Listener Strategies in the Perception of Dysarthric Speech

A thesis submitted in partial fulfilment of the requirements for the Degree of Master of Speech Language Therapy, Department of Communication Disorders, University of Canterbury

by Sharon Broadmore

April 2011

Table of Contents

List of Tables

List of Figures

Acknowledgments

Firstly I would like to thank my primary supervisor, Dr Megan McAuliffe for her patient support in the development of this thesis. Dr McAuliffe has been generous in giving of her time, expertise and advice so that this work could be completed. Secondly I wish to extend my thanks to Stephanie Borrie who gave me encouragement, mentoring and advice throughout this process. Her skills in research design and writing have been of great benefit to my understanding of the subject and completion of the paper. I wish to thank Dr Gregory O'Beirne for his guidance with the research method, his patience in helping me understand the preparation of acoustic signals for presentation to participants and his technical expertise. Lastly I would like to thank my partner who has been unyielding in her support for me during my studies.

Abstract

When listeners are presented with stimuli from multiple speakers versus single speakers in a perception experiment, decreased speech recognition accuracy and increased response time results. These findings have been demonstrated in studies that have employed normal (Creelman, 1957; Mullennix & Pisoni, 1990; Nygaard & Pisoni, 1998) and accented speech (Bradlow & Bent, 2008). It is thought that perceptual normalisation processes are, in part, responsible for this perceptual cost (Bladon, Henton, & Pickering, 1984; Johnson, 2009; Magnuson & Nusbaum, 2007; Mullennix, Pisoni, & Martin, 1989). Interestingly, studies are yet to examine whether these same findings occur when listeners encounter dysarthric speech – a naturally degraded speech signal associated with neurological disorder or disease. It has also been found that when listeners are exposed to multiple speakers with dysarthria, they generally adapt to the dysarthric signal over time; resulting in an improved ability to decipher the signal (Liss, Spitzer, Caviness, & Adler, 2002; Tjaden & Liss, 1995a). However, the rate of this adaption when listeners are exposed to a single speaker is yet to be examined.

This study aimed to determine: (1) whether the intelligibility of dysarthric speech (in this case, hypokinetic dysarthria associated with Parkinson's disease) varied across single versus multi-speaker conditions; and (2) whether intelligibility increased over time when a listener was exposed to a single speaker with dysarthria. To answer these questions, sixty young healthy listeners were randomly allocated to one of four experimental conditions, one multiple speaker and three single speaker conditions. Each listener transcribed 60 three to five word phrases over one session and the results were examined for percent words correct. Contrary to expectations, there was no significant difference in percent intelligibility scores of the listener group who transcribed in a multi-speaker versus transcriptions from the single speaker listener conditions. In addition, perceptual learning effects across the rating period were identified for two out of the three single speaker listener groups only. The absence of significant findings in the multi-speaker versus single speaker transcripts may be explained by further analysis of within speaker variability. Acoustic analysis of the speakers may also shed light on the reduced perceptual learning that occurred in one of the single speaker groups. Greater numbers of speakers and experimental phrases would be beneficial in expanding trends seen in intelligibility of the single speaker groups.

Introduction

Dysarthria and Speech Intelligibility

Dysarthria refers to "a collective name for a group of speech disorders resulting from disturbances in muscular control over the speech mechanism due to damage of the central or peripheral nervous system. It designates problems in oral communication due to paralysis, weakness, or incoordination of the speech musculature." (Darley, Aronson, & Brown, 1969b, p. 246). Dysarthria commonly occurs as a result of traumatic brain injury, trauma at birth, stroke or progressive neurological disease such as Parkinson's disease (PD) (Miller, Noble, Jones & Burn, 2006), or progressive supranuclear palsy (PSP) (Schrag, et al., 2003). Dysarthria can result from lesions in both the central and peripheral nervous system pathways (Murdoch, 1998).

Dysarthria is traditionally characterised using the perceptual classification system of Darley, Aronson and Brown (1969a) which divides the speech disorder into seven subtypes: hyperkinetic, hypokinetic, ataxic, spastic, flaccid, unilateral upper motor neuron and mixed. Each of these subtypes has its own distinct speech features and neurological site of origin. Some of these perceptual speech characteristics include reduced intensity, imprecise articulation, monopitch, monoloudness, and excess and equal stress. One or more of these characteristics may lead to reduced speech intelligibility.

The reduced intelligibility exhibited by individuals with dysarthria can result in communication breakdown—with the listening partner unable to comprehend the speaker's utterances. This, in turn, has negative effects on conversations, interaction with others and the ability to make one's self understood. The presence of dysarthria has been reported to have significant negative effects on quality of life (Miller, et al., 2006) and results in reduced functioning in social, physical and workplace contexts (Hustad, Beukelman, & Yorkston, 1998; Mackenzie & Lowit, 2007; Yorkston, Beukelman, Strand, & Bell, 1999).

Speech intelligibility is defined as "the degree to which a listener understands the acoustic signal produced by a speaker" (Duffy, 2005, p. 96). It is dependent upon the interplay between speaker and listener and is influenced by the acoustic output of the speaker and the various cognitive- perceptual processes activated by the listener to decipher the signal. Measures of speech intelligibility are one of the primary outcome measures in the assessment and treatment of dysarthria and one of the key references for speech-oriented approaches to rehabilitation (Duffy, 2005).

Measurement of Speech Intelligibility in Dysarthria

Currently two main types of intelligibility measurement exist: intelligibility rating scales and orthographic-based measures (including single word and phrase or sentence tests). Intelligibility measures are a useful tool for quantifying both the outcomes of behavioural speech intervention and in tracking disease progression through the analysis of intelligibility decline over time.

Rating scales.

Speech-language therapists have used two types of rating scales to measure perceptual manifestations of voice and speech associated with dysarthria. Partition scaling and magnitude scaling have both been employed to measure perceptual attributes such as loudness, pitch, severity and intelligibility. Partition scaling, such as Equal Appearing Interval (EAI) and categorical scales, have been used in the areas of voice to quantify characteristics like vowel roughness (Toner & Emanuel, 1989), nasality (Whitehill, Lee, & Chun, 2002; Zraick & Liss, 2000) and pleasantness (Eadie & Doyle, 2005). Partition scaling involves a listener rating a particular speech characteristic on a linear continuum that is equally partitioned with either numerals or descriptors (Schiavetti, 1992). Partition scales have been used in dysarthria assessment and intervention to provide a measure of intelligibility as well as information pertaining to subtype classification. For differential diagnosis, speech-language therapists in clinical practice, and some researchers, frequently use variants of the original Darley, Aronson and Brown rating scales published in their seminal Mayo Clinic studies (Darley, et al., 1969a, 1969b). Using this form of scale, an independent listener analyses the speech signal for a number of given characteristics, for example loudness, articulation, intelligibility and/or prosody. These are rated on a one to four scale—one being normal and four being severely deviant (Duffy, 2005, p. 90).

Magnitude scales are another form of scaling commonly used in the field of speech disorders. One such scale is the Direct Magnitude Estimation scale (DME) (Stevens, 1975), wherein listeners assign a number to particular speech or voice attributes proportional to the magnitude (deviation from normal) of that attribute. DME is thought to provide a reliable measure of the intelligibility of the signal as a whole (Schiavetti, Metz, & Sitler, 1981; Stevens, 1975). The DME scale has been utilised to ascertain the degree of intelligibility impairment in hearing-impaired speakers (Schiavetti, et al., 1981), those who stutter (Schiavetti, Sacco, Metz, & Sitler, 1983), and in individuals with dysarthria (Turner, Tjaden, & Weismer, 1995; Turner & Weismer, 1993). Using DME, listeners are commonly instructed to rate "the ease with which speech could be understood, with attention to the precision of articulation" (Tjaden & Wilding, 2004, p. 771). DME scaling of intelligibility can be done in one of two ways—with the use of a referent (modulus scaling) (Weismer, Jeng, Laures, Kent, & Kent, 2001) or without (free modulus rating) (Schiavetti, 1992). When a modulus is used, it is commonly selected as a mid-category exemplar—that is, a speech sample that is about mid-range intelligibility. Listeners are then instructed to assign the modulus a number and scale each subsequent speech stimuli either above or below that modulus. Speech intelligibility can be affected by one, or a number of subsystem speech anomalies including speaking rate, fundamental frequency, and intensity. While DME provides a reliable measure of intelligibility, it does not provide an indication of how much of the signal is actually understood or comprehended by the listener, or the types of errors that are commonly made by the speaker — word substitutions or articulation errors. However, orthographic-based measures do provide percentage intelligibility scores and, therefore, are commonly used as a tool for both clinical and research purposes.

Orthographic measures.

Both in clinical practice and research (Dagenais, Watts, Turnage, & Kennedy, 1999; McAuliffe, Ward, & Murdoch, 2005, 2007) orthographic transcription is one of the more commonly used forms of assessment of intelligibility. It is used to rate severity of dysarthria as well as measuring outcomes. It provides a quantitative measure of correctly transcribed words. One commonly used assessment based on orthographic measures is the Assessment of Intelligibility of Dysarthric Speech (AIDS) (Yorkston, Beukelman, & Traynor, 1984). This assessment consists of single-word and sentence-level (increasing in length) transcriptions that are generally transcribed by a single naive listener. An assessor then compares listener transcriptions with target productions. Results are tallied and the final intelligibility score is calculated as a percentage. Numerous factors may influence intelligibility scores including: equal weighting in the scoring of accurately transcribed words (e.g., all words are scored for accuracy without consideration of linguistic class), listener familiarity with dysarthric speech, variation in the number of speakers or types of dysarthria in a rating task and the perceptual learning of the listener (Mattys & Liss, 2008; Mullennix, et al., 1989). These factors need to be considered for reliability of rating scales or orthographic intelligibility scores in research and clinical practice.

