listener group is the only significant interaction. The positive coefficient associated with the compactness index shows that bursts with more-compact spectra were perceived as more /k/-like than were sounds with more-distributed spectra. The positive coefficient associated with the total loudness by group interaction indicates that the clinicians

perceived bursts with louder peak frequencies to be more /k/-like, while there was no association for the naïve listeners. Figures 5a through 5e show the relationship between the VAS ratings and the psychoacoustic measures of peak loudness, peak ERB, and compactness index for the naïve listeners and the clinicians. Diverging regression lines show a different predictive relationship between psychoacoustic measures and VAS ratings. Figure 5a graphically illustrates the relationships for /t/-/k/ in front vowel contexts. The Peak Loudness graph shows that higher peak loudness levels are associated with more /k/-like ratings and that the relationship is steeper for clinicians than for the laypeople. The Compactness Index shows a similar relationship but not nearly as pronounced.

The model for back-vowel context of /t/-/k/ is shown in Table 5b. As this table shows, none of the main effects were significant and the compactness index to listener group is the only significant interaction. This shows that, for the clinicians, most-compact burst spectra were perceived as more /k/-like and that there were no significant predictors for the naïve listeners. Figure 5b graphically illustrates the relationships for /t/-/k/ in back vowel contexts. The Peak ERB graph shows a slightly stronger tendency for clinicians to rate lower peak ERB values as more /k/-like. The Compactness Index graph shows that naïve listeners were more likely to rate higher compactness values as more /t/-like.

Overall, the association between psychoacoustic measures and ratings for /t/-/k/ stimuli was weak for both groups of listeners. Linear mixed-effects models do not provide measures of effect size directly analogous to those in Ordinary Least Squares

(OLS) regression. To give the reader a sense of the effect size, four OLS multiple regressions were run predicting the mean ratings for each of the items from the psychoacoustic measures. Again, separate analyses were conducted for front- and back-vowel contexts on average ratings for the naïve listeners and the clinicians. The R^2 for the back-vowel regressions were 8.4% and 8.0% for the naïve listeners and the clinicians, respectively, while those for the front-vowel regressions were 13.2% and 19.1%, respectively.

The model for front-vowel context of /d/-/g/ is shown in Table 5c. As this table shows, the peak ERB, listener group, and total loudness were significant main effects and peak ERB to listener group and total loudness to listener group were significant interactions. The coefficients for this model show that sounds were rated as more /g/-like if the loudest ERB in the burst was at a higher frequency and had a higher peak loudness. It also shows the clinicians rated sounds as more /g/-like overall than did the naïve listeners. The interaction terms showed that the clinicians' ratings were more strongly influenced by total loudness of the burst and peak ERB in the burst than the naïve listeners' were. Figure 5c graphically illustrates the relationships for /d/-/g/ in front vowel contexts. These graphs reveal that clinicians were more likely to rate higher values of all three psychoacoustic properties as more /g/-like.

The model for back-vowel context of /d/-/g/ is shown in Table 5d. As this table shows, the peak ERB, listener group, and total loudness were significant main effects and that total loudness by listener group and compactness index by listener group were significant interactions. The coefficients showed that sounds with higher peak ERB were

more likely to be rated as /d/-like, that louder bursts were more likely to be rated as /g/-like, and that the clinicians overall rated sounds as more /g/-like than did the naïve listeners. The interactions showed that the clinicians were more strongly influenced by total loudness and by the compactness index than were the naïve listeners. Again, OLS multiple regressions were run to get a rough measure of the amount of variance in the two groups' ratings that was accounted for by the psychoacoustic measures. The R^2 for front-vowel contexts for clinicians and naïve listeners was 22.7% and 15.2%, respectively, while the R2 for back-vowel contexts was 24.3% and 19.7%, respectively. Figure 5d graphically illustrates the relationships for /d/-/g/ in back vowel contexts. The Peak Loudness and Compactness Index graphs show a significantly stronger relationship to the VAS ratings for the clinicians.

The model for /s/-/θ/ is shown in Table 5e. As this table shows, all of the main effects and interactions were significant. The main effect of group essentially replicates the effect found in the ANOVA on mean data, and shows that the clinical listeners rated things as more /θ/-like. The negative coefficients for peak ERB, total loudness, and the compactness index showed that sounds were rated as more /θ/-like if they had lower peak ERBs, lower overall loudness, and more-diffuse spectra. The negative coefficients for the interaction terms shows that the relationship between the psychoacoustic measures and the ratings were even more strongly negative than were those for the naïve listeners. Again, OLS regressions were run to gauge effect sizes for these relationships for the two groups. The variance accounted for by the clinician's ratings was 54.8%, while that for the naïve listeners was 51.4%. Figure 5e graphically illustrates the relationships for /s/-

/θ/. These graphs reveal that clinicians were more likely to rate higher values of all three psychoacoustic properties as more /s/-like.

Discussion

The results of this experiment showed three main differences between the way clinicians and naïve listeners perceive children's productions of /t/, /d/, /s/, /k/, /g/, and /θ/. One interesting finding is that the clinicians were more willing to click closer to both ends of each scale than were laypeople. That is, clinicians were more willing than the naïve listeners to rate a sound as if it were closer to an ideal exemplar of the target sound. Additionally, clinicians were more likely to rate the stimuli as /θ/, /k/, and /g/ than were the naïve listeners. This tendency is particularly well illustrated by figures 3a, 3b, and 3c. In all three conditions there appears to be more variability towards the end of the spectrum for these three sounds, which may indicate that the acoustics are more variable, making them more challenging for all of the participants to identify. One possible explanation for why the clinicians were more likely to rate closer to the /θ/, /k/, and /g/ ends of the spectrum than naïve listeners is that they had a better perception of the psychoacoustic measures than the naïve listeners and rated the stimuli correspondingly. It is noteworthy that the naïve listeners always defaulted to the sound that was more frequently occurring in real words. Perhaps the clinicians' experience of working with clients on less-commonly occurring sounds gives them greater familiarity with the acoustic properties. However, another possible explanation is that the clinicians were more confident in their perceptions, and since at the beginning of the test all of the

participants were encouraged to use the whole VAS spectrum, the clinicians simply were more willing than the naïve listeners to click closer to the ends for sounds that were more difficult to distinguish.

Another interesting finding is that clinicians had higher intra-rater reliability than the naïve listeners. Figures 4a, 4b, and 4c provide a visual representation of the reliability results of the clinicians and naïve listeners. For the stimuli that were repeated for reliability checking, the clinicians' ratings were closer to their original ratings. The group difference is strongest for the /s/-/θ/ stimuli, weakest for the /t/-/k/ stimuli, and intermediate for the /d/-/g/ stimuli. This is an expected and important finding because it shows that the clinicians were more systematic in their judgments than were the naïve listeners. Put differently, their responses more likely correlated with what they were actually hearing. Reliability is an important characteristic for clinicians to have because they need to be able to provide consistent assessment, treatment, and feedback to their clients. For example, the greater degree of reliability revealed in this study would indicate that they are more likely to correct a client the same way from one session to another when that client is exhibiting a particular covert contrast.

Finally, this study found that clinicians show a tighter relationship between the acoustic properties and the VAS ratings than naïve listeners. This is demonstrated through the linear mixed-effects models, which are illustrated by scatterplots in figures 5a through 5e. Both groups were sensitive to the acoustic predictors, but the clinicians' ratings had a stronger correlation with the acoustic properties. This is another positive outcome of this study, because it is vitally important for clinicians involved in treating

phonological and/or articulation disorders to have well-tuned listening skills. When treating speech problems, clinicians who are able to precisely identify acoustic characteristics of incorrect sounds are arguably at an advantage in developing effective treatment plans over those who can perceive things as merely 'correct' or 'incorrect'. In some cases they may be able to visually detect incorrect placement of articulators, but in many cases, such as the multitude of /r/ sounds, they must be guided by their ears alone.

The above findings are important for clinical research because even though previous studies demonstrate that it is possible for people in general to detect fine-grained assessments of speech without the use of intricate instrumentation, these studies do not indicate to what extent these skills can be learned or improved upon. Based on this study's results, clinicians are better able to perceive fine phonetic details of children's speech. This demonstrates that people can learn or improve this ability, at the very least through clinical experience. Phoneticians and transcribers involved in speech studies could potentially be trained to more accurately and consistently perceive covert contrasts and other intermediate forms. This would allow them to add a finer level of detail to the data that is collected: data that can approximate psychoacoustic analysis but can be gathered more cost effectively. If researchers were able to learn and improve on these skills, they would have more accurate data, which could lead to better understanding and improved models of speech development.

