University of Alberta

Effects of Talker Severity and Repeated Presentations on Listener Judgments of the Speech Intelligibility of Young Children with Dysarthria

by

Kimberley Joy Coté-Reschny ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Science in Speech-Language Pathology

Department of Speech Pathology and Audiology

Edmonton, Alberta
Spring 2007

Your file Votre référence
ISBN: 978-0-494-29947-0
Our file Notre référence
ISBN: 978-0-494-29947-0

ABSTRACT

Effects of talker severity and number of presentations of a stimulus and their potential interaction, on listeners' abilities to understand the speech of children with dysarthria, presented in noise (+10 dB SNR), were investigated. Main effects of talker severity and presentation number were found for two dependent variables; listeners' word identification scores (i.e., lowest for most severe dysarthria, higher with repeated presentations) and their response times (i.e., longest for most severe dysarthria, shorter with repeated presentations). This study also investigated the effect of talker severity on a third dependent variable, listeners' ratings of the effort required to understand the children's words, using Direct Magnitude Estimation. A main effect of talker severity was found (i.e., highest effort ratings for most severe dysarthria). Relationships between pairings of the three dependent variables were investigated. Theme analyses of listeners' beliefs related to the ease or difficulty of the word identification task were conducted.

Acknowledgements

This thesis was funded in part by the Natural Sciences and Engineering Research Council of Canada.

I would first like to thank my supervisor, Dr. Megan Hodge, for all the hard work that she put into this entire project. Her encouragement, dedication and enthusiasm have been admired since before this journey began. I would not be where I am today without her support.

I would also like to thank Bill Hodgetts for his suggestions for the design and analysis of the study and the use of background noise, his help with obtaining the noise file and technical support with normalizing the audio files. Thank you to Carrie Gotzke for her never-failing supportive advice and her willingness to lend a hand, regardless of the task, whenever it was needed. I would also like to thank Helen Southwood for her valuable advice regarding the design and analysis of the DME task and Chris Westbury for his assistance with the phonological neighbourhood and word frequency values. In addition, thank you to Jason Daniels for creating the program to my very particular specifications. I would also like to thank my committee members for taking the time to provide me with valuable suggestions and their commitment to this project and the listeners who participated in this study for taking time away from beautiful summer days to come sit in a sound booth.

Finally, I would like to thank my husband Josh for his understanding and support during this process. He has always been there for me, being the rock through the tough times and helping celebrate the good. Thank you.

This thesis is dedicated to my grandparents, Fred and Connie Keet.

Table of Contents

List of Tables

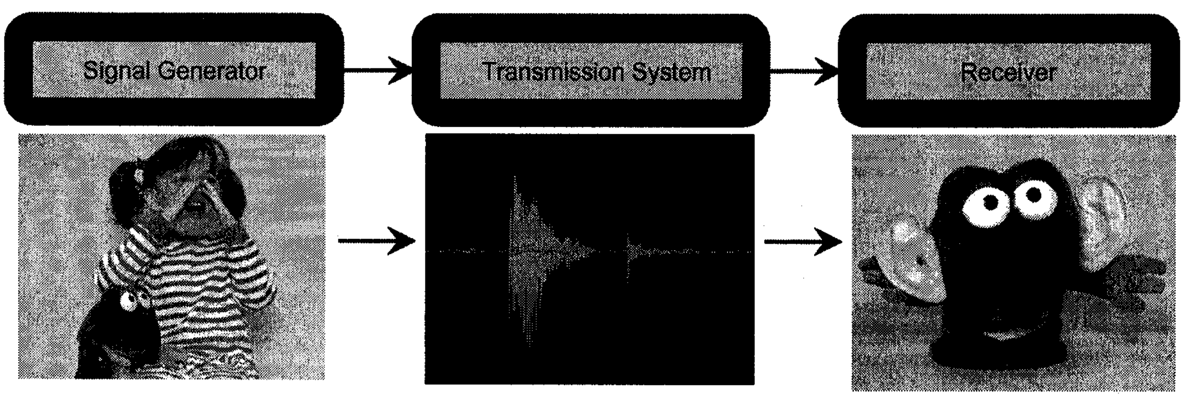
List of Figures

INTRODUCTION

Overview

Spoken communication is a process that involves sending and receiving messages that are coded in an acoustic (sound) signal. This type of communication requires a signal generator (talker), a transmission system (the medium, or environment, that the speech signal travels through), and a receiver (listener). These components are illustrated in Figure 1.