The current study is clinically motivated. It focuses on two of the factors that have been shown in the speech perception literature to influence perceptual processing—specifically, speaker variation and perceptual learning. To date, research in the field of dysarthria has not empirically tested how these two processes may affect clinical outcome measures. Therefore the subsequent components of this review will focus on the literature related to speaker variation and perceptual learning.

Speaker Variation

The human speech perception system shows remarkable flexibility in its ability to adapt to different voices, pronunciations of various phonemes, variations in speaking rate, and other speaker-dependent and independent variation (Peelle, 2005; Peelle & Wingfield, 2005). Individuals are able to comprehend the content of messages, even in the face of this substantial variation. Given the ability of listeners to adjust to these variations in production, theorists have postulated that a process of perceptual adjustment or “normalization” occurs (Bladon, et al., 1984; Johnson, 2009; Magnuson & Nusbaum, 2007) when there is variability in the speech signal. This may be variability within the speaker or the variability that occurs across multiple speakers. The process of normalisation has been shown to begin during the early part of a child’s development. As the child is learning information about their native language from the language that surrounds them, it is necessary for them to able to identify the same word spoken by more than one person. In order to do this the process of adjustment to different voices is necessary (Jusczyl, Pisoni, & Mullennix, 1992; Kuhl & Miller, 1982). Normalisation studies have provided evidence that although a listener’s perceptual system develops the capacity to accommodate to variation in the speech signal, the accommodation processes that are required come at a cost. The cost results in increased latency in perceptual processing times and reduced accuracy when the system is challenged (Mullennix, et al., 1989).

Creelman (1957) was one of the first to examine the effects of speaker variability on spoken word recognition. His interest was prompted by an earlier investigation by Davis, Bidduph and Balashek (1952) who were designing a mechanism for recognition of spoken telephone digits. Davis et al. found that the circuit they had created for this task was able to adjust accurately for a single speaker over ten different spoken digits, however, in order for it to recognise the same digits spoken by other speakers the circuit had to be adjusted to each voice separately. Creelman investigated whether individuals performed in a similar manner as the circuitry when perceiving more than one voice—that is, whether the human perceptual system adjusted between speakers. Monosyllabic and phonetically balanced words were spoken by 25 males and 25 females and arranged to provide five stimuli tapes, each of 25 words. Tape one had one speaker, tape two had two speakers, tape three had four speakers, tape four had eight speakers and tape five had 60 speakers. In all there were 16 tapes in each set. In a between-subjects design, words were randomly presented to five listeners assigned to one of the five experimental stimuli sets. The words were played randomly through headphones at an interval of 3.5 seconds between words. The words were mixed with random noise at -6, 0 and +6 dB signal to noise ratios (SNR) and listeners were instructed to make a response every time a word was spoken, guessing if required. Findings of the research indicated that as the number of speakers in the stimuli set increased, intelligibility scores decreased significantly. The effect was greater when listeners were taxed by noise—particularly in the 0 and +6 dB SNR conditions and, interestingly, female voices were more intelligible than male voices. Creelman theorised that this effect was the result of a process of 'perceptual adjustment' whereby listeners must have to adapt to differences in voices, which in turn negatively affected intelligibility scores.

In addition to the measurement of speech recognition accuracy, research has also investigated response times (RT) as a measure of the normalisation process when more than one speaker is heard. RT's have been shown to provide a strong indication that processes occur in the multi-speaker groups that do not occur, or occur differently, in single speaker groups (Cole, Coltheart, & Allard, 1974). Cole et al. investigated aspects of memory as well as processes of speech perception using multi-speaker and single-speaker stimuli. In their research response times were found to be faster when a phoneme was produced by the same speaker than when spoken by a different speaker in a naming task of consonants or vowels. The basis of Cole et al.'s study was the concept of two different types of memory storage for visual stimuli (the physical code, which stores letters or phonemes based on their physical characteristics, and the name code, which encodes the verbal characteristics into memory) postulated by Posner and colleagues (Posner, Boies, Eichelman, & Taylor, 1969; Posner & Keele, 1967). Cole et al. examined if similar memory storage capacities were utilised for auditory stimuli. They used a naming task of phonemes spoken by a single speaker (either male or female), or multiple speaker (both male and female). The researchers sought to examine the effects of these two conditions on the recall of phonemes. They postulated that if response times were faster when the two phonemes were spoken by the same voice then auditory memory did not utilise these two different memory stores. Results showed that the responses to the same or different letter judgements were faster in the single speaker (same voice) condition and interestingly were also faster when naming same rather than different. They suggest from these results that, for auditory stimuli, a process occurs whereby when the presentation of the letters in the same voice occurs only the physical code is accessed to identify if the letters are the same. However, if the letters are presented in different voices analyses of the name code is also necessary. This additional process in a mismatch of voices would result in a slower processing time therefore longer RT's.

In another speaker variability study using measures of response time Mullennix, Pisoni and Martin (1989) examined the effect of multiple speakers on spoken word recognition. They sought to determine if listeners' word recognition accuracy and response latency times differed across single speaker versus multi-speaker samples and if there was an effect of perceptual accuracy in high versus low density words. Over four experiments using word recognition and response times as dependent variables they clearly demonstrated that the multi-speaker group was disadvantaged due to the occurrence of a perceptual adjustment process. The researchers manipulated variables of word frequency, lexical density and a distorted speech signal. Two different speech distortion methods were used: 1) signal to noise ratio (SNR) where the speech signal is superimposed with white noise and 2) manipulation of the acoustic signal with noise. Throughout the experiment listeners heard consonant-vowel-consonant (CVC) words spoken by 15 different speakers in the various conditions.

In the first phase Mullenix et al., (1989) found that listeners in the multi-speaker condition exhibited significantly reduced intelligibility scores compared to listeners in the single-speaker condition across both noise (SNR) and lexical density. In phase two, both word naming accuracy and response times were negatively affected in multi- speaker conditions even without noise. Further, analysis of response latencies found that they decreased over the trials - evidence of a practice or perceptual learning effect. Phase three again demonstrated faster response times in the single speaker condition and, further, that high frequency words were accurately identified more often and faster than low frequency words. Finally, phase four again showed a significant effect on speaker variability in favour of the single speaker, word frequency in favour of the high frequency words and an effect of reduced performance as the signal became more degraded. Of interest, specifically to the current study, there was a significant interaction between speaker variability and signal degradation specifically at a degradation level from10 to 20%. This is relevant for the current study as the level of noise degradation may represent intelligibility levels in a naturally degraded speech system. Taken together, results of the study show a consistent and robust effect of speaker variability. Speaker variation has a detrimental effect on spoken word recognition and response times, specifically when the signal is degraded by noise. The findings support the idea that some form of perceptual normalisation or adjustment had occurred across trials resulting in decreased accuracy and longer response time. In addition, analysis of results from phase two showed a clear perceptual learning effect.

The notion that a normalisation process does occur in speech perception and that the process incurs a cost to the perceptual system has been established in the literature (Creelman, 1957; Mullennix, et al., 1989). The question of where in the perception process adjustment to variability occurs was investigated by Mullennix and Pisoni (1990). They examined the relationship between the processing of the acoustic-phonetic information of the spoken word and that of the encoding of the vocal characteristics when more than one speaker is heard. The aim of the research was to determine whether the two processes, adjustment to different voices (speaker normalisation) and acoustic-phonetic encoding are interrelated or whether they occur separately. If they are separate processes then it was necessary to determine which one would take precedence. Response times were used to determine at what stage the two processes occur, and whether this was affected by variable speakers. Voice identification (male or female) and phonetic identification (initial /p/ or /b/) in monosyllabic English words were manipulated in a selective attention procedure. Four different exposure conditions were employed with listeners asked to attend to the voice and identify whether it was male or female, or attend to the word and identify whether it had an initial /p/ and /b/. Results of the investigation showed that the two processes were closely related and mutually dependent. Further, the study identified that the perceptual encoding of voice information, or adjustment, was mandatory and that variable voices have a negative effect on phonetic encoding resulting in increased response time. Results showed that when words were presented in different voices (male and female) listeners had greater difficulty in making decisions about the initial phoneme of the word than when asked to make a judgement about the voice when the initial phoneme was varied (initial /p/ or /b/). The authors theorised that the two perceptual processes, encoding of the voice and encoding of the phonetic information, function in parallel. However, when greater numbers of perceptual processes are required (e.g., variation in voices and variation in initial phoneme) the processes occur in a hierarchical manner in which the normalisation process takes precedence.