These results are also relevant for practicing clinicians because they could help them provide more effective therapy to their clients. If it is possible for researchers to improve their accuracy and consistency in perceiving fine phonetic detail then clinicians

can also be taught this skill. If training in hearing intermediate sounds was paired with a thorough understanding of the physiological processes that are generating those sounds, clinicians would be able to more accurately determine what is happening in the speech mechanism of clients. This would allow them to better identify problems and develop treatments.

One potential limitation of this study is that the stimuli used consisted solely of segments of words. While that decision was necessary at this stage in the research, it may not tell the whole story. One possible reason why naïve 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 speech sounds than naïve 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. This is because human beings tend to automatically categorize 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 naïve listeners rate children's speech when whole words are used as stimuli. Researchers could weigh the affects of categorizing sounds based on their contexts. It would be interesting to see if experienced listeners and naïve listeners' perceptions of whole word utterances would exhibit the same relationships to the acoustic characteristics as the results in this study. 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 naïve listeners.

This study only looked at three contrasting pairs of speech sounds. Additional research is needed to examine clinicians and naïve listeners' ability to distinguish fine phonetic detail of other troublesome sounds, such as liquids and glides. Those sounds are more vowel-like and would have a very different set of psychoacoustic properties than the fricatives and stops used 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 clinical experience. Even though we have evidence that shows that covert contrasts exist, we cannot just expect clinicians and researchers to automatically hear, identify, and record them. As stated in the introduction, clinicians are learning 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. Speech is far too important to let trial and error determine how it is treated. The ability to communicate is one of the most important aspects of being human and has a profound affect on a person's quality of life. This is why we need to incorporate covert contrasts and intermediate sounds into clinician training and practice. I encourage

researchers and clinicians to build upon the limited symbols of the International Phonetic Alphabet to develop new ways of transcribing covert contrasts 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 phonological disorders.

References

Arbisi-Kelm, T., Beckman, M. E., Kong, E., & Edwards, J. (2008). 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, 10–14 November 2008.

Baayen, R. H., Davidson, D. H., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390 – 412.

Bijur, P. E., Sliver, W., & Gallagher, E.J. (2001). Reliability of the visual analog scale for measurement of acute pain. *Journal of Academic Emergency Medicine, 8(12):* 1153-1157.

de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Babbling and first words in four languages. *Language, 67*, 297-319.

Edwards, J., & Beckman, M. E. (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in the acquisition of consonant phonemes. *Language Learning & Development*, 4, 122-156.

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

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303–306.

Jusczyk, P.W., Rosner, B.S., Cutting, J.E., Foard, C.F., & Smith, L.B. (1977). Categorical

perception of nonspeech sounds by 2-month-old infants. *Perception and Psychophysics, 21 (1)*: 50-54.

Kaiser, E., Munson, B., Li, F., Holliday, J., Beckman, M., Edwards, J., & Schellinger, S. (2009). Why do adults vary in how categorically they rate the accuracy of children's speech? *Journal of the Acoustical Society of America, 125*, 27-53.

Li, F., Edwards, J., & Beckman, M. E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics, 37(1):* 111-124.

Macken, M. & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language, 7*, 41-74.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition 29*, 143-178.

Moore, B. C., Glasberg, B. R., & Baer, T. (1997). A Model for the prediction of thresholds, loudness, and partial loudness. *Journal of Audio Engineering Society, 45*, 224-240.

Moore, D. R. (2002). Auditory development and the role of experience. *British Medical Bulletin 63*, 171–181.

Munson, B., Edwards, J., & Beckman, M. (in press). Phonological representations in language acquisition: Climbing the ladder of abstraction. Downloaded on April 24, 2010 from http://www.tc.umn.edu/~munso005/MunsonEdwardsBeckman_

LabPhonHandbook_Revision_08January2010.pdf.

Munson, B., Edwards, J., Schellinger, S. K., Beckman, M. E., & Meyer, M. K. (2010). Deconstructing Phonetic Transcription: Covert Contrast, Perceptual Bias, and an Extraterrestrial View of Vox Humana. *Clinical Linguistics and Phonetics, 24*, 245-260.

Oller, R.K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child Phonology I: Production*. New York: Academic Press.

Sharf, D., Ohde, R., & Lejman, M. (1988). Relationship between the discrimination of /w-r/ and /t-d/ continua and the identification of distorted /r/. *Journal of Speech and Hearing Research, 31*, 193-206.

Schellinger, S., Edwards, J., Munson, B., & Beckman, M. E. (2008). Assessment of phonetic skills in children 1: Transcription categories and listener expectations. Paper presented at the 2008 ASHA Convention, Chicago, 20-22 November 2008. Downloaded on April 3, 2010 from http://www.ling.ohio-state.edu/~edwards/ASHA08_SchellingerEtal.pdf.

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

Smit, A. B., Freilinger, J. J., Bernthal, J. E., Hand, L., & Bird, A. (1990). The Iowa

articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*, 779-798.

Stoel-Gammon, C. (2001). Transcribing the speech of young children. *Topics in Language Disorders, 21 (4):* 12-21.

Urberg-Carlson, K., Munson, B., & Kaiser, E. (2009). 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, 125*, 25-29.

Wolfe, V., Martin, D., Borton, T., & Youngblood, H.C. (2003). The effect of clinical experience on cue trading for the /r-w/ contrast. *American Journal of Speech-Language Pathology, 12*, 221-228.

Appendix A: Tables

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,	15	Apraxia, Dysarthria, Articulation,	Elementary school (25)

Years of Experience	Work Status	Current Work Environment	Current Clientele	Years at Current Job	Clinical Population Disorders	Previous Work Environments & Years in Each Environment
			Elementary, Secondary		Phonological, Autism, Structural amoralities, Auditory processing, Learning, SLI	
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	
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	6^{th}-8^{th} 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)

Table 2. Clinician Self-reported Expertise Questionnaire

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

Table 3. Results of Clinician Self-reported Expertise

Subject Number	Accurately transcribe	Phonological vs. Apraxia	Literacy education	Evidence based practice	Opinions of colleagues	Standardized tests	Regularly transcribe	Regularly audio record
1	N	SA	A	A	A	SA	D	D
2	A	A	SA	SA	A	SA	D	SA
3	A	N	A	A	A	SA	A	A
4	A	A	SA	A	SA	SA	SA	SA
5	A	A	A	A	N	A	N	D
6	SA	N	SA	SA	SA	SA	A	A
7	SA	A	A	A	A	SA	N	N
8	A	SA	SA	A	A	SA	A	D
9	SA	N	SA	SA	A	SA	A	A
10	SA	A	A	A	A	A	N	N
11	N	A	SA	A	A	A	A	D
12	A	A	SA	A	A	SA	N	SA
13	SA	A	A	A	N	SA	SA	N
14	A	D	SA	A	A	SA	D	D
15	A	A	A	N	A	A	A	A
16	A	A	A	A	A	SA	A	N
17	A	N	NA	A	A	A	D	D
18	A	SA	SA	SA	SA	SA	N	SA
19	SA	SA	SA	SA	N	SA	N	D
20	A	A	SA	A	SA	SA	A	N
21	A	A	A	A	N	A	A	N

SA=strongly agree, A=agree, N=neutral, D=disagree, SD=strongly disagree, NA=no response

Table 4a. 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 4b. Acoustic Characteristics of /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 4c. Acoustic Characteristics of /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	0.20	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 4d. 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	0.20	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 4e. 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	0.20	0.03	0.20	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 5a. Front-vowel /t/-/k/ Relationship Between Acoustic Characteristics of Stimuli and Perception by Listeners

Factor	Coefficient	St. Err.	t (df=1)	p-value
(Intercept)	122.32	110.289	1.11	0.2675
Peak ERB	-2.87	3.462	-0.83	0.4072
Listener Group	-104.08	55.556	-1.87	0.0612
Compactness Index	727.64	353.158	2.06	**0.0395**
Total Loudness	1.42	1.385	1.03	0.3053
Peak ERB *x* Listener Group	0.09	1.851	0.05	0.9609
Total Loudness *x* Listener Group	1.92	0.770	2.49	**0.0128**
Compactness Index *x* Listener Group	115.51	160.916	0.72	0.4730