Figure 1. Components of the process of spoken communication include the talker (signal generator), the environment (transmission system) and the listener (receiver).



Pictures from *Test of Children's Speech Plus* (http://www.tocs.plus.ualberta.ca/).
Used with permission from M. Hodge.

For most individuals, this complex process is carried out with ease. However, for talkers who produce speech that is difficult to understand, for spoken communication that takes place in a noisy environment or for listeners who have a hearing impairment, this process can be a very difficult one. It is apparent then, that the effort that both talkers and listeners invest in a conversation will influence the success of their communicative interaction. Ways to attempt to quantify the effort that individuals give to a communicative interaction include measuring the amount of physical and mental energy talkers use and the amount of mental energy that listeners need to expend to understand what was said. Breakdowns in spoken communication can occur at any of these levels and the success with which the breakdown is repaired has an impact on the success of the current and future interactions. How much of a message is understood by a listener can be defined both in terms of speech intelligibility and in terms of comprehensibility.

Speech intelligibility refers to how much of the sound signal that the talker produces is understood by the listener (Kent, Weismer, Kent, & Rosenbek, 1989; Yorkston & Beukelman, 1980). Comprehensibility refers to "the extent to which a listener understands utterances produced by a speaker in a communication context" (Barefoot, Bochner, Johnson & Eigen, 1993). Comprehensibility reflects the condition where the listener uses everything that is available to him or her to understand the message, including the sound signal produced by the talker. Therefore, comprehensibility can be influenced by the predictability of the message in the context where the communicative interaction occurs (e.g., topics of conversation in a classroom may be different than topics of conversation at a restaurant) and the participants (e.g., topics of conversation with a friend may be different than topics of conversation with a grandparent).

Comprehensibility can also be influenced by the listener's motivation to understand the talker's message, the quality of visual cues that the talker provides (e.g., lip reading cues, gestures that support what is being said, written cues such as the first letter of the word being attempted) and the listener's familiarity with the talker, among other factors (see Bradlow & Pisoni, 1999; Hustad & Cahill, 2003; Kent, Miolo, & Bloedel, 1994). Factors that affect speech intelligibility and comprehensibility are of interest in making recommendations to children who have reduced speech intelligibility due to dysarthria about how to repair communication breakdowns.

Literature Review

Factors Affecting Speech Intelligibility

Speech intelligibility refers to how successfully listeners can understand a talker's spoken message (Kent, 1988; Kent et al., 1989; Yorkston & Beukelman, 1980). For the purpose of this study, speech intelligibility refers to how much a listener can understand when given only the auditory (i.e., acoustic) signal, such as when listening to recorded speech being played back (Kent et al., 1989). Characteristics of the talker, the environment and the listener can all affect the clarity, or intelligibility, of the message that is being conveyed (Kent, 1992). Breakdowns in spoken communication that result in a mismatch between the message intended and the message received can occur at any of these three components or some combination.

The signal may be distorted if the talker has a speech disorder. This disorder may affect articulatory accuracy and the loudness, vocal quality, rate and prosody of speech, leading to a signal that is difficult to understand because the clarity of the signal has been reduced (Hodge & Wellman, 1999). The signal may be distorted by intervening background noise in the environment as is evidenced by the difficulty that people experience when trying to listen to a conversation in a noisy room. This background noise can mask or overpower the signal and make it hard to decipher. The signal can also be more difficult to understand if there are distractions in the environment, people conversing are not facing each other, the talker and listener are far away from each other or they are not maintaining eye contact (Berry & Sanders, 1983; Erber, 1996). These factors make it difficult to focus on what is being said and extract the meaning accurately. Finally, the specific characteristics that the listener brings to the communication environment affect how much of the message is understood. For example, individuals with a hearing impairment receive an altered or distorted acoustic signal because the acoustic signal being produced is not sensed accurately by the impaired auditory system. This distorted signal can then lead to misunderstandings of the intended speech signal. Each of these components of the communication process and their potential for reducing speech intelligibility are elaborated in the following sections.