The above studies have provided evidence that the perceptual system has the ability to adapt to different voices, albeit with a cost to the perceptual system. They also provide evidence to show that adjustment to the voice is an initial and mandatory process. The studies were completed using neurologically healthy speakers that spoke the same language as the listeners. However, in everyday life an individual may encounter a number of different accents of their native speech or individuals with speech disorders. Studies that have incorporated noise to degrade the speech signal have provided some insight into how the perceptual system copes with the challenges of a distorted speech signal (Creelman, 1957; Mullennix & Pisoni, 1990). Other studies on perceptual adjustment or adaptation using a speech signal that is an accented version of the listener's native language contribute to the understanding of how the perceptual system adapts under such challenges.

Top-down and Bottom-up Perceptual Processes

Prior to introducing literature that investigates listener adaptation to accented speech, it is helpful to discuss the concept of top-down and bottom-up processing as the literature refers to this concept. This is especially important as we move into research that uses connected speech (i.e., sentence or phrases) as opposed to phonemes or single words. Recognition of sentence form provides the listener with suprasegmental and contextual information that may be used in the perceptual process. Top-down processing refers to the perceptual system's ability to utilise elements such as context, environment, knowledge about the speaker, and semantics—all of which are frequently found in an everyday conversation. On the other hand, bottom-up processing utilises the phonemic structure of the word in both word and sentence recognition.

Recent models of speech perception indicate an agreement between researchers that spoken word and sentence recognition requires a hierarchical system of processing. Speech perception models such as Cohort (Marslen-Wilson & Welsh, 1978), Shortlist (McClelland & Elman, 1986), and the Neighbourhood Activation Model (NAM) (Luce, 1987), support the concept that speech perception has an initial bottom-up process. The listener's perceptual system analyses the phonological segments as the word unfolds – this activates the words in the lexicon that closely resemble the incoming signal (Samuel, 1981). Higher level information such a context and semantics influence the selection process by inhibiting those words in the cohort that are less likely to be the target word. Top-down perception processing, where sentential context has an influence over word recognition in a sentence, has been evidenced by a number of researchers. Paradigms such as shadowing, mispronunciation (Marslen-Wilson & Welsh, 1978) and gating (Grosjean, 1980) have been utilised to provide such evidence. When examining how individuals use speech perception strategies as outlined in speech perception models, it raises the question of how a listener uses bottom-up processing to perceive the impaired speech signal when acoustic properties of the signal are impaired and perhaps unrecognisable.

Nygaard and Pisoni (1998) investigated whether: (1) listeners could identify a talker from hearing only single words in a normal and a degraded speech signal and (2) learning of sentence length utterances has an influence on single word retrieval from familiar and unfamiliar speakers and (3) learning at sentence level influenced the accuracy of single word retrieval and sentence length utterances of distorted speech of familiar and unfamiliar speakers. They used implicit learning (exposure only to the voices) and explicit training to examine perceptual learning processes. Explicit training involved three phases: (1) the *familiarisation task* where listeners heard single words from named speakers (2) the *recognition task* where the listeners heard single words spoken by the same speakers and were asked to identify the speaker while receiving feedback to inform them if they were correct of not and (3) the *test phase* where listeners heard single words from the same speakers and were again asked to identify the speaker, however in this phase they were not provided feedback. They used this training paradigm in single words in one experiment and sentence length utterances in the other. The results showed a large variation in the listener's perceptual abilities. However overall, the results showed that familiarity with a talker's voice through single word training influenced linguistic processing and that listeners were able to identify single words produced by the familiar speakers in a degraded speech signal (tested at +5, 0 and -5 dB SNR). Results also showed that listeners needed fewer training sessions when trained in sentence length utterances to identify speakers correctly and had higher intelligibility scores at sentence length utterances of familiar speakers. Familiarity with the speaker also resulted in higher intelligibility levels at lower SNR levels.

Overall, the findings of Nygaard and Pisoni provided evidence that learning the characteristics of voices assisted with intelligibility. Indexical characteristics and linguistic properties of speech appear to be intertwined or linked in the process of speech perception. When the speech signal is degraded or the environment has a deleterious effect on listening the ability of listeners to use top-down information to aid intelligibility may be limited. In that case familiarity with the speaker's voice may be the most helpful in successful communication. This study has brought some understanding of the effects of perceptual learning on sentence level utterances and how a listener can quickly learn to perceive the speech of an unfamiliar speaker. Other studies have provided evidence of rapid perceptual learning using accented speech.

Listener Adaptation to Accented Speech

When the speech signal is challenging to the listener, such as in the case of accented speech, studies can provide insight into the bottom-up and top-down processing strategies enacted to perceive speech. Clarke and Garrett (2004) identified the speed at which perceptual learning took place in accented speech. In a series of experiments, native English speakers listened to non-native English speakers (Spanish and Chinese). Response times were measured when listeners were exposed to single probe words from the prior sentence length utterance. It was found that the native English listeners started to adapt to the accented speech rapidly (within two to four sentence level phrases). Listeners were exposed to low probability sentences from a female American-English speaker and a female native speaker of Mexican-Spanish. Fifteen of the listeners reported having had previous experience with accented speech, for example contact with close friends or family with a Spanish accent. To counter this, the listeners with high experience and those with low experience were investigated as separate groups. They found that initially those with experience had shorter response times than those with low experience (although not statistically significant) however, by the presentation of the fourth (final) block this difference had attenuated. Overall, initially the listeners were slower at responding to the accented speech than to the native English speakers but this effect reduced after only one minute of hearing the accented speech, representing exposure to only two to four sentences. The investigators tested the same listeners with Chinese accented speech as the listeners were less likely to have been exposed to Chinese accented speakers due to where they lived. The investigators conducted this experiment for two reasons: (1) to control for the possibility that the listeners may have had unknown prior adaptation to the Mexican-Spanish accent through ambient exposure to the population during their lives, and (2) to investigate whether the listeners were accessing a stored template or whether their adaptation was the result of the "online learning of phonological patterns" (Clarke & Garrett, 2004, p. 3653). Similar results were found with reaction times for the Chinese accented speaker. From the presentation of the first block of stimuli to the fourth block, reaction times had reduced by just over 130ms and significant improvements occurred over the first few sentences in the first block. Their results showed that the listeners adapted to the accented speech they heard, and they theorise that adaptation to one voice may generalise to a new voice.

The concept of generalisation of previously heard accented speech to novel voices with the same accent was investigated by Bradlow and Bent (2008). They aimed to determine the extent to which access to higher lexical information influenced adaptation to foreign accented speech. The first component of the study examined talker-dependent adaptation – using a single speaker transcription paradigm. American-English speakers heard a speaker of Chinese with low, medium or high English intelligibility and one Slovakian speaker with medium English intelligibility. Listeners were assigned to one of the four groups and were presented with 64 Bamford-Kowal-Bench (BKB) sentences (Bamford & Wilson, 1979) read by one of the above speakers in the assigned groups. Perceptual adaptation was measured using intelligibility scores at the first, second, third, and fourth quartile. The results revealed that, regardless of the intelligibility levels of the speakers, there was an increase in the overall intelligibility in all groups over time. However, as the intelligibility of the speaker decreased, longer periods of exposure were required to for adaptation to occur. Part two of the study examined speaker-independent perceptual learning. Listeners heard the stimuli sentences mixed with white noise in five conditions: multi-speaker, which incorporated all five of the of the Chinese accented speakers speech (16 sentences each); a speakers speech specific (the same training a speakers speech as the post-test a speakers speech); four single speaker groups (four Chinese accented speakers of differing intelligibility scores); two control conditions, training with native English speakers and no training. The results of part two showed that listeners exposed to the multi-speaker condition could generalise their learning to other speakers of the same accented speech. The researchers suggest from these findings that "exposure to multiple talkers of a foreign-accent may be a highly effective means of enhancing speech communication between native and non-native speakers" (Bradlow & Bent, 2008, p. 721). It was further suggested that the greater the source variation the listener is exposed to, the possibility of robust adjustments and generalisation is likely to be greater.

Listener Adaptation, Perceptual Learning and Generalisation in Distorted Speech

Several forms of speech distortion methods have been used to examine the effects of adaptation or perceptual learning, including noise –vocoding (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005), time compression (Dupoux & Green, 1997), synthetic (Greenspan, Nusbaum, & Pisoni, 1988), and spectral shifting (Rosen, Faulkner, & Wilkinson, 1999).

Davis et al., (2005) aimed to determine whether perceptual learning occurred at segmental level or if higher lexical processes were involved in the process of adaptation. The authors demonstrated that listeners completing an orthographic transcription task adapted to noise-vocoded speech over 30 sentences, and that this adaptation to vocoded sentences was more efficient when listeners heard the non-vocoded sentences first. This demonstration of the "pop-out" phenomenon (whereby listeners exposed to a clear version of a distorted sentence suddenly find the distorted version more intelligible the second time they hear it) also occurred when additional written feedback was provided, leading authors to suggest the effect may not be reliant on an acoustic signal, but rather that the processes that support this effect may be at the phonological or an even higher level. Further experiments by this group showed that training on vocoded non-words that conformed to English phonology patterns produced no benefit, whereas training on vocoded real words did, leading the researchers to conclude that higher level support such as semantic, lexical and/or syntactic information aids listeners in intelligibility of noise-vocoded speech. To investigate which one of these higher level components aids intelligibility, noise vocoded sentences were presented to listeners following a period of training on prose sentences to determine if the semantic information was utilised, or so-called "jabberwocky" (or nonsense) sentences to determine if syntactic structure made a difference. Training with the prose sentences and the English sentences had similar results. Training with jabberwocky sentences was less effective than real words in improving performance. The researchers conclude from these results that neither semantic nor syntactic speech components are necessary for perceptual adaptation to occur, rather the lexical components guide the learning. Findings from this complex study showed that substantial perceptual learning of a distorted speech signal occurs rapidly, over as little as 30 to 40 sentences. The participants were correctly transcribing words they had not heard before in vocoded form indicating generalisation and the perceptual system had not learned whole words in the training, rather had learned parts of words which they were able to apply to novel words. It was proposed that perceptual learning occurs at a pre-lexical level and that the learning at this level is reliant on lexical level information to inform and make alterations to the pre-lexical level.