Table 5b. Back-vowel /t/-/k/ Relationship Between Acoustic Characteristics of Stimuli and Perception by Listeners

Factor	Coefficient	St. Err.	t (df=1)	p-value
(Intercept)	532.21	155.514	3.42	**0.0006**
Peak ERB	-5.22	4.101	-1.27	0.2029
Listener Group	-52.42	56.521	-0.93	0.3538
Compactness Index	-497.02	666.780	-0.75	0.4561
Total Loudness	-1.17	1.468	-0.80	0.4248
Peak ERB *x* Listener Group	-2.26	1.545	-1.46	0.1446
Total Loudness *x* Listener Group	0.54	0.589	0.92	0.3561
Compactness Index *x* Listener Group	502.09	236.739	2.12	**0.0340**

Table 5c. Front-vowel /d/-/g/ Relationship Between Acoustic Characteristics of Stimuli and Perception by Listeners

Factor	Coefficient	St. Err.	t (df=1)	p-value
(Intercept)	-207.81	116.528	-1.78	0.0747
Peak ERB	8.42	3.752	2.25	**0.0248**
Listener Group	174.02	45.310	3.84	**0.0001**
Compactness Index	437.43	327.546	1.34	0.1818
Total Loudness	3.74	1.296	2.89	**0.0039**
Peak ERB *x* Listener Group	-3.33	1.433	-2.33	**0.0201**
Total Loudness *x* Listener Group	-1.61	0.493	-3.26	**0.0011**
Compactness Index *x* Listener Group	-92.12	128.109	-0.72	0.4721

Table 5d. Back-vowel /d/-/g/ Relationship Between Acoustic Characteristics of Stimuli and Perception by Listeners

Factor	Coefficient	St. Err.	t (df=1)	p-value
(Intercept)	188.90	101.736	1.86	0.0634
Peak ERB	-9.69	2.644	-3.67	**0.0003**
Listener Group	139.06	37.375	3.72	**0.0002**
Compactness Index	483.66	427.080	1.13	0.2575
Total Loudness	5.12	1.180	4.34	**<0.0001**
Peak ERB *x* Listener Group	1.67	0.935	1.79	0.0741
Total Loudness *x* Listener Group	-2.68	0.417	-6.42	**<0.0001**
Compactness Index *x* Listener Group	-354.18	151.458	-2.34	**0.0194**

Table 5e. /s/-/θ/ Relationship Between Acoustic Characteristics of Stimuli and Perception by Listeners

Factor	Coefficient	St. Err.	t (df=1)	p-value
(Intercept)	557.51	26.149	21.32	**<0.0001**
Peak ERB	-2.42	0.698	-3.46	**0.0005**
Listener Group	124.21	14.290	8.69	**<0.0001**
Compactness Index	-592.60	75.721	-7.84	**<0.0001**
Total Loudness	-106.69	19.946	-5.36	**<0.0001**
Peak ERB *x* Listener Group	-0.87	0.259	-3.36	**0.0008**
Total Loudness *x* Listener Group	-31.55	7.414	-4.26	**<0.0001**
Compactness Index *x* Listener Group	-243.17	28.146	-8.64	**<0.0001**