Talker

When the acoustic speech signal is distorted, difficulties understanding the message arise. Individuals speaking a second language, for example, have lower intelligibility scores than native speakers of that language (Rogers, Dalby & Nishi, 2004). When identifying content words spoken in sentences, listeners correctly identified more words spoken by native speakers (74%) than by non-native speakers with high speech intelligibility (55%) and listeners identified more words spoken by non-native speakers with high speech intelligibility than non-native speakers with low speech intelligibility (38%). Furthermore, non-native speakers with high speech intelligibility were most affected by background noise. They were similar to native speakers in quiet conditions, but at an SNR of – 5 dB, speech intelligibility scores were more similar to non-native speakers with low speech intelligibility. Results reported by Rogers et al. (2004)

suggested that the effect of the speech signal being distorted (e.g., a non-native speaker talking), is greater in a less than ideal listening environment (e.g., in background noise).

Motor speech disorders result from damage to components of the nervous system that control the production of speech. These include the neurons in the brain that plan motor movements and the neural pathways in the central and/or peripheral nervous systems to the target muscles. Dysarthria is a motor speech disorder characterized by imprecise articulation, diminished breath group lengths and stress patterns and/or abnormal voice and resonance. It results from damage to the neural centres and pathways that innervate muscular structures that produce speech, or damage to the muscles themselves (Hodge & Wellman, 1999).

Production of the speech signal.

At the signal generator component of the speech intelligibility process, problems with the talker's ability to generate a sound (sound source issues) and shape the vocal tract to filter this sound into the desired sound patterns of the language (vocal tract filter issues) can lead to distorted speech signals that are less readily understood by a listener. The talker's sound source can be defined as the place in the vocal tract anatomy that vibrates, or causes air turbulence, to create sound. For example, producing /s/ (e.g., sit) requires that the tongue be almost touching the alveolar ridge in the mouth to create a constriction in the vocal tract. This becomes a turbulent sound source when a pressurized air stream is forced through it. Producing a /z/ (zap) requires that the tongue be almost touching the alveolar ridge in the mouth to create a turbulent sound source, and that the vocal folds also vibrate to produce phonation (i.e., voice). In the latter example, the speech sound has two sound sources, noise produced at the constriction between the tongue and the upper alveolar ridge, and voice, produced at the level of the vocal folds. The quality and loudness of the speech signal are largely determined at the level of the sound source. The quality of the signal depends on the regularity of the oscillatory movements of the vocal folds to produce phonation. The loudness of the speech signal depends upon the adequacy of the respiratory system to generate sufficient positive pressure below the vocal folds and the adequacy of the vocal folds to valve the pressurized air stream flowing from the lungs. The greater the air pressure below the vocal folds, the louder the voice that is produced if the vocal folds valve the air flow efficiently. Breathing during speech requires a quick inspiration and a slow expiration that has enough air pressure to create vibrations and turbulent air flows, but not so much that air is lost very quickly and only a few words can be said. Control of the expiratory muscles is also required to place appropriate stress on words and syllables (Hodge & Wellman, 1999).

The vocal tract filter characteristic can be defined as the shape the vocal tract takes on during speech production. For example, different configurations and movements of the soft palate, pharynx, tongue, jaw and lips create different sizes and shapes of the vocal tract resonating cavities and therefore, different acoustic properties, and thus different sounds. Impairments in any of the muscles that are used to generate the sound sources for speech or the filter characteristics of the vocal tract can occur in neurological conditions such as cerebral palsy. These impairments may reduce intelligibility if they limit the child's ability to articulate speech sounds accurately, to generate and sustain adequate respiratory pressures for appropriately loud speech, stress contrasts and breath group lengths and/or to appropriately control the vocal folds to valve the pressurized air from the lungs for phonation of adequate quality, pitch and loudness (Hodge & Wellman, 1999).