Adaptation and Perceptual Learning in Dysarthric Speech

How the listener adjusts to, and becomes better able to understand speech that deviates from normal is a question of theoretical and clinical interest to speech-language therapists. Within a perceptual rating period, either in the research or in the clinical arena, listeners are likely to adapt to the disordered signal. That is, their ability to comprehend the disordered speaker will become better with time. As yet, this has received limited attention in the dysarthria literature. Speech perception research examining this process of adjusting to distorted signals (Davis, et al., 2005) has shown that listeners rapidly adapt to a distorted or accented signal, while subsequent attempts to decipher that signal are more successful (Bradlow & Bent, 2008; Clarke, 2002; Hervais-Adelman, Davis, Johnsrude, Taylor, & Carlyon, 2010).

The research discussed thus far has utilised normal speech, artificially distorted and accented speech, and both single and multi-speaker comparisons. However, a growing number of researchers are contributing to the understanding of speech perception and perceptual learning using disordered speech. Research on speech perception of disordered speech helps to gain an understanding of the processes involved in adaptation and perceptual learning when the speech signal is organically distorted.

Hypokinetic dysarthria is typically the result of Parkinson's disease (PD) and exhibits perceptual characteristics such as fast rate of speech, monoloudness, monopitch and imprecise articulation (Darley, et al., 1969a; Duffy, 2005). When examining how individuals use speech perception strategies as outlined in speech perception models, it raises the question of how a listener perceives the disordered speech signal. Speech perception models have described bottom-up processing where the segmental features of words, such as phonemes and syllables in continuous speech, provide a strong link to the target word in the lexicon. In situations where segmental information is insufficient, top-down processes (i.e., context) may assist speech perception.

Top-down processing, where sentential context has an influence of word recognition has been evidenced by a number of researchers (Grosjean, 1980; Marslen-Wilson & Welsh, 1978). However when the speech signal is disordered, such as the case of hypokinetic dysarthria, intelligibility is compromised (Duffy, 2005; Liss, 2007). Reduced intelligibility has an undermining effect on perceptual learning rates (Bradlow and Bent (2008)). Liss posits that the perceptual processing of disordered speech may be the same as that for healthy speech, however, the perceptual challenges listeners encounter with disordered speech limits success (Liss, 2007).

A recent study by Mattys and Liss (2008) compared the voice specificity effects of speakers with normal, mild and severe dysarthria. Voice specificity effects are those effects described earlier in Nygaard and Pisoni's study—that words are better recalled if they are heard in the same voice than when they are heard in a different voice. Mattys and Liss also tested the Luce, McLennan and Charles-Luce's (2003) *time-course hypothesis* (as cited in Mattys & Liss, 2008) when applied to disordered speech, whereby indexical details are relied on more by a listener in less than optimal listening conditions. A further aim of their study was to show that using naturally degraded speech, such as dysarthria, alongside normal speech can be used to explore alternative areas of speech perception without the need for artificial intervention of the speech signal. Exploration of this type is fundamentally important to the clinical population and professionals that provide intervention for individuals and their communication partners. Mattys and Liss presented listeners with two blocks of 60 words, 30 spoken by males and 30 by females. Three groups of two speakers, one male and one female produced the speech stimuli: two normal speaking controls, two speakers with mild hyperkinetic dysarthria and two speakers with severe hyperkinetic dysarthria. The words were organised in the two blocks in order that half of the words in block one were repeated in block two, however half of the repeated words were in the same voice as block one and half were in a different voice. The listeners were to identify, via a yes/no response, if the words they heard in block two were the same as those from block one. Results demonstrated two categorical effects relevant to the current study: (1) voice difference had an effect on both accuracy and response latency and (2) as the voice impairment increased the percent recall error increased. Interestingly, recall of words that were presented in the same voice remained relatively high across the control, mild and severe conditions. Mattys and Liss infer from this that if the voice is constant the listener is able to tolerate a greater degree of the degraded speech signal. An effect of latency was found in the same and different voice conditions whereby as the impairment increased the latency periods also increased across the conditions. However, the latency period for the different voice condition was higher than the same voice condition in both the mild and severe dysarthric speech. In relation to *the time-course hypothesis* the authors divided the listeners into fast and slow respondents and analysed the results according to the two impairment groups. They found the fast respondents showed no voice effect or effect of interaction between the voice and impairment level, however the slow respondents exhibited both a voice effect and a voice-impairment interaction effect. They posit that this result confirms the time course hypothesis whereby when the listening conditions were good (mild dysarthria) and the responses are fast, the retrieval of "stimulus idiosyncrasies is minimal" however when the conditions are less than optimal (severe dysarthria) and responses are slow, "similarity in the surface characteristics of the stimuli proves a helpful aid to recalling" (p. 1240).

While there have been a number of researchers investigating aspects of speech perception in dysarthria (D'Innocenzo, Tjaden, & Greenman, 2006; Liss, et al., 2002; Liss, Spitzer, Caviness, Adler, & Edwards, 1998; Liss, Spitzer, Caviness, Adler, & W., 2000) to our knowledge Mattys and Liss is the only investigation to investigate the effects of different voices versus the same voices in dysarthric speech perception. Research into the influence of familiarity on disordered speech, and specifically on dysarthric speech have been conducted (DePaul & Kent, 2000; Hustad & Cahill, 2003; Klasner & Yorkston, 2005; Liss, et al., 2002; Patel & Schroeder, 2007; Pennington & Miller, 2007; Tjaden & Liss, 1995b) however, manipulation of variation of numbers of dysarthric speakers a listener is exposed to in a cohort, or variation in type or severity of dysarthria is sparse. It would be beneficial from a clinical and research perspective to know if speaker variability has an effect on intelligibility ratings. In addition, evidence of the implications of the effects of different speakers may need to be considered in research cohorts.

Klasner and Yorkston (2005) sought to identify the barriers to intelligibility and what strategies listeners used when attempting to two dysarthria types: mixed flaccid dysarthria secondary to Amyotrophic Lateral Sclerosis (ALS) and hyperkinetic dysarthria secondary to Huntington's disease (HD). Their aim was to determine if there was a variation in the barriers and strategies used by listeners in the two dysarthria types. The speech stimuli were 22 acoustically recorded sentences taken from the Assessment of Intelligibility of Dysarthria (Yorkston, et al., 1984). The sentences were judged as being between 60 to 90% intelligible by speech-language therapists experienced in assessment and treatment of dysarthric speech. Listeners between the ages of 18 to 35 were allocated to four focus groups, two groups heard sentences from the speakers with HD and two heard sentences from the speakers with ALS. The listeners tasks were: (1) listen, (2) orthographically transcribe what they hear, (3) compare the transcription to the target and (4) describe (through group discussion) what experiences influenced their understanding of the sentences. The researchers used four categories for the content analysis of the transcriptions: segmental-when commenting on speech sounds, suprasegmental-when commenting on rate, rhythm and/or prosody, linguistic-when commenting on content and meaning and cognitive-when commenting on listener attention and effort. Results showed that listeners judged segmental and linguistic barriers the most detrimental to intelligibility in both ALS and HD dysarthria, both of which are important to the application of bottom-up processing (Lindblom, 1990). The listeners highly endorsed cognitive strategies as the most helpful listening strategy in both dysarthria types however all four categories scored highly in the strategy scales. The researchers found that the listeners played an active role when trying to comprehend the sentences by using all of the strategies at their disposal. This study provides evidence the listeners may utilise both top-down and bottom-up processing in challenging listening situations such as disordered speech signals.

Similarly Hustad (2000) found that intelligibility and comprehensibility of dysarthric speech was increased when listeners had access to extrinsic information, such as topic cues to supplement their communication with dysarthric speakers. The author found that mixed cues that incorporated phonological information in the form of the speaker pointing to the initial phoneme of the word and topic information were the most powerful in increasing intelligibility. The influences of top-down and bottom-up perceptual processing working together, especially when the speech signal is disordered and additional information is needed to decipher speech is evident in research.