Appendix B: Graphs and Figures

Figure 1. Example Visual Analog Scale

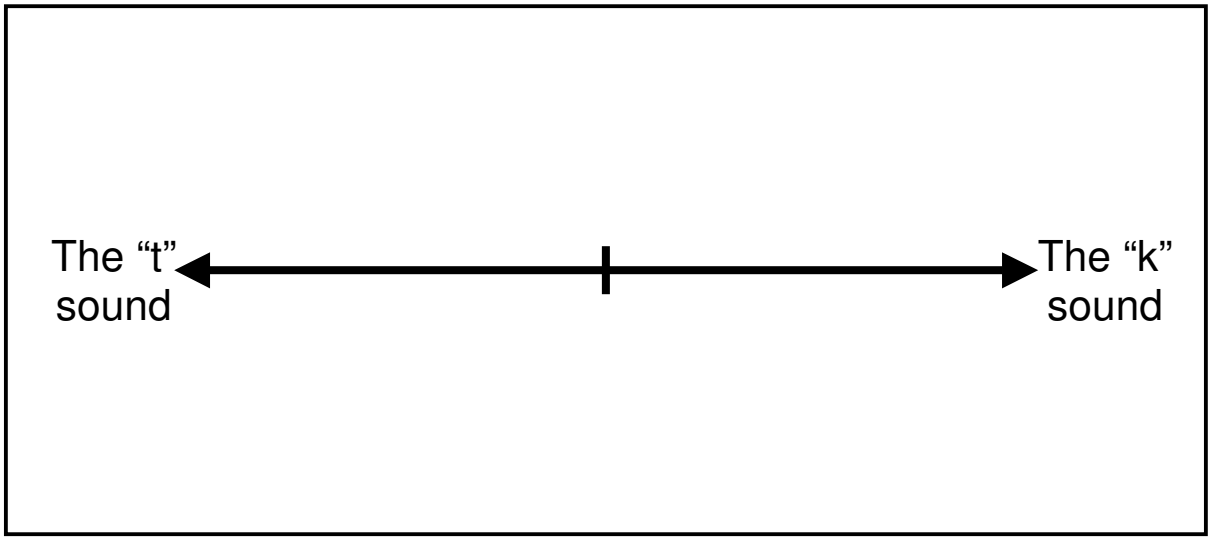

Figure 2a. /t/ and /k/ Mean VAS Ratings for Each Transcription Category by Group

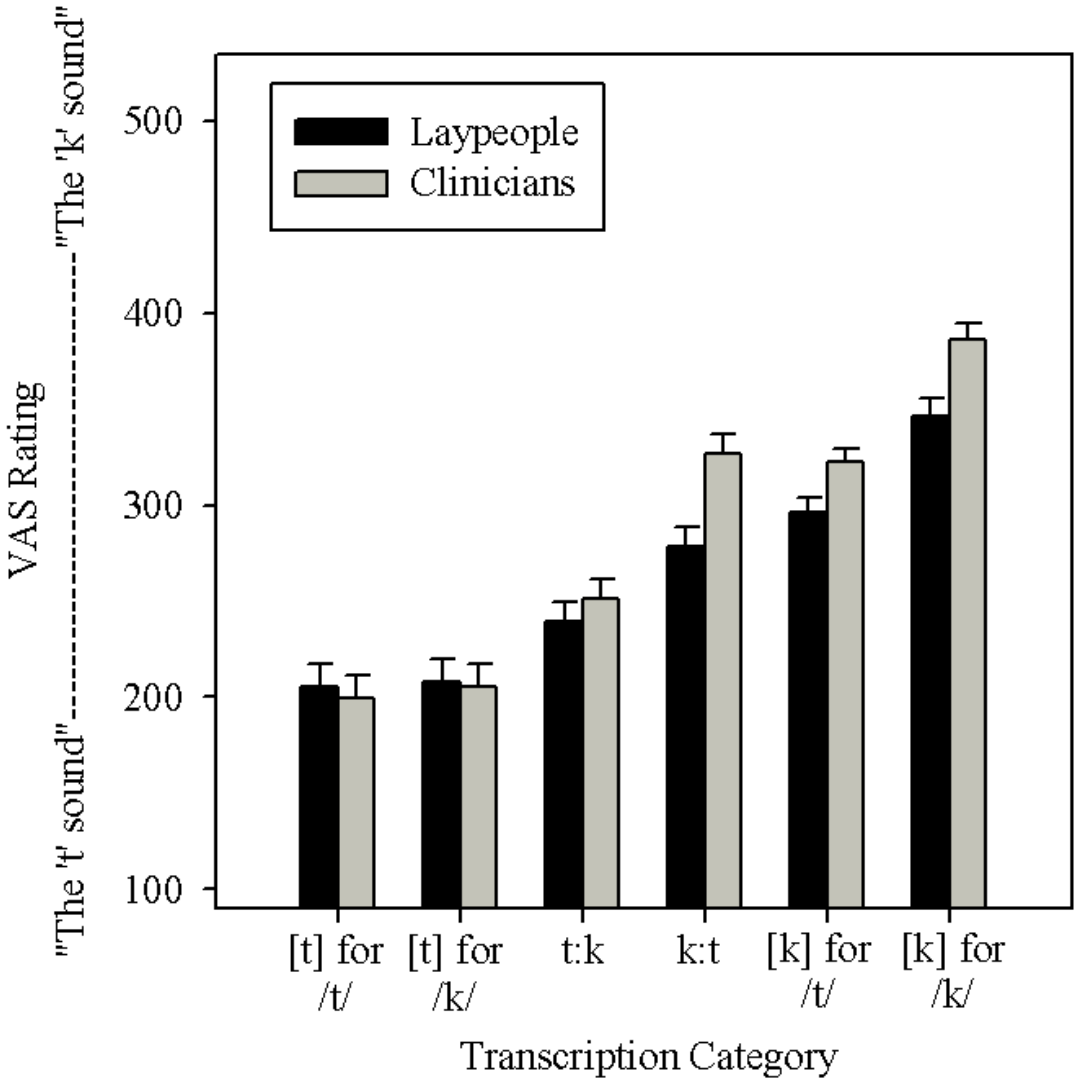

Figure 2b. /s/ and /θ/ Mean VAS Ratings for Each Transcription Category by Group

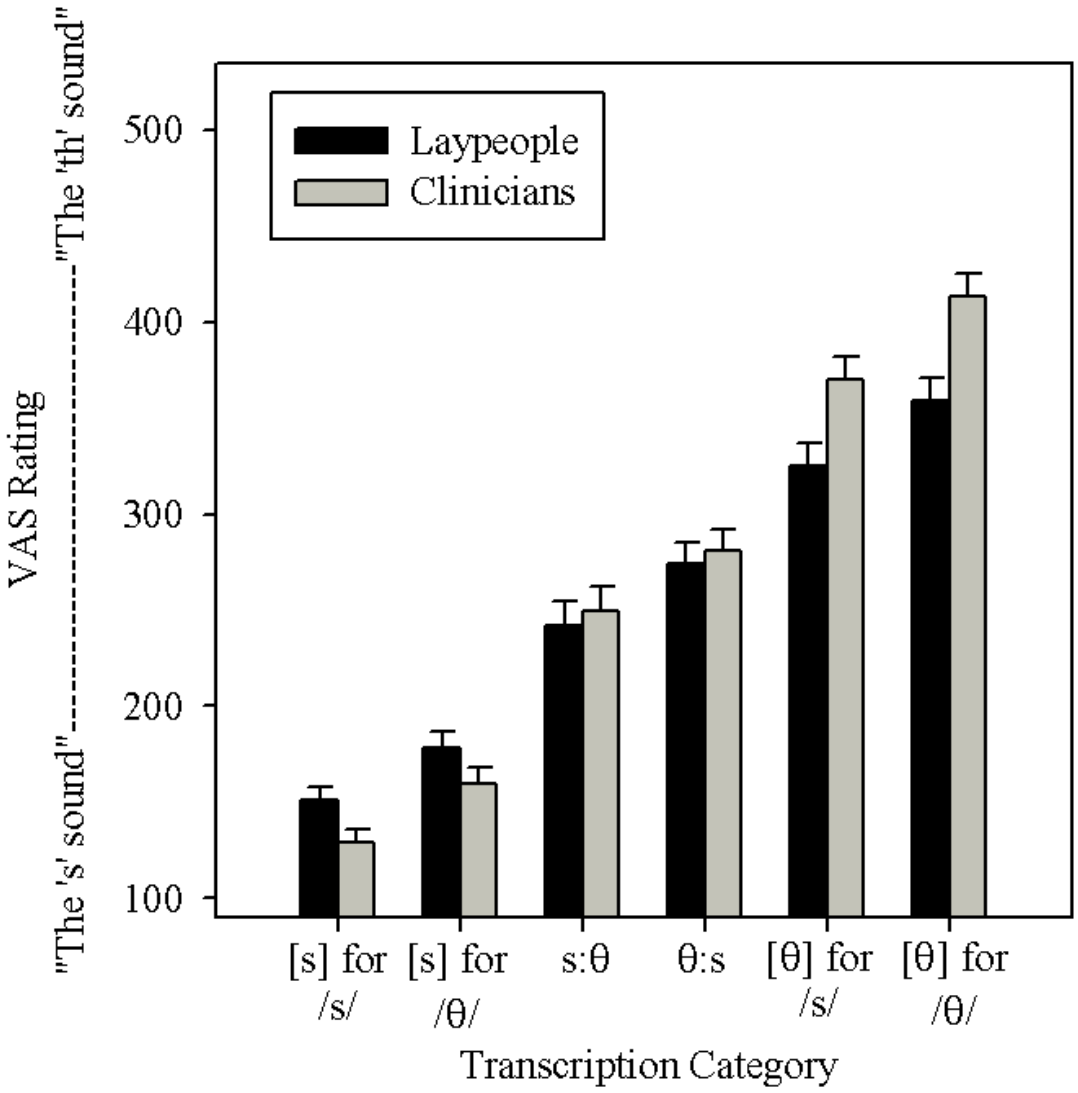