Severity of speech disorder.

Muscular structures used for speech articulation include the soft palate, posterior, middle and tip of the tongue, the mandible, the lips, and for the /h/ sound, the vocal folds. Damage to the innervation of the muscles in these structures affects their actions for speech (e.g., not being able to bring the lips together to produce /b/, /p/ and /m/; difficulty getting the correct placement of the tongue for sounds including /s/, /t/ and /k/; limited movement for the tongue, jaw and lips for dynamic sounds like the diphthong /au/ in "h<u>ow</u>"; inadequate closure of the soft palate on high pressure consonants like /s/ so that air escapes through the velopharyngeal opening into the nasal cavity or on voiced oral consonants (e.g., /b/ and /d/) so that the nasal cavities act as a resonator and these consonants sound like their nasal counterparts (e.g., /m/ and /n/). Damage to the innervation of inspiratory and expiratory muscles can result in altered prosody by reducing the number of syllables that can be said in one breath group (known as shortened breath groups), not being able to place extra stress on the appropriate syllables or words to signal meaning and inappropriate loudness (e.g., too quiet). Damage to the innervation of the vocal folds can result in voice disorders where the individual has difficulty initiating phonation (vocal fold vibration for voiced consonants and vowels), reduced quality of phonation, reduced loudness control, and/or timing phonation for producing both voiceless and voiced sounds in the same word or phrase (Hodge & Wellman, 1999). The specific features of the speech disorder depend on the type and severity of dysarthria exhibited. This is determined by the location and extent of the damage in the nervous system that caused the dysarthria.

There are several locations in the nervous system that, when damaged, can result in dysarthria. These include the upper motor neurons that originate in the cortex, direct activation pathways from the cortex to lower motor neurons in the brainstem and spinal cord, indirect pathways from the cortex to the lower motor neurons via control circuits that include the basal ganglia, associated subcortical motor centres and the cerebellum, and the lower motor neurons themselves, including the muscle fibres they innervate (Hodge & Wellman, 1999). This damage can be manifested as weakness, slowness, limited range of motion, reduced coordination and/or extraneous movements of speech muscles. In *spastic dysarthria*, for example, bilateral damage along the activation pathways from the upper motor neurons in the cortex to the brain stem or spinal cord can result in spastic paralysis characterized by weakness, increased muscle tone, hyperactive reflexes and limited range and slowness of movements (Hodge & Wellman, 1999; Yorkston, Beukelman, Strand & Bell, 1999).

With regard to severity, a person with mild dysarthria may show some imprecise articulation of sounds that require use of a single articulator (e.g., lips) if only the muscles of that articulator are affected by the neural damage. On the other hand, a person with widespread neural damage may have all muscle systems affected, even to the extent that no speech can be produced (Hodge & Wellman, 1999). For example, if innervation to muscles of the articulators, respiratory system, and vocal folds are damaged, one might observe 1) imprecise articulation (e.g., difficulties producing many sounds due to reduced or altered innervation to the lips, tongue jaw and soft palate), 2) reduced ability to vocalize appropriately (e.g., producing voiced and voiceless sounds at appropriate times, initiating phonation and/or poor vocal quality), 3) poor breath support for speech (e.g., short breath groups, reduced ability to signal word stress and/or reduced loudness) and, 4) a resonance disorder if there is consistent inadequate closure of the velopharynx during speech.