Study Aims

A number of perceptual processes are active during the perception of normal, distorted and disordered speech. The preceding discussion identified two critical factors that influence listeners' perception of both normal and disordered speech: (1) speech signal variability due to a number of different speakers or variation of a single speaker, and (2) perceptual learning through familiarisation to a voice and speech patterns. Signal variability has been shown to have a detrimental effect on speech perception whereas perceptual learning has been demonstrated to increased intelligibility, in some instances over a short period of time. While both of these factors have been reported in a number of investigations, research into the effects of speaker variability and perceptual learning in disordered speech, specifically dysarthria, have only recently gained momentum. The interest in listener strategies in deciphering disordered speech is of clinical interest. Clinicians and researchers recognise the importance of the listener in successful communication intervention and the need to expand therapy to include the communication partners. Therefore the primary aim of this study is to determine the effect of speaker numbers upon the listener's perception of dysarthric speech. It is hypothesised that intelligibility scores will be greater for listeners transcribing speech stimuli produced by a single speaker relative to multiple speaker productions. Such a prediction is based on evidence that a multi-speaker condition may negatively influence the accuracy of speech perception in normal, artificially distorted and accented speech (Cole, et al., 1974; Creelman, 1957; Horii, House, & Hughes, 1971; Mullennix, et al., 1989). The secondary aim is to investigate perceptual learning effects for listeners exposed to single-speaker condition, specifically, the degree to which intelligibility changes across a phrase transcription set. Given the prior studies in perceptual learning, for example accented speech (Bradlow & Bent, 2008; Clarke & Garrett, 2004), distorted speech (Davis, et al., 2005; Dupoux & Green, 1997; Greenspan, et al., 1988) and disordered speech (Liss, et al., 2002; Liss, et al., 1998; Tjaden & Liss, 1995b) it is hypothesised that intelligibility of dysarthric speech will increase over phrase presentation.

Method

Overview

This study utilised a between groups design to investigate the effect of multiple versus single speaker conditions upon the intelligibility of dysarthric speech. A secondary within groups analysis of the data was completed to investigate whether perceptual learning was evident within the rating period. Sixty listeners were randomly assigned to one of four listener conditions and completed a transcription task. One group transcribed phrases from a stimulus list comprised of six different speakers with dysarthria. The remaining three groups transcribed phrases from a stimulus set composed of a single speaker only.

Speech Stimuli Selection

The speaker stimuli were gathered as part of a larger research project. The speech stimuli of six individuals with hypokinetic dysarthria associated with PD were selected from the Motor Speech Disorders Laboratory database. All six were native speakers of New Zealand English (NZE) and had been diagnosed with PD by a neurologist. All speech stimuli were collected in a sound attenuated booth. During data collection, each speaker wore a head-mounted microphone positioned approximately five centimetres from the mouth. Their speech output was digitally recorded using Sony Sound Forge (v9.0, Madison Media Software, Madison, WI) at a sample rate of 48 kHz and saved in waveform audio file format (.wav). All speech was elicited through a reading task – the stimuli for reading were positioned at a comfortable location for the speaker. Speakers were encouraged to use their normal everyday speaking voice during the recording.

The speakers completed a number of speech tasks (e.g., conversational speech, passage reading) as part of the larger investigation. However, for the purposes of the current study, only the phrase stimuli were extracted for use and, therefore, are discussed. Each participant read sixty phrases based on the prior work of Cutler and Butterfield (1992). All phrases were 6 syllables in length and contained between 4 and 5 words. The phrases were of low predictability to reduce the potential for semantic or contextual clues, and all were grammatically correct.

Experimental stimuli.

Speech stimuli were selected from six speakers with hypokinetic dysarthria. The speakers were selected based on their speech characteristics. All exhibited salient speech features representative of hypokinetic dysarthria. Details of the participants are presented in table one.

Table 1.

Biographical details of the individuals with Parkinson's disease

Speaker	Primary Diagnosis	Age	SIT score	Predominant Speech Characteristics
1	PD	77	65%	Moderate articulatory imprecision Moderate breathy vocal quality Short, rapid rushes of speech Mild-moderate reduced volume Mild-moderate hypernasality
2	PD	70	70%	Moderate impaired articulation Moderate breathy vocal quality Short, rapid rushes of speech Intermittent palilalia
3	PD	70	75%	Moderate impaired phoneme production Fatiguing vocal behaviour resulting in fading out before the end of an utterance Mild-moderate reduced volume Intermittent short rushes of speech
4	PD	70	84%	Mild imprecise articulation Reduced volume Breathy vocal quality Mild reduced volume Fast rate of speech
5	PD	69	92%	Mild imprecise articulation Breathy vocal quality Reduced volume Fast rate of speech
6	PD	64	98%	Mild imprecise articulation with slurred consonants Reduced volume Fast rate of speech

Note: PD= Parkinson's Disease.

Of the 6 speakers, 3 were selected for inclusion as "experimental" speakers (speakers 1 to 3). They exhibited moderate reductions in intelligibility evidenced through scores of between 55% - 75% on the Sentence Intelligibility Test (SIT) (Yorkston, et al., 1984). The remaining three participants, speakers 4 to 6, were included as fillers and exhibited mild to moderate reductions in intelligibility on the SIT. From each of the three experimental stimuli 10 phrases were selected for use in the listening experiment. To ensure experimental rigour, the 10 experimental phrases were balanced across speakers for the presence of alternating phrasal stress patterns. So five exhibited a weak-strong pattern and five phrases exhibited a strong-weak phrasal stress patterning. Those phrases selected for use in the listening experiment were screened to ensure that they exhibited no mispronunciations or hesitations.

In addition, it was confirmed by perceptual analysis that all exhibited the presence of the perceptual characteristics outlined previously.

Multiple and single speaker phrase sets.

Four sets of 60 phrases were constructed from the selected speech stimuli. Phrase set one was comprised of the ten selected experimental phrases from each of the three "experiential" speakers with PD (i.e., 30 phrases) and ten selected experimental phrases from each of the "filler" speakers with PD (i.e., 30 phrases) – this phrase set was termed the "multiple speaker condition". Phrase sets two through four were termed the "single speaker conditions". Each phrase set was comprised of the stimuli from one of the experimental speakers only. Therefore, phrase set two comprised the 60 speech stimuli of speaker one, phrase set three the 60 speech stimuli of speaker two, and phrase set three the 60 speech stimuli of speaker three.

Listeners

Sixty participants aged eighteen to forty years (M = 23.5 years) completed the listening experiment. All participants were native speakers of NZE, reported no language, cognitive or learning disability, and no significant experience or contact with persons with dysarthria. All passed a pure-tone hearing screen at 20 dB HL at 1000, 2000, and 4000Hz and 30 dB HL at 500Hz. The listener participants were recruited from the student body of the University of Canterbury and the researcher's family and contacts. The speech-language therapy students recruited for the study did not receive any incentives for participation as there is a professional expectation that they would participate in research to further their understanding of the process. The student body from the psychology department undergraduate degree received course credit for their participation and all other participants were offered a $10 gift voucher in return for their participation.

Procedure

The 60 listener participants were randomly assigned to one of the four phrase set groups. The participants completed the experiment individually in a sound attenuated room. The speech stimuli were presented via a custom-designed experimental programme through high quality headphones (Sennheiser HD 280). Prior to the experiment, all phrases were calibrated to loudness levels within ±0.1dB. They were presented at 65 dB across all listener participants.

The experiment was completed using software designed by Dr Greg O'Beirne and programmed in LabVIEW (8.20 (National Instruments, Tx, USA). Participants were told that they would hear a series of phrases. These phrases may be difficult to understand and although they contained real English words, the phrases would not make sense. Participants were further instructed to listen carefully to each phrase and type into a text box, via keyboard, what they thought they had heard. If they were unable to understand the words in the phrase they were instructed to make a guess and, if they could not guess, to type a slash (/) or cross (x) where they thought they heard a word. These instructions were presented as part of the custom-designed experimental programme and, therefore, were identical across participants.

Transcription Analysis

The experiment resulted in a total data set of 60 transcripts of 60 phrases totalling 3600 transcripts. This included 15 transcripts of 60 phrases in each of the four phrase set experimental conditions. The researcher independently coded each experimental transcription for percent words correct. The phrases were coded for percent words correct using the coding criteria of Liss and colleagues (2002). Specifically, a word was scored as correct if: 1) it was an exact match; 2) it differed only by tense or plural suffix and did not differ in any other syllable; or 3) the word was an 'a' - 'the' substitution and did not differ in any other syllable.

Reliability of Transcription Coding

Ten percent of the transcripts were randomly selected to calculate inter-judge and intra-judge reliability. The transcripts were re-analysed by the researcher to determine intra-rater reliability and a second rater so that inter-judge estimates of reliability could be obtained. Pearson correlation analyses were conducted on the percent intelligibility scores. A correlation of 0.97 was achieved for the intra-judge reliability (significant at $p < 0.01$), and a correlation of 0.96 was achieved for the inter-judge reliability (significant at $p < 0.01$).

Statistical Analysis

Independent group *t*-tests were used to compare the intelligibility scores of the multi-speaker group and the single speaker groups. A mixed between-within subjects ANOVA was conducted to explore the effects of time (first 10 versus the last 10 experimental phrases) in the within-subjects variable and the between-subjects (speaker one, two, or three), on the dependent variable of percent correct. For all comparisons, the alpha level was set at $p < 0.05$ and effect sizes reported for significant results.