Figure 2c. /d/ and /g/ Mean VAS Ratings for Each Transcription Category by Group

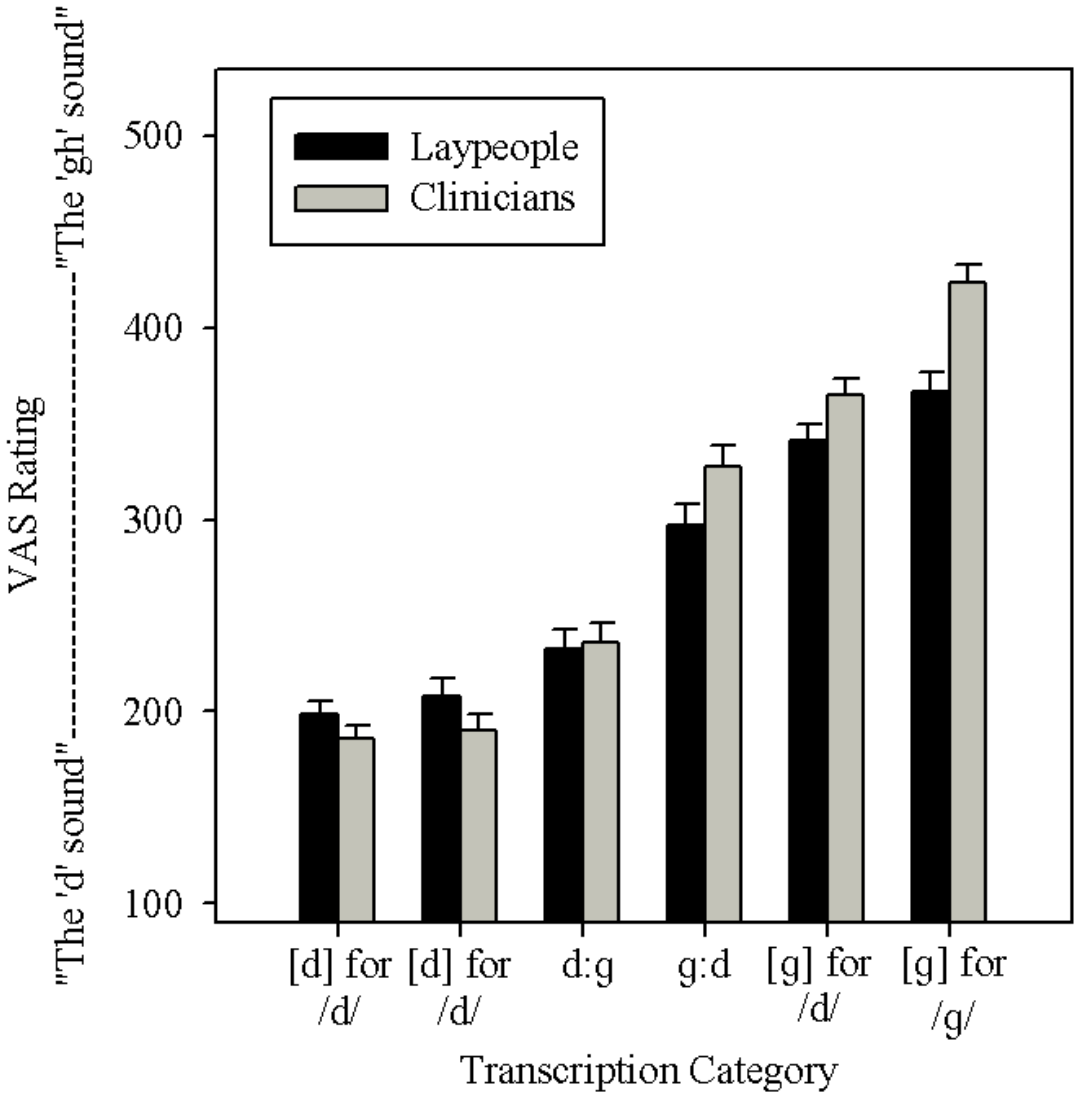

Figure 3a. /t/ and /k/ Mean Laypersons' VAS rating to Mean Clinicians' Ratings

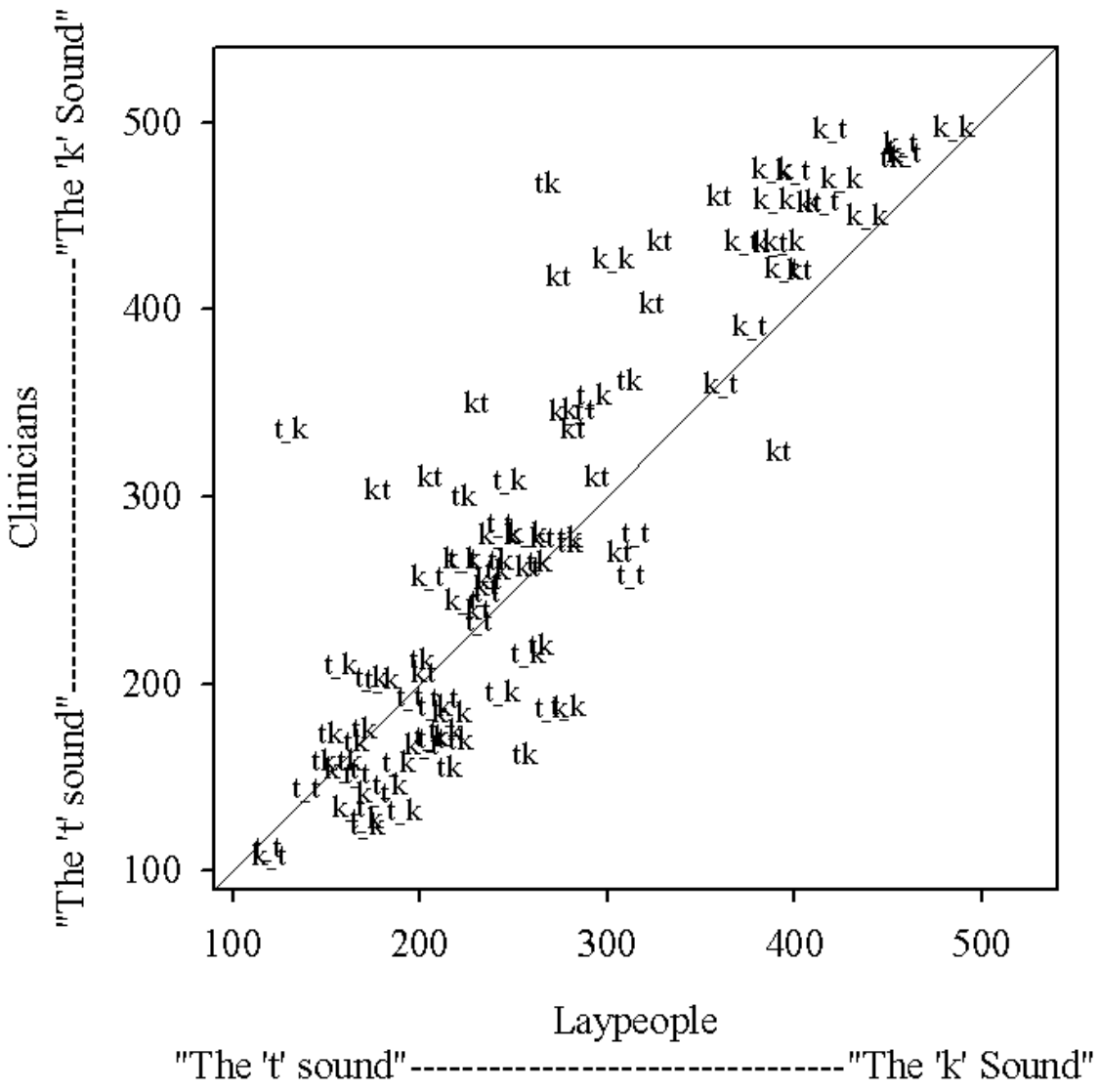

Figure 3b. /s/ and /θ/ Mean Laypersons' VAS rating to Mean Clinicians' Ratings

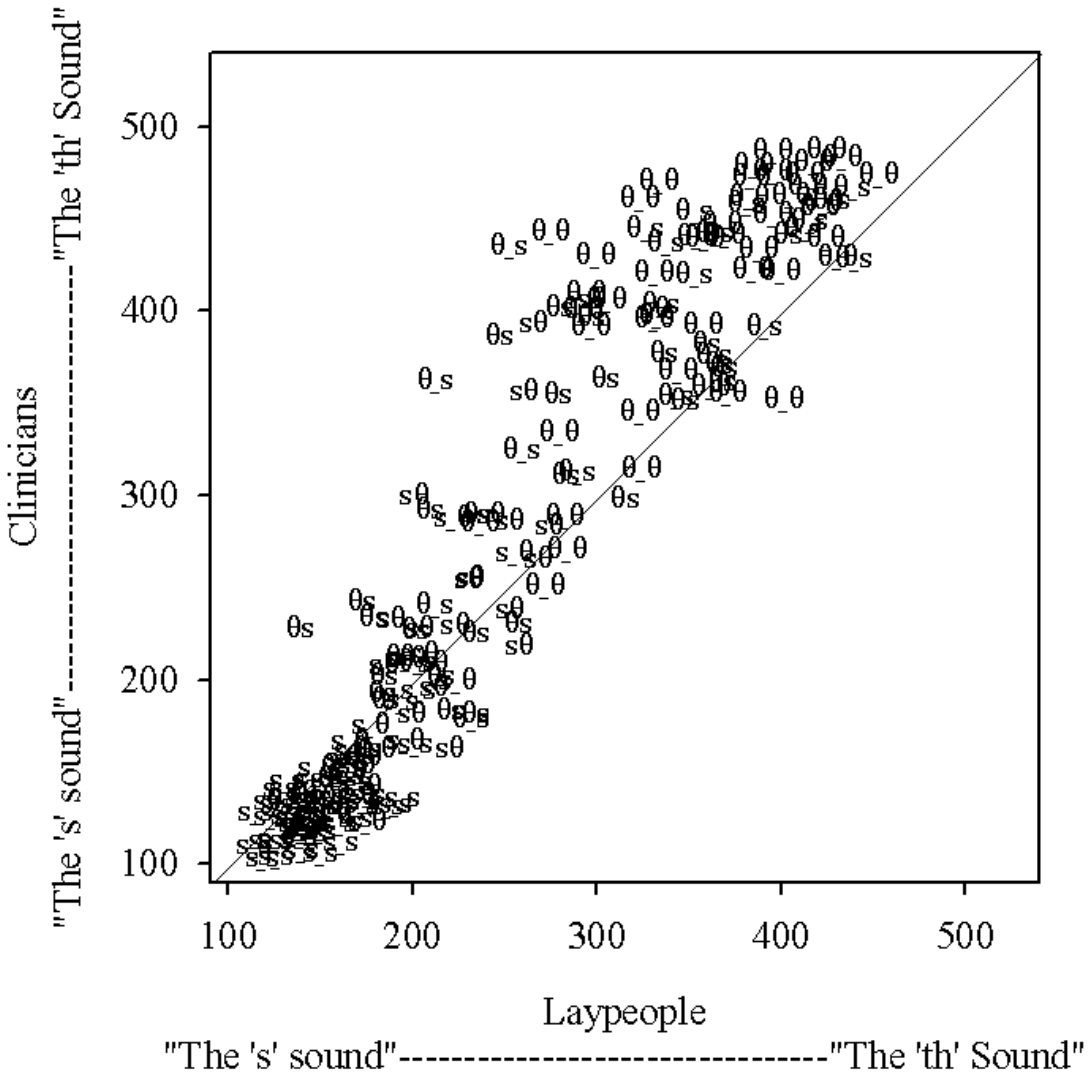