Shriberg and Kwiatkowski (1982) developed a method to assess the severity of a speech disorder using a spontaneous speech sample and rating scales. The percent consonants-correct (PCC) measure involves identifying each spoken consonant as pronounced correctly or incorrectly when compared to an adult-like production of the intended sound. Number of consonants spoken correctly is divided by the total number of spoken consonants and multiplied by 100. Severity ratings are assigned as follows: 85-100% (mild), 65-85% (mild-moderate), 50-65% (moderate-severe) and less than 50% (severe). They found that the PCC measure was closely related to listener perceptions of severity as measured by a 9-point equal-appearing interval rating scale. The percent-vowels-correct (PVC) measure involves identifying each spoken vowel as pronounced correctly or incorrectly (Austin & Shriberg, 1997). Children with dysarthria may have both the production of consonants and vowels affected and so a combination of PCC and PVC may be a more representative measure of severity of their speech disorder. Speech intelligibility scores from spoken sentences can also be used to assess the severity of a speech disorder. Because these represent "connected" speech, they may be more sensitive to difficulties with prosody and breath support. The classification of speech disorder severity level (mild to severe), of the four children with spastic dysarthria whose word recordings were used in the current study, was based on the word identification scores of adult listeners with normal hearing identifying spoken sentences. These listeners did not know the children and had minimal experience listening to children with dysarthria.

Signal content.

The lexical content of the signal, or the actual words being said, can also affect how much of the signal the listener can understand. These factors include the frequency of the word(s), or how often each word is used in the language (Brown & Rubenstein, 1961) and how phonologically similar each word is to other words, or if it has many or few 'neighbours' (Bradlow & Pisoni, 1999). Signal content characteristics also include the type of stimuli being understood and the phonetic characteristics of the words.

A word's phonological neighbourhood[1] and word frequency can affect the degree to which they are identified. These characteristics of words are often measured differently in studies; however, the basic pattern is consistent. Phonological neighbourhood refers to how many other words are phonologically similar to the target word. Goldinger, Luce and Pisoni (1989) presented words (auditory signal only) to listeners in + 5 dB SNR. Words were organized by neighbourhood density[2]. Subjects were asked to type in the word they heard. Words with a sparse neighbourhood density (few other phonetically similar words) were identified more accurately than words with a dense neighbourhood density. Grainger, Muneaux, Farioli and Ziegler (2005) identified a word's phonological neighbours as the number of words that differ in only one phoneme, keeping the same number of phonemes in the word and respecting phoneme position. Words with less than eight neighbours were identified as having a sparse phonological neighbourhood and

---

[1] The terms phonological neighbourhood and phonetic neighbourhood are used throughout the literature as labels of variables measuring the same phenomenon.
[2] The measure used here was a frequency-weighted neighbourhood score, which includes a calculation of phonologically similar words and the word frequency of those phonologically similar words.

words with more than eight neighbours were identified as having a dense phonological neighbourhood.

In terms of word frequency, multiple researchers have identified an advantage for high frequency words over low frequency words (see Luce & Pisoni, 1998). Balota & Chumbley (1985) used written word frequency values from Kučera and Francis (1967). Words were identified as either high frequency (more than 36 occurrences per million) or low frequency (less than 7 occurrences per million) words. Printed words were presented to participants who were asked to pronounce the word. Participants were faster at naming high frequency words compared with low frequency words. Goldinger et al. (1989) found that high frequency words were identified more accurately than low frequency words when stimuli were presented auditorily and participants were asked to type the word they heard.

The length of utterance that is presented to a listener to identify (e.g., a single word, an entire sentence or a conversation) also affects speech intelligibility measures. In general, speech intelligibility scores tend to be higher for conversational speech samples and sentences than for words (Gordon-Brannan & Hodson, 2000). This difference can be attributed to the increase in syntactic and semantic context that comes with saying complete sentences. Yorkston and Beukelman (1978) found an interaction between understanding single words or whole sentences and intelligibility scores. For talkers with higher speech intelligibility scores, listeners understood more when given an entire sentence compared to when given single words. On the other hand, for talkers with lower speech intelligibility, listeners understood more when listening to single words compared to listening to entire sentences.