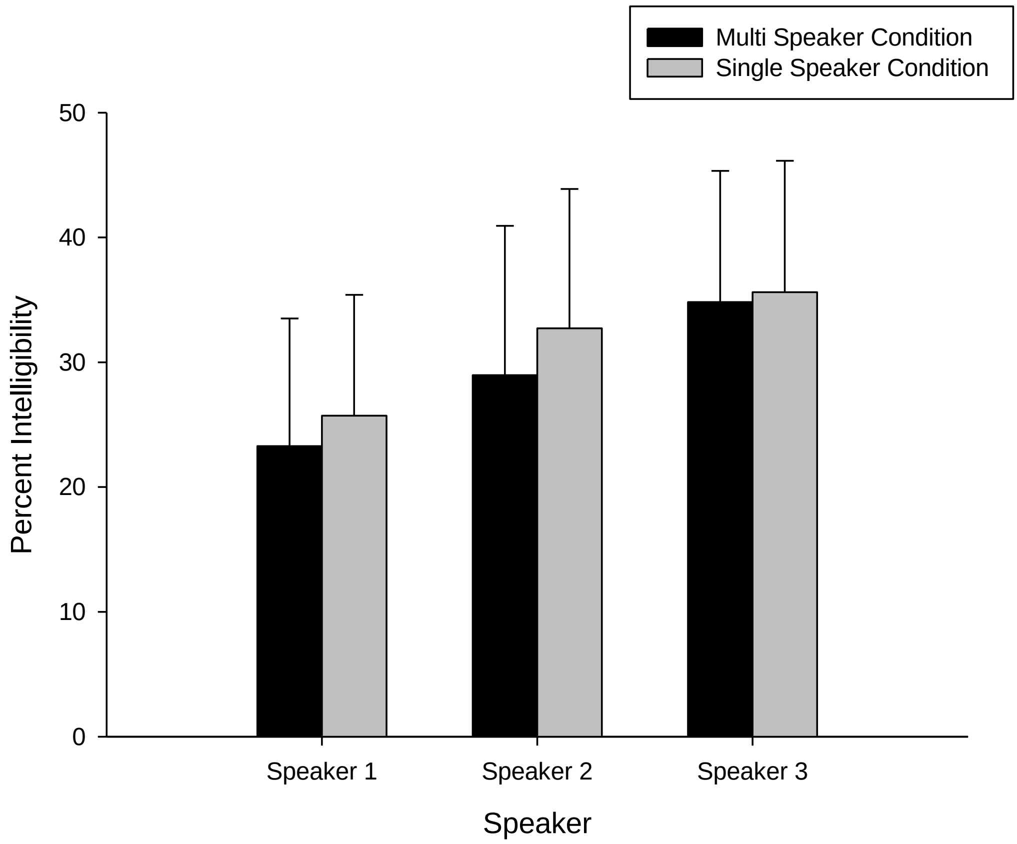
Results

Phrase Intelligibility

Figure one shows the mean percent intelligibility scores for each of the three experimental stimuli speakers when transcribed by listeners who heard the speaker in the multiple-speaker versus single-speaker conditions. While a trend towards higher intelligibility scores in the single-speaker conditions was observed, statistical analysis using independent groups *t*-tests showed that percent intelligibility scores were not significantly different when obtained in the multiple-speaker and single-speaker conditions from speakers one, $t\ (28) = .67$, $p = .51$, two, $t\ (28) = .89$, $p = .38$, and three, $t\ (28) = .21$, $p = .84$, respectively.

Figure 1.

Comparison of mean percent correct scores for the ten experimental phrases selected from the single and multiple speaker conditions.

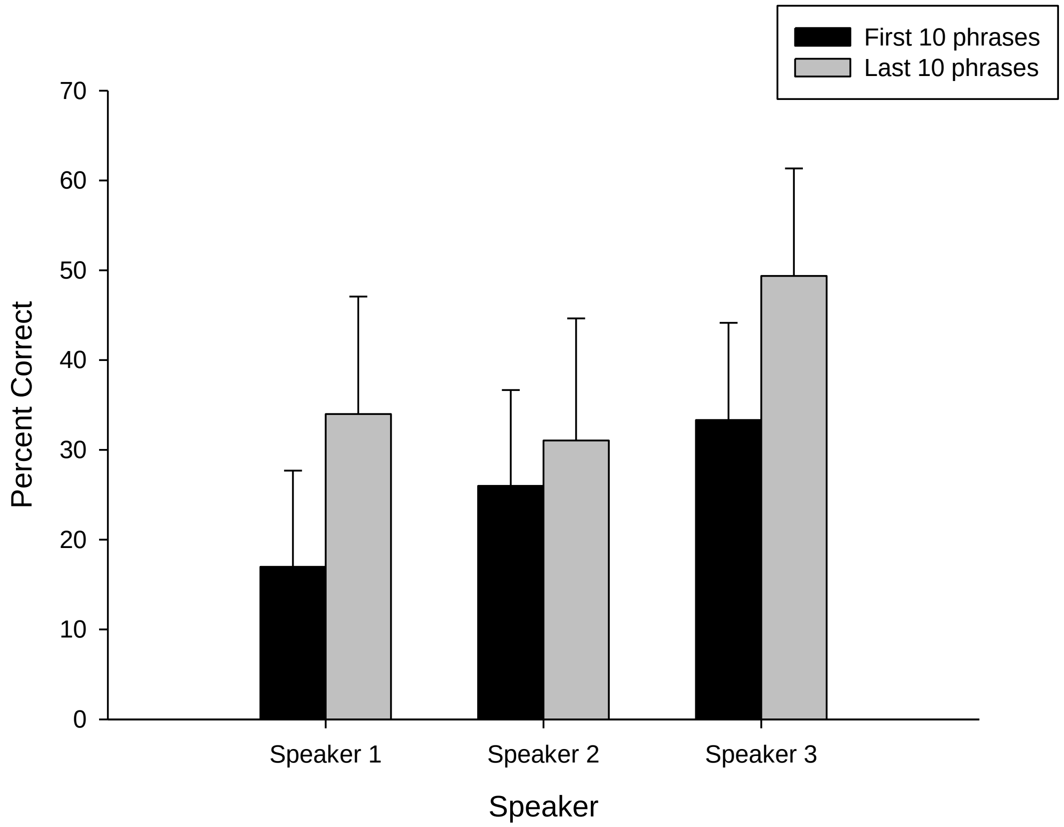


Perceptual Learning

Figure two shows the mean percent intelligibility scores derived from listener transcripts obtained in the single speaker conditions only. It compares the average percentage intelligibility of the first ten phrases with the last ten phrases across the three single-speaker conditions (i.e., assigned to speaker one, speaker two, and speaker three respectively). A mixed between-within subjects ANOVA was conducted to explore the effects of time (first 10 versus last 10 phrases) and group (assigned to speaker one, two, or three) on the dependent variable of percent correct. There was a significant main effect of time, F (1,42) = 61.63, $p <$ 0.001, $\eta p^2 = .60$, showing that overall, the percent correct for the last ten phrases was significantly higher than for the first ten phrases. In addition, there was a statistically significant main effect of group, F (2,42) = 9.55, p <0.001, $\eta p^2 = .32$, demonstrating, as expected, variation in overall intelligibility existed across speakers one, two, and three. The time X group interaction was also significant, F = 5.61, p = 0.007, $\eta p^2 = .21$. Post hoc examination of the interaction effect using paired *t*-tests indicated that a clear perceptual learning effect was demonstrated for the listeners who transcribed speakers one, t (14) = -7.20, p <0.001, d = -2.71, and three, t (14) = -5.20, p < 0.001, d = -1.94. However, while a trend was observed, there was no significant difference between intelligibility scores in the first and last 10 phrases for listeners who heard speaker two, t (14) = -1.70, p = .11.

Figure 2.

Comparison of mean percent correct scores for the first ten and the last ten phrases in the single speaker conditions



Discussion

The aims of this study were twofold: (1) to determine if listeners exhibited more difficulty transcribing phases when they were exposed to a multi-speaker (six different speakers) condition as opposed to a single speaker condition; and (2) to examine perceptual learning over phrases in a cohort of speakers with dysarthria. In relation to these aims no significant difference in average intelligibility scores was found when phrases transcribed in the multi-speaker and single speaker groups were compared. Further, a significant learning effect was observed with the intelligibility of the first 10 phrases significantly less than the final ten phrases for two of the three single speaker groups. However, this same effect was not observed for transcripts from the speech samples of a third speaker.

Multi-speaker versus single speaker

Contrary to expectations, intelligibility scores did not differ significantly for listeners transcribing phrases in the single versus multi-speaker conditions. While a general trend revealed an advantage in the single speaker groups, this was non-significant. Based on previous research, our expectation was that there would be a single-speaker advantage in disordered speech. Research has clearly shown a single speaker advantage in normal speech (Cole, et al., 1974; Creelman, 1957; Mullennix, et al., 1989) and distorted speech (Mullennix, et al., 1989) however, for this particular cohort of listeners and speakers that was not the case.

There are two possible reasons for this lack of single-speaker advantage. First, it is possible that across-trial variation in the speech characteristics of the individuals with PD may have resulted in the single-speaker groups needing to adjust to variation similar to the listeners in the multi-speakers group (across-speaker variation). This may have had an equalising effect on the end result. Adjustment may not occur at all for listeners who hear disordered speech of a severity over a certain level as it does for normal and accented speech. Liss (2007) suggested that listeners perceive disordered speech in the same way as they perceive normal speech, however the perceptual challenge results in less success. Perhaps the speakers chosen for this study had speech characteristics that challenged all of the listeners, both in the single and multi-speaker groups, so that both groups were equally disadvantaged by variability. Intra-participant variability in dysarthric speech has been investigated and variability in certain characteristics such as utterance duration and voice onset time of the speech signal have been evidenced, specifically in dysarthria associated with Huntington's disease (Hertrich & Ackermann, 1994). Future research into the variability of the speech signal within one speaker over the course of a set of phrases would be valuable to understand more about the adjustment process that needs to occur in speech perception of dysarthric speech. Research has established perceptual adjustment occurs when speakers change, however, given the variability that may occur in the speech of a dysarthric speaker perhaps the normalisation process does not occur as it does in normal speech. If the normalisation process was inhibited for some reason, analysis of perceptual learning becomes important. Consequently, analysis of the single speaker groups becomes of integral interest to this current study.