Figure 3c. /d/ and /g/ Mean Laypersons' VAS rating to Mean Clinicians' Ratings

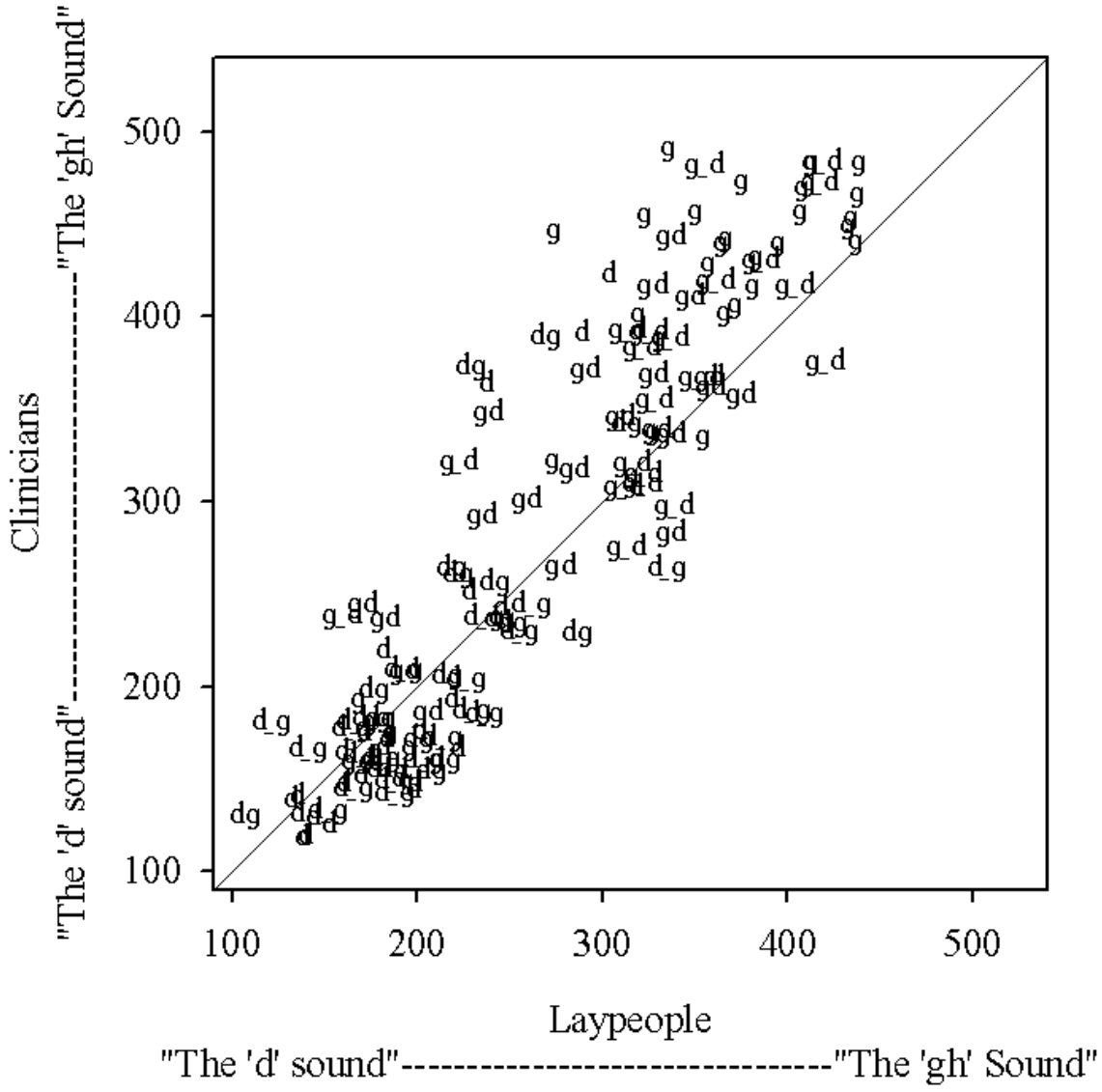

Figure 4a. /t/ and /k/ Reliability Measurements

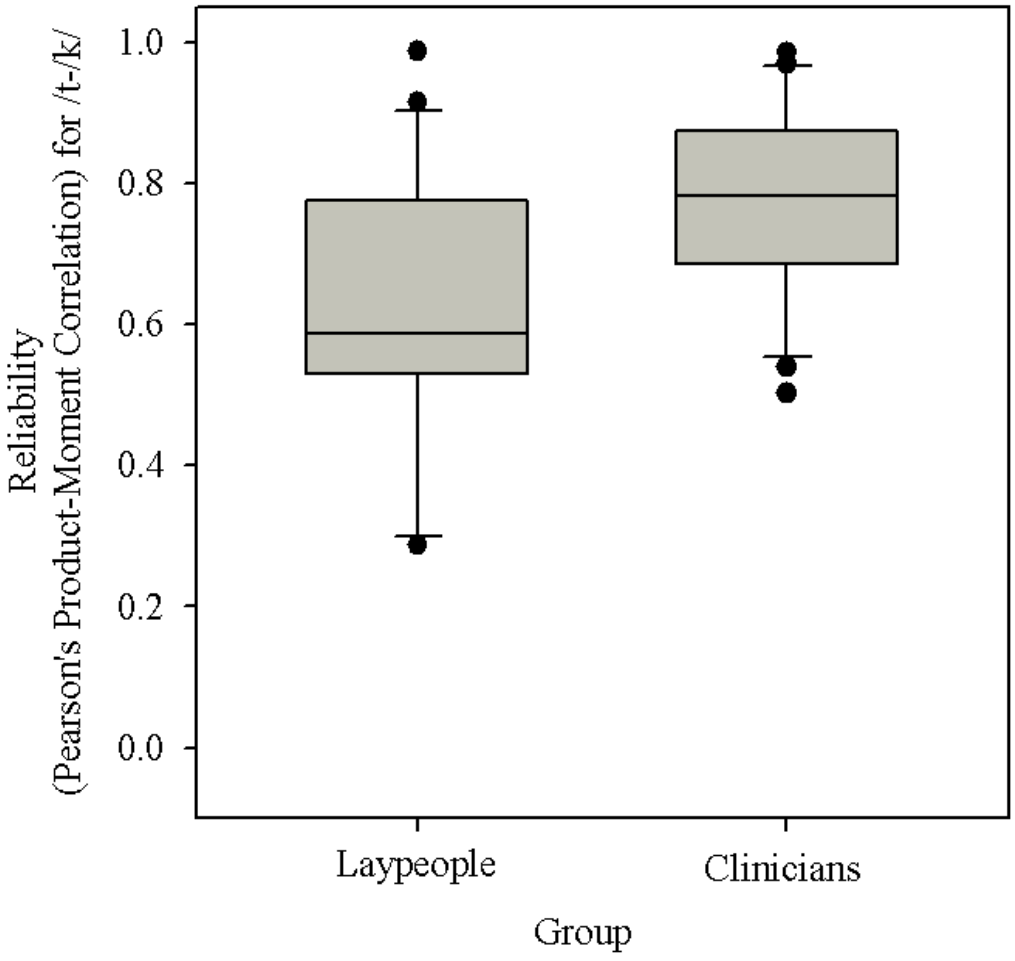

Figure 4b. /s/ and /θ/ Reliability Measurements

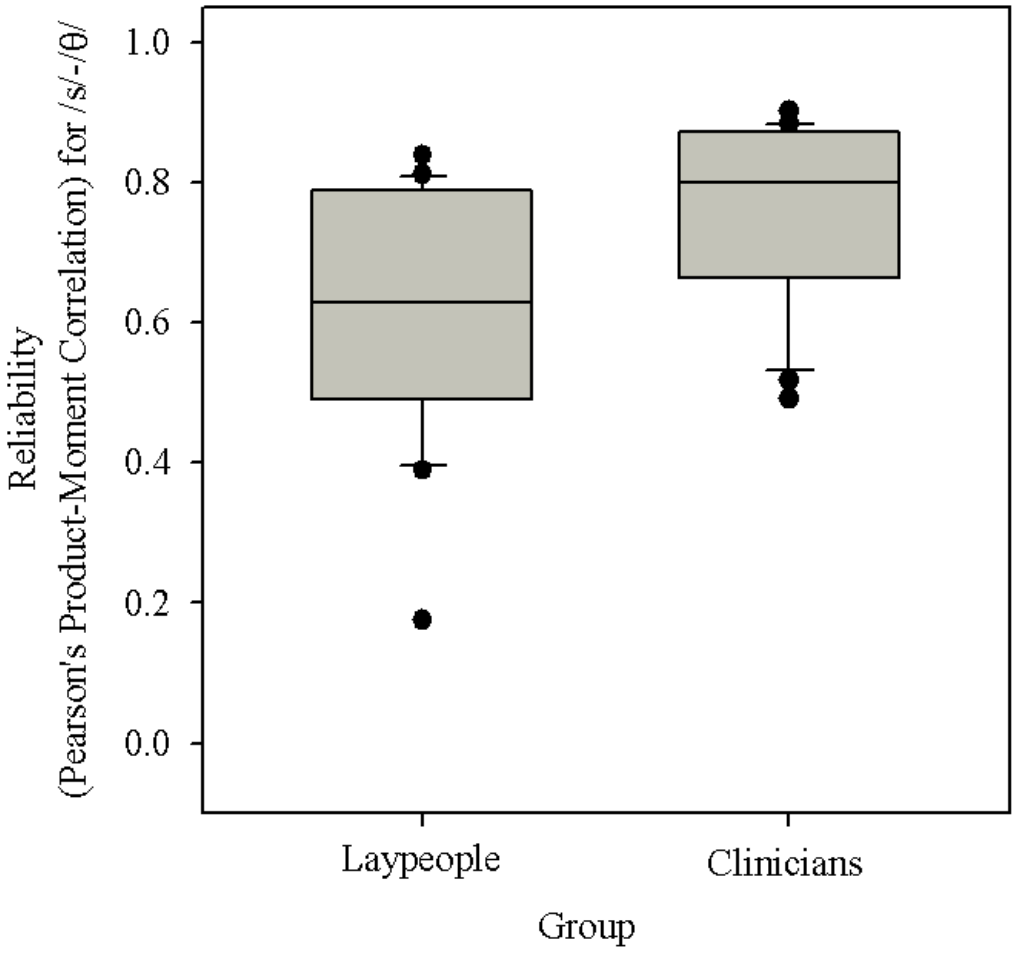

Figure 4c. /d/ and /g/ Reliability Measurements

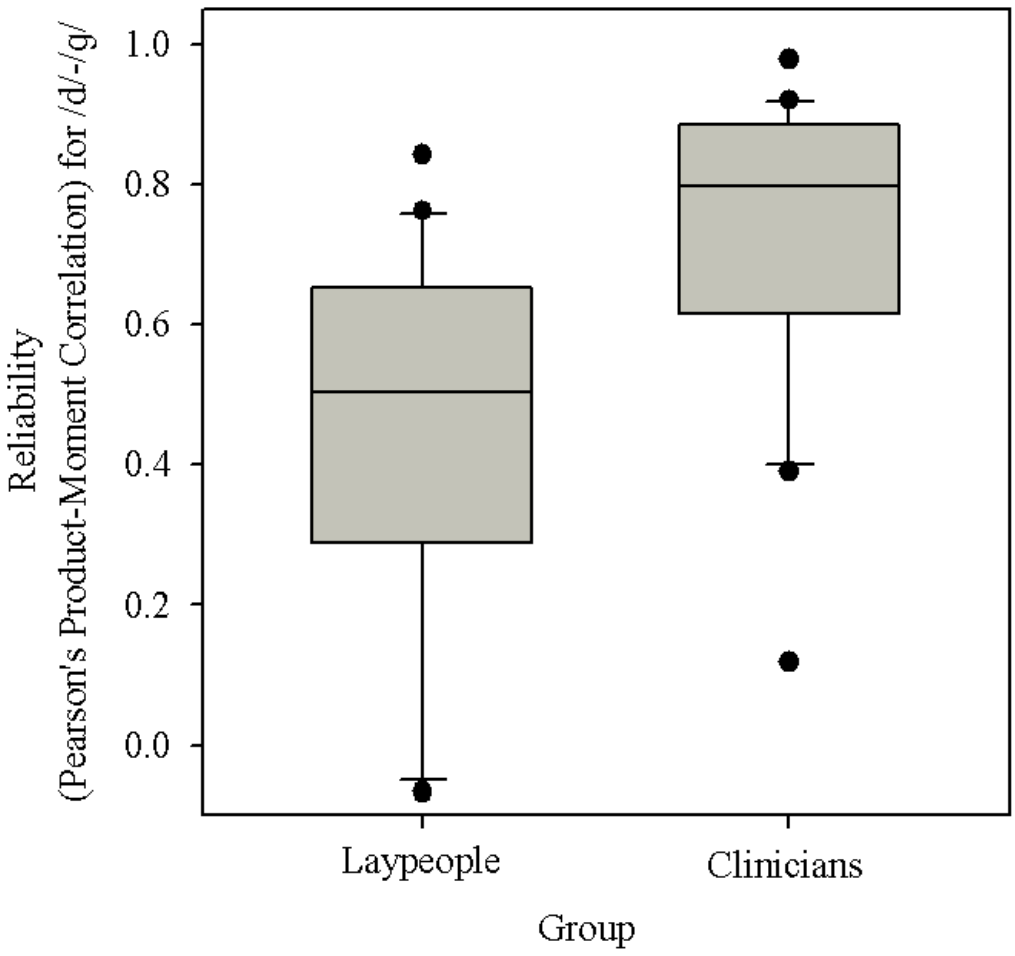