The specific speech sounds in the message and their relative intelligibility have been shown to interact with noise. Specifically, Benki & Felty (2005) found that consonants were more vulnerable to noise than vowels. Manner of articulation was especially vulnerable for bilabial consonants (e.g., confusing /f/ for /p/ or vice versa) and consonant deletion was highest for final nasals. With regard to the effects of noise on vowels, the tense/lax distinction was most vulnerable. Both consonants and vowels were identified correctly more often when they were in initial rather than final word position. Miller and Nicely (1955) created confusion matrices for consonants in noise. They found that the features most robust and least vulnerable to noise were voicing and nasality distinctions. These could still be perceived at - 12 dB SNR. Place distinctions, on the other hand, were vulnerable to noise and a + 6 dB SNR was required for accurate identification. Place of articulation is also the easiest feature to see. Miller and Nicely (1955) used only audition, therefore their results relate directly to speech intelligibility in noise and not to overall comprehensibility.

These results indicate that any word lists created for auditory identification should be balanced for word frequency, phonological neighbourhood and have the same consonants in the same word positions. When the order of word lists is randomized for each listener, these steps to equate the word lists will help ensure that any effects seen between word lists are due to talker-specific factors rather than differences between the lists.

Environment

Factors in the environment can contribute to an inaccurate message being received. These include the amount and type of distracters present, for example, visual distracters that result in the listener not looking at the talker's face. Background noise that masks the acoustic signal or distracts the listener can also negatively influence how much of the speech signal is understood. Since communication rarely takes place without some kind of interfering background noise, measures of speech intelligibility should include this factor if the results are to be generalized to performance in a real-life situation.

Cues.

What we see when we are looking at someone talking has large effects on what we hear (McGurk & MacDonald, 1976) and not looking at the person talking can decrease how much of the speech signal is understood. Monsen (1983) found that when listeners were asked to identify words spoken by talkers with a hearing impairment, word identification scores were 14% higher when the listener both saw the face and heard the speech sample of the talker compared with when listeners were only able to hear the speech sample. The environment can also provide cues about what the person is talking about. Listeners may be able to guess what the talker is talking about based on their gestures or eye gaze (e.g., pointing or looking at the object).

Signal-to-noise ratio.

A commonly used measure of the relationship between the acoustic speech signal and the background noise is signal-to-noise ratio or SNR. A positive SNR indicates that the signal of interest is at a more intense level than the background noise, while a negative SNR indicates that the signal is at a less intense level than the background noise. It is apparent then that when the signal is speech, a higher SNR should result in more accurate speech identification scores when a listener is asked to understand a spoken message.

Different types of background noise can differ in their impact on speech intelligibility. Understanding speech in a room full of other people talking, for example, may be more difficult than understanding speech in noise from nearby machinery that is at the same sound level. Larsby, Hällgren, Lyxell and Arlinger (2005) studied the effects of background noise on speech for 1) noise that was similar in temporal features to speech without real words, 2) noise that contained context like speech (e.g., real words), but not the temporal features (e.g., multi-talker babble), 3) noise that was another person talking and 4) no noise. They found that when the interfering background noise had temporal characteristics like those similar to speech, subjects performed worse and had longer reaction times when making decisions about a spoken word than either of the other noise conditions. This suggests that when the temporal characteristics of the background noise resemble speech, it makes understanding speech in noise the most difficult.

The level of background noise found in a typical classroom varies depending on many different factors. For example, the location of the room relative to loud areas in the school (e.g., a gymnasium) and relative to outside (e.g., near a loud highway) can have a notable effect on the average noise level in the classroom. The age of the children can also affect background noise levels, with younger children, in general, having louder classrooms than older children (Picard & Bradley, 2001). Jamieson, Kranjc, Yu and Hodgetts (2004) reported average noise levels found in classrooms to be anywhere from 44 to 94 dB A[3]. Ross (1992), along with other researchers, has identified the level of background noise in a traditional occupied classroom as being between 58 and 60 dB A

[3] The A-weighting acts like the sensitivity of the human ear. For example, the human ear is less sensitive to lower frequency sounds as compared to higher frequency sounds, so the A-weighting mimics this relationship (Crandell & Smaldino, 2002; Ross, 1992).

or 62 and 65 dB C[4] and the level of a teacher's voice in a classroom as being between 65 and 70 dB SPL[5] at three feet from the talker. Crandell, Smaldino and Flexer (1995) reported SNRs in classrooms as ranging from + 5 to – 7 dB. These estimates leave an expected SNR of – 7 to + 12 dB as reasonable measurements of the SNR for a teacher's voice in classrooms at three feet from the talker.