Secondly, it is possible that the small number of experimental phrases and speakers did not allow for an effect to be seen. The results show a clear trend of a single speaker advantage, therefore the possibility that increasing both number of experimental phrases and number of speakers may have enabled the effect to be more pronounced.

Comparison between Early and Later Transcriptions in a Single Speaker Condition

This study also asked whether familiarity with voice and speech patterns of dysarthric speech results in improved transcription accuracy over time. We expected to see a perceptual learning effect occurring over the duration of the 60 phrases produced by a single speaker, for all three speakers. This type of perceptual learning adjustment would be reflected in increased intelligibility scores toward the end of the phrase set for each speaker. While significant intelligibility gains were observed between the first and last 10 phrases for two of the three speakers; the gains over time observed for speaker two were not statistically significant.

Prior research had reported that intelligibility levels affect the length of exposure required for listeners to make perceptual learning gains. In Bradlow and Bent's study on accented speech (2008), perceptual learning gains were made by listeners over all levels of foreign accented speech, however those listeners in the lower intelligibility groups required a longer exposure to the speech signal to make the same gains as those in the higher intelligibility groups. One might expect a similar result in the current study had the speakers been of differing intelligibility levels. However, the three speakers in the current study were of a similar intelligibility level as seen by the Sentence Intelligibility Test scores. This may suggest that in disordered speech perceptual learning may not occur in the same way as accented speech. Davis et al. (2005) found that intelligibility in noise-vocoded speech utilised top-down processing to assist the word recognition at the pre-lexical level. The BKB sentences used in the Barlow and Bent study provided contextual information that would assist in transcription, whereas the phrases used in the current study were of a low probability level. If this were the case, again, a similar pattern would be seen for all three speakers. Although variability within a single speaker has not been addressed in this study, this has been the subject of other studies. Kraljic, Samuel and Benman (2008) investigated perceptual learning with a variation in a speaker's production of two phonemes. They found that learning occurred when the listener perceived the change to be characteristic; however it did not occur when the listener perceived the change as incidental. Individual speaker variability may also provide some explanation for the current findings. Speaker two perhaps presented with a more variable speech signal than that of the other two speakers resulting in a lower perceptual learning condition. Additionally, variability within a speaker may have a detrimental effect on the normalisation process. If speaker two has more variability in the speech signal than the other two, perhaps the listeners were consistently adjusting to the speech rather than gaining perceptual learning advantages. The fact that variability within a single speaker has not been addressed is a significant limitation to this study. Acoustic analysis of each of the speech signals to identify what variability's may have contributed to the result and may increase the understanding of why this anomaly exists.

Limitations and Future Research

This study raises a number of questions. Why was the normalisation difference between multi-and single speaker groups so small where prior research provides evidence that there is a marked difference in these two conditions in normal and distorted speech signals? Why did a significant perceptual learning effect exist for only two out of the three speakers? Previous studies in adjustment in speech perception have had a higher number of listeners and a larger set of phrases. The current study had 60 phrases, however only ten out of the 60 were analysed in the intelligibility statistics. It is possible that ten phrases was not an adequate enough number to obtain statistically significant effects of normalisation. An increased number of experimental phrases would benefit future research in this area to determine if normalisation does occur at a significant level.

Acoustic analysis has not been completed on the speaker phrases for this project, therefore perceptual analysis was the sole measure of the characteristics of the speech signal. Acoustic measures of phrase duration, fundamental frequency and intensity variation (Liss, et al., 1998) may have offered additional insight into why the effects of perceptual learning for speaker six were reduced. Future studies investigating acoustic variability within a single subject with hypokinetic dysarthria would benefit the understanding of both normalisation and perceptual learning in disordered speech.

Although perceptual learning did occur between the first ten and last ten phrases, we have not identified how quickly the learning occurred, only that it had occurred over the course of 60 phrases. Research in noise-vocoded speech shows a perceptual learning effect after just 30 phrases (Davis, et al., 2005), and two to four sentences of accented speech (Clarke & Garrett, 2004). It may have been beneficial to have examined a set of ten phrases taken from the middle of the phrase set to determine if perceptual learning had occurred earlier. This would also explore the possibility of an effect of fatigue that may have occurred after listening and transcribing 50 speech disordered phrases. Future research that measures how quickly listeners to dysarthric speech show evidence of perceptual learning would be important to establish minimal baselines for therapy.

Clinical Implications

This study highlights the effects of perceptual learning in disordered speech. Perceptual learning may have a detrimental effect in scoring intelligibility in standardised tests such as SIT. In best practice, transcription may be completed by a naive listener, however even if the listener has had no prior experience with the speaker, perceptual learning is likely to have an effect on the sentence transcription that follows the transcription of 30 words and may result in inflated intelligibility scores. In typical practice situations, transcription of the SIT may be completed by a clinician already familiar with the disordered speech signal. This may result in the clinician arriving at scores that are not a true representation of intelligibility. Consequently caution needs to be taken in the clinical arena when identifying intelligibility through orthographic transcription alone.

Conclusions

This study found that normalisation or adjustment to different speakers compared to single speakers does not occur similarly for naturally disordered speech as it does for normal and other distorted speech signals. A trend towards single speaker advantage was shown however the difference was small. Furthermore, perceptual learning effects were found however, these were not consistent of the three speakers. The results show that further investigation is required into the acoustic variability within a single speaker to ensure that normalisation is accounted for in research designs. In addition, normalisation and perceptual learning are intrinsically intertwined in the perception of speech. Consideration needs to be given to this in intelligibility analysis in clinical practice.

References

Bamford, J., & Wilson, I. (1979). Methodological considerations and practical aspects of the BKB sentence list. In J. Bench & J. Bamford (Eds.), *Speech-hearing tests and the spoken language of hearing-impaired children* (pp. 148-187). London: Academic Press.

Bladon, R. A. W., Henton, C. G., & Pickering, J. B. (1984). Towards an auditory theory of speaker normalization. *Language & Communication, 4*(1), 59-69.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707-729. doi: DOI: 10.1016/j.cognition.2007.04.005

Clarke, C. (2002). *Perceptual adjustment to foreign-accented English with short term exposure*. Paper presented at the 7th International Conference on Spoken Language Processing, Denver, Colorado, USA. http://www.isca-speech.org/archive/icslp_2002/i02_0253.html

Clarke, C., & Garrett, M. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America, 116*(6), 3647-3658.

Cole, R. A., Coltheart, M., & Allard, F. (1974). Memory of a speaker's voice: Reaction time to same- or different-voiced letters. *Quarterly Journal of Experimental Psychology, 26*(1), 1-7.

Creelman, C. (1957). Case of the Unknown Talker. *The Journal of the Acoustical Society of America, 29*(5), 655-655.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31*(2), 218-236. doi: Doi: 10.1016/0749-596x(92)90012-m

D'Innocenzo, J., Tjaden, K., & Greenman, G. (2006). Intelligibility in dysarthria:Effects of listener familiarity and speaking condition. *Clinical Linguistics & Phonetics, 20*(9), 659-675.

Dagenais, P. A., Watts, C. R., Turnage, L. M., & Kennedy, S. (1999). Intelligibility and acceptability of moderately dysarthric speech by three types of listeners. *Journal of Medical Speech-Language Pathology, 7*(2), 91-96.

Darley, F., Aronson, A., & Brown, J. (1969a). Clusters of Deviant Speech Dimensions in the Dysarthrias. *J Speech Hear Res, 12*(3), 462-496.

Darley, F., Aronson, A., & Brown, J. (1969b). Differential Diagnostic Patterns of Dysarthria. *Journal of Speech Hearing Research, 12*(2), 246-269.

Davis, K. H., Biddulph, R., & Balashek, S. (1952). Automatic Recognition of Spoken Digits. *The Journal of the Acoustical Society of America, 24*(6), 637-642.

Davis, M., Johnsrude, I., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the

Comprehension of Noise-Vocoded Sentences. *Journal of Experimental Psychology: General, 134*(2), 222-241. doi: 10.1037/0096-3445.134.2.222

10.1037/0096-3445.134.2.222.supp (Supplemental)

DePaul, R., & Kent, R. D. (2000). A longitudinal case study of ALS: Effects of listener familiarity and proficiency on intelligibility judgments. *American Journal of Speech-Language Pathology, 9*(3), 230-240.

Duffy, J. (2005). *Motor Speech Disorders: substrates, Differential Diagnosis, and Management* (2nd ed.). Missouri: Elsevier Nosby.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 914-927. doi: 10.1037/0096-1523.23.3.914

Eadie, T. L., & Doyle, P. C. (2005). Scaling of Voice Pleasantness and Acceptability in Tracheoesophageal Speakers. *Journal of Voice, 19*(3), 373-383. doi: DOI: 10.1016/j.jvoice.2004.04.004

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 421-433. doi: 10.1037/0278-7393.14.3.421

Grosjean, F. (1980). Spoken word recognition processs and the gating paradigm. *Perception and Psychophysics, 28*(4), 267-283.