Figure 5a. Naïve listeners (maroon) and clinicians' (gold) /t/-/k/ ratings in front vowel contexts by total loudness (left), Peak ERB (middle) and compactness index (right)

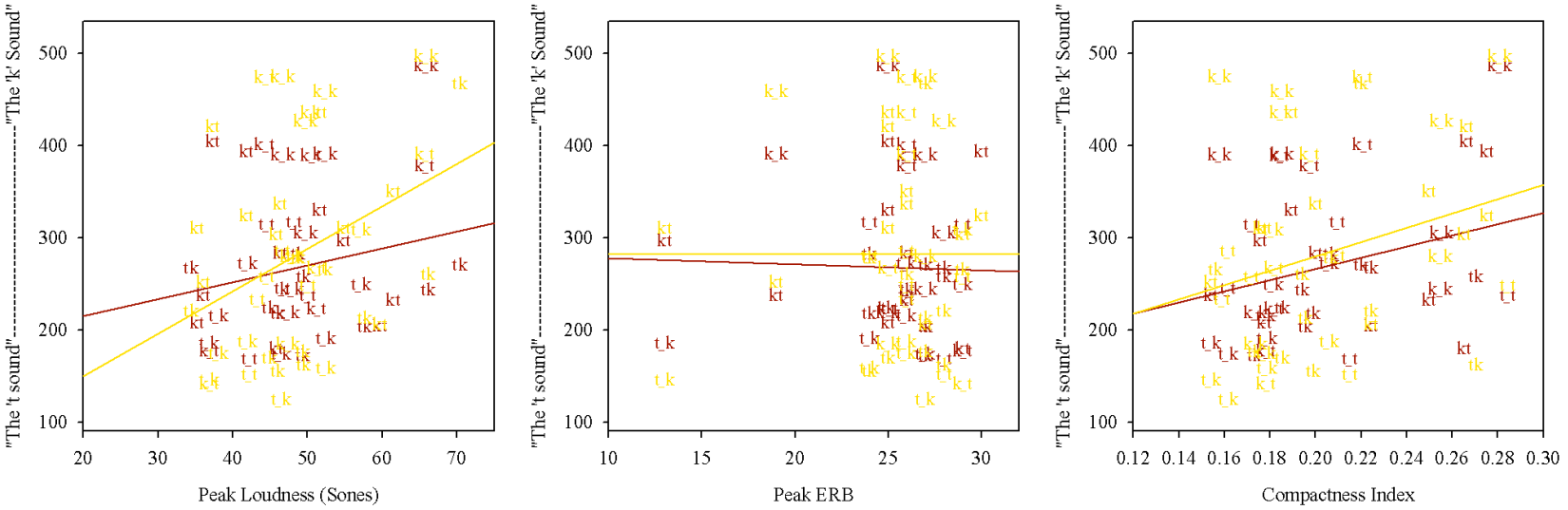

Figure 5b. Naïve listeners (maroon) and clinicians' (gold) /t/-/k/ ratings in back vowel contexts by total loudness (left), Peak ERB (middle) and compactness index (right)