When identifying content words in sentences spoken by native speakers of English, listeners with normal hearing were 95% correct at + 10 dB SNR and only 30% correct at – 5 dB SNR with multitalker babble as the background noise (Rogers et al., 2004). At a + 12 dB SNR, with the background noise being a conglomeration of voices and clattering dishes in a cafeteria, normal hearing individuals could correctly discriminate among individual words spoken by a typical talker with 90% accuracy. At this same SNR, subjects with varying degrees of sensorineural hearing loss could correctly discriminate, on average, 66% of the words (Cooper & Cutts, 1971). Furthermore, scores decreased an average of 3.47 to 3.57% per dB HL that the noise was increased for hearing impaired and normal hearing subjects respectively, resulting in a predictable average performance for the specific word lists at varying levels of noise. However, the variability of these scores increased with lower SNR levels (Cooper & Cutts, 1971). When recognizing monosyllabic words in multitalker babble noise, it is possible for some individuals with mild sensorineural hearing loss to have comparable performance to normal hearing listeners. For example, when the speech signal was presented at an optimal intensity (e.g., 72 dB versus 60 dB SPL) and the SNR was 12 dB higher (+ 24 versus + 12 dB SNR), one individual achieved word identification scores

[4] The C-weighting measures the intensity of the sound equally across the frequencies, therefore, it responds better to lower frequency sounds than does the A-weighting.
[5] Sound Pressure Level

similar to normal hearing listeners (Dirks, Morgan & Dubno, 1982). These normal hearing individuals could recognize 50% of the words at an SNR of about – 10 dB with a speech signal at 60 dB SPL. For many hearing impaired individuals to achieve scores comparable to normal hearing individuals, the SNR needed to be greater by + 4 to + 12 dB, depending on the level of hearing loss (Crandell et al., 1995, Dirks et al., 1982).

The type of signal being presented can also affect the SNR required to correctly identify 50% of the signal. Specifically, an SNR of - 14 dB is required to correctly identify 50% of spoken digits, - 4 dB for identifying words in sentences and + 3 dB for identifying nonsense syllables (Miller, Heise, & Lichten, 1951). It can be expected that the SNR also needs to be greater for a listener to comprehend a talker with distorted speech, or reduced signal clarity. The distance from the talker to the listener and the reverberation time of a room, or how long it takes the signal to decrease by 60 dB, also affect the SNR (Crandell & Smaldino, 2002). Lower reverberation times (e.g., the sound is absorbed faster) result in increased speech intelligibility. The factors of SNR, reverberation time and distance will be held constant in the proposed study.

As classroom SNRs have been shown to range from – 7 dB to + 12 dB SNR, an SNR of + 10 dB is one that is typical of a more ideal classroom environment. It would be expected to result in highly accurate speech discrimination scores with a normal hearing listener and a highly intelligible talker, but would be expected to reduce accuracy and increase response time and perceived effort, when listening to talkers who have reduced speech intelligibility.

Listener

Listeners play an integral role in the communication process. They are affected by multiple factors including the nature of the task being performed, their listening capabilities and their engagement in the task.

Task.