Hertrich, I., & Ackermann, H. (1994). Acoustic Analysis of Speech Timing in Huntington's Disease. *Brain and Language, 47*(2), 182-196. doi: DOI: 10.1006/brln.1994.1048

Hervais-Adelman, A., Davis, M., Johnsrude, I., Taylor, K., & Carlyon, R. (2010). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology: Human Perception and Performance*. doi: 10.1037/a0020772

Horii, Y., House, A. S., & Hughes, G. W. (1971). A Masking Noise with Speech-Envelope Characteristics for Studying Intelligibility. *The Journal of the Acoustical Society of America, 49*(6B), 1849-1856.

Hustad, K. (2000). *Effects of context on intelligibility and comprehensibility of severely dysarthric speech.* 60, ProQuest Information & Learning, US. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2000-95010-346&site=ehost-live Available from EBSCOhost psyh database.

Hustad, K., Beukelman, D., & Yorkston, K. (1998). Functional outcome assessment in dysarthria. *Seminars in Speech and Language, 19*, 291-302.

Hustad, K., & Cahill, M. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology, 12*(2), 198-208. doi: 10.1044/1058-0360(2003/066)

Johnson, K. (2009). Speaker normalization in speech perception. In D. B. Piosn & R. E. Remez (Eds.), *The Handbook of Speech Perception*. Malden, USA: Blackwell Publishing.

Jusczyl, P. W., Pisoni, D. B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition, 43*(3), 253-291.

Klasner, E. R., & Yorkston, K. M. (2005). Speech Intelligibility in ALS and HD Dysarthria: The Everyday Listener's Perspective. *Journal of Medical Speech-Language Pathology, 13*(2), 127-139.

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science, 19*(4), 332-338. doi: 10.1111/j.1467-9280.2008.02090.x

Kuhl, P., & Miller, J. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perceptions & Psychophysics, 31*(3), 279-292.

Lindblom, B. (1990). On the communication process: Speaker-listener interaction and the development of speech*. *Augmentative and Alternative Communication, 6*(4), 220-230. doi: doi:10.1080/07434619012331275504

Liss, J. M. (2007). The role of speech perception in motor speech disorders. In G. Weismer (Ed.), *Motor Speech Disorders: Essays for Ray Kent* (pp. 187-219). San Diego: Pleural Publishing Inc.

Liss, J. M., Spitzer, M., Caviness, J. N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America, 112*(6), 3022-3030.

Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *The Journal of the Acoustical Society of America, 104*(4), 2457-2466.

Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., & W., E. B. (2000). Lexical boundary error analysis in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America, 107*(6), 3415-3424.

Luce, P. (1987). The neighborhood activation model of auditory word recognition. *The Journal of the Acoustical Society of America, 81*(S1), S1-S1.

Luce, P., McLennan, C., & Charles-Luce, J. (2003). Abstractness and specificity in spoken word recogntion: indexical and allophonic variability in long-term repetition priming. In J. Bowers & C. Marsolek (Eds.), *Rethinking Implicit Memory*. New York: Oxford University Press.

Mackenzie, C., & Lowit, A. (2007). Behavioural intervention effects in dysarthria following stroke: communication effectiveness, intelligibility and dysarthria impact. *International Journal of Language & Communication Disorders, 42*(2), 131-153. doi: doi:10.1080/13682820600861776

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance, 33*(2), 391-409.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*(1), 29-63.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(4), 676-684.

Mattys, S., & Liss, J. (2008). On building models of spoken-word recognition: When there is as much to learn from natural "oddities" as artificial normality. *Perception & Psychophysics, 70*(7), 1235-1242. doi: 10.3758/pp.70.7.1235

McAuliffe, M. J., Ward, E. C., & Murdoch, B. E. (2005). Articulatory Function in Hypokinetic Dysarthria: An Electropalatographic Examination of Two Cases. *Journal of Medical Speech-Language Pathology, 13*(2), 149-168.

McAuliffe, M. J., Ward, E. C., & Murdoch, B. E. (2007). Intra-participant variability in Parkinson's disease: An electropalatographic examination of articulation. *Advances in Speech Language Pathology, 9*(1), 13-19.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

Miller, N., Noble, E., Jones, D., & Burn, D. (2006). Life with communication changes in Parkinson's disease. *Age and Ageing, 35,* 235-239.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47*(4), 379-390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America, 85*(1), 365-378.

Murdoch, B. E. (1998). The neuroanatomical framework of dysarhtria. In B. E. Murdoch (Ed.), *Dysarthria: A Physiological Approach to Assessment and Treatment*. Cheltenham: Stanley Thornes (Publishers) Ltd.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & psychophysics, 60*(3), 355-376.

Patel, R., & Schroeder, B. (2007). Influence of familiarity on identifying prosodic vocalizations produced by children with severe dysarthria. *Clinical Linguistics & Phonetics, 21*(10), 833-848. doi: 10.1080/02699200701559476

Peelle, J. E. (2005). *Flexibility in speech comprehension: Evidence from time-compressed speech.* 65, ProQuest Information & Learning, US. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2005-99008-272&site=ehost-live Available from EBSCOhost psyh database.

Peelle, J. E., & Wingfield, A. (2005). Dissociations in Perceptual Learning Revealed by Adult Age Differences in Adaptation to Time-Compressed Speech. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1315-1330. doi: 10.1037/0096-1523.31.6.1315

Pennington, L., & Miller, N. (2007). Influence of listening conditions and listener characteristics on intelligibility of dysarthric speech. *Clinical Linguistics & Phonetics, 21*(5), 393-403. doi: 10.1080/02699200701276675

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology, 79*(12), 1-16. doi: 10.1037/h0026947

Posner, M. I., & Keele, S. W. (1967). Decay of Visual Information from a Single Letter. *Science, 158*(3797), 137-139.

Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America, 106*(6), 3629-3636. doi: 10.1121/1.428215

Samuel, A. G. (1981). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance, 7*(5), 1124-1131.

Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R. Kent (Ed.), *Intelligibility in Speech Disorders* (Vol. 1). Philadelphia: John Benjamins North America.

Schiavetti, N., Metz, D. E., & Sitler, R. W. (1981). Construct Validity of Direct Magnitude Estimation and Interval Scaling of Speech Intelligibility: Evidence from a Study of the Hearing Impaired. *J Speech Hear Res, 24*(3), 441-445.

Schiavetti, N., Sacco, P. R., Metz, D. E., & Sitler, R. W. (1983). Direct Magnitude Estimation and Interval Scaling of Stuttering Severity. *J Speech Hear Res, 26*(4), 568-573.

Schrag, A., Selai, C., Davis, J., Lees, A., Jahanshahi, M., & Quinn, N. (2003). Health-related quality of life in patients with progressive supranuclear palsy. *Movement Disorders, 18*(12), 1464-1469.

Stevens, G. (1975). *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects*. USA: John Wiley & Sons, Inc.

Tjaden, K., & Liss, J. (1995a). The Influence of Familiarity on Judgments of Treated Speech. *Am J Speech Lang Pathol, 4*(1), 39-48.

Tjaden, K., & Liss, J. (1995b). The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics & Phonetics, 9*(2), 139-154. doi: doi:10.3109/02699209508985329

Tjaden, K., & Wilding, G. (2004). Rate and Loudness Manipulations in Dysarthria: Acoustic and Perceptual Findings. *J Speech Lang Hear Res, 47*(4), 766-783. doi: 10.1044/1092-4388(2004/058)

Toner, M. A., & Emanuel, F. W. (1989). Direct Magnitude Estimation and Equal Appearing Interval Scaling of Vowel Roughness. *J Speech Hear Res*, *32*(1), 78-82.

Turner, G., Tjaden, K., & Weismer, G. (1995). The Influence of Speaking Rate on Vowel Space and Speech Intelligibility for Individuals With Amyotrophic Lateral Sclerosis. *J Speech Hear Res, 38*(5), 1001-1013.

Turner, G., & Weismer, G. (1993). Characteristics of Speaking Rate in the Dysarthria Associated With Amyotrophic Lateral Sclerosis. *J Speech Hear Res, 36*(6), 1134-1144.

Weismer, G., Jeng, J.-Y., Laures, J., S, Kent, R., D., & Kent, J., F. . (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatrica et Logopaedica*, *53*(1), 1.

Whitehill, T. L., Lee, A. S. Y., & Chun, J. C. (2002). Direct Magnitude Estimation and Interval Scaling of Hypernasality. [Article]. *Journal of Speech, Language & Hearing Research, 45*(1), 80.

Yorkston, K., Beukelman, D., Strand, E., & Bell, K. (1999). *Management of Motor Speech Disorders in Children and Adults*. Austin, Texas: Pro-Ed.

Yorkston, K., Beukelman, D., & Traynor, C. (1984). Assessment of Intelligibility of Dysarthric Speech.

Zraick, R. I., & Liss, J. M. (2000). A comparison of equal-appearing interval scaling and direct magnitude estimation of nasal voice quality. *Journal of Speech, Language & Hearing Research, 43*(4), 979-988.