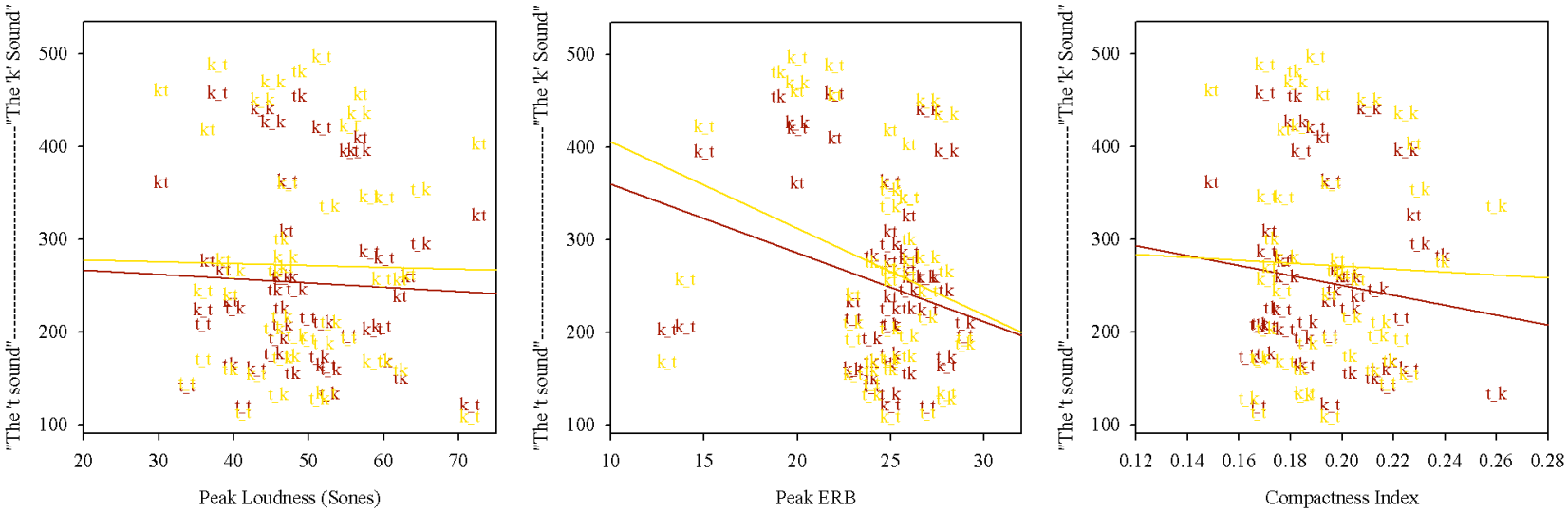

Figure 5c. Naïve listeners (maroon) and clinicians' (gold) /d/-/g/ ratings in front vowel contexts by total loudness (left), Peak ERB (middle) and compactness index (right)

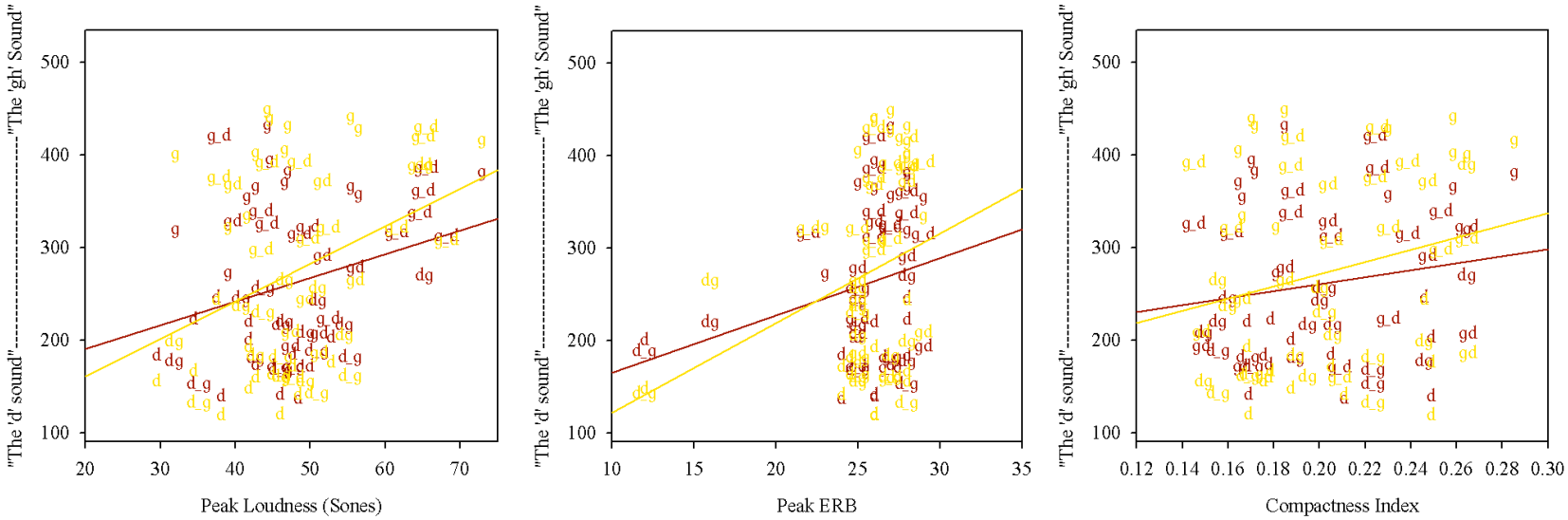

Figure 5d. Naïve listeners (maroon) and clinicians' (gold) /d/-/g/ ratings in back vowel contexts by total loudness (left), Peak ERB (middle) and compactness index (right)

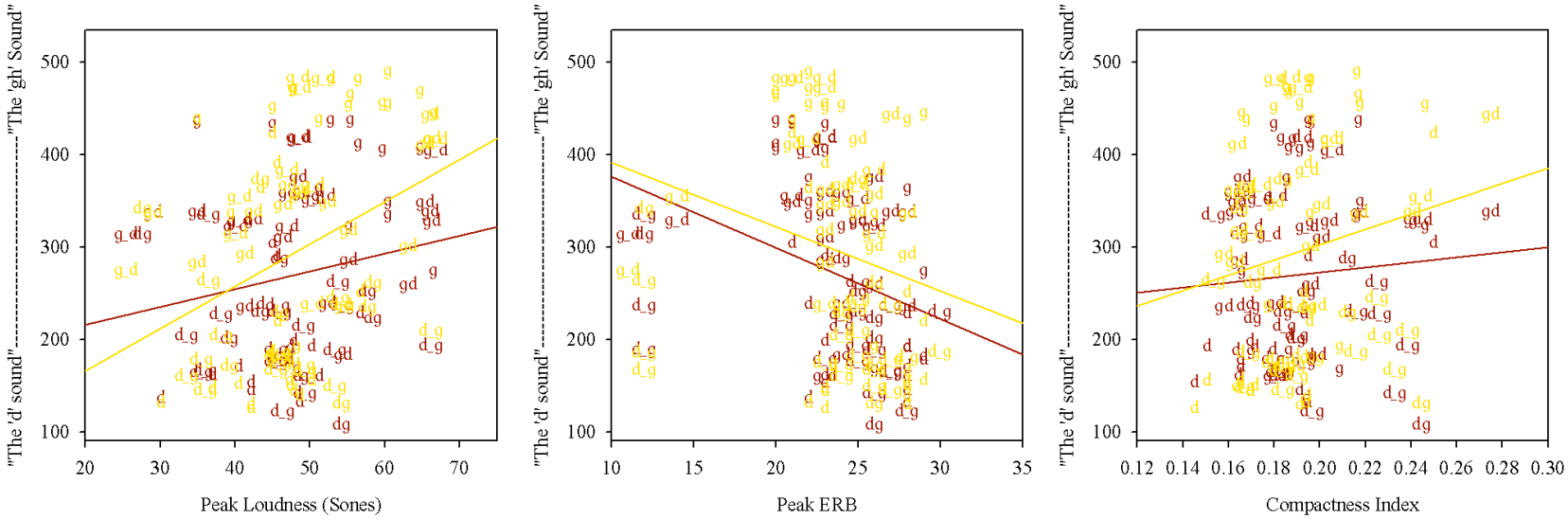

Figure 5e Naïve listeners (maroon) and clinicians' (gold) /s/-/θ/ ratings by total loudness (left), Peak ERB (middle) and compactness index (right)