A judging task where the listener rates how understandable the talker is on a scale can give different results compared to a task where the listener identifies the words spoken. A rating scale, for example, is a quick and easy way of judging how understandable a talker is. Word identification, on the other hand, takes more listening time. Furthermore, whether a listener is identifying words spoken singly, in sentences or in spontaneous speech also affects identification scores. Using a multiple choice format versus an open-set task to identify what is heard also affects scores (see Kent et al., 1994). Monsen (1983) found that when individuals listened to sentences spoken by talkers with hearing impairment and were given the context of the sentence (e.g., "There is no need to worry about theft" before the sentence "Our car is safe"), word identification scores increased by 14%. Hustad and Beukelman (2001) found that intelligibility scores for sentences spoken by adults with dysarthria increased when listeners were given extra cues. Mean intelligibility scores when given no cues was 18%. When listeners were given a topic cue (e.g. "The topic of this sentence is purchasing a new vehicle" before "Jason needed to buy a car"), mean scores increased by 10%. When listeners were given the first letter of each word in the sentence as the talker was saying it, mean scores increased by 18%. When listeners were given both cues together, mean scores increased by 33%. The complexity of the utterance, including its linguistic structure, length, familiarity and predictability also affect identification scores (Kent, 1992; Kent et al., 1994).

Capability.

The listener's ability to understand a spoken message is influenced by a number of factors including familiarity with the spoken language, hearing sensitivity and familiarity with the talker or others with a similar speech pattern. Listeners who are trying to understand speech in their non-native language have more difficulty than when the speech is in their native language (Bradlow & Pisoni, 1999). Listeners with a hearing impairment typically perform worse on speech identification tasks than those without hearing impairment (Larsby et al., 2005). For listeners with a hearing impairment, the signal must be sufficiently above their auditory threshold so they are able to extract the acoustic cues effectively. Finally, if listeners are familiar with the talker (e.g., a spouse or parent), with other talkers with a similar speech disorder or with the material the talker is saying, they are more accurate at identifying what the talker is saying (Beukelman & Yorkston, 1979; DePaul & Kent, 2000; Hustad & Cahill, 2003; Monsen, 1983).

Engagement.

Listeners' degree of attention, motivation and focus when trying to understand the speech signal can influence how accurately they identify what is being said. Intelligibility is often measured using rating scales or word identification scores, specifically, accurate identification of spoken words (Kent et al., 1989). However, measures of the effort that listeners invest in the process of making decisions about the identity of a word are also of interest. The mental effort that listeners expend in identifying spoken messages can also be measured. Downs (1982) reported that people trying to understand a distorted speech signal due to a hearing loss can achieve the same word identification accuracy scores when listening to a speaker as someone without a hearing loss, but this requires a great amount of effort and increased processing time. For a normal hearing listener, it is hypothesized that understanding the distorted speech of a child with a motor speech disorder will also require greater effort and longer processing times. Measures of listener effort are discussed in a later section on Measuring Effort (see p. 40).

Measuring Speech Intelligibility

Scaling

Speech intelligibility has been measured in a number of different ways. In one approach, interval scaling, listeners give a subjective rating of how much of the speech signal they can understand (e.g., 1 – "completely not understandable" to 5 – "completely understandable"), or how intelligible the speaker was (e.g., 1 – "essentially unintelligible" to 5 – "essentially intelligible") (Gordan-Brannan, & Hodson, 2000). Samar and Metz (1988) compared rating scales to methods where the listener writes down what they heard to measure speech intelligibility in adults with hearing impairment. The interval scaling method (e.g., 1 – "speech is completely unintelligible" to 5 – "speech is completely intelligible") provides an evaluation of speech intelligibility that is quick. However, there are questions about the reliability of interval scaling, especially for talkers with speech intelligibility that is in the midrange, in comparison with the write-down method (Samar & Metz, 1988). For example, 95% of talkers assigned a rating of 3 had actual word identification scores ranging between 25% and 90%.

Direct magnitude estimation (DME) has been shown to be a better measure of the speech intelligibility of hearing impaired individuals than the interval scaling method

(Schiavetti, Metz, & Sitler, 1981). In the DME scaling procedure, the listener rates the speech intelligibility of a talker by comparing the speech sample against a 'standard' stimulus using a ratio. For example, listeners would hear a speech sample from the 'standard' talker, be given a number that corresponds to that talker (or asked to assign a number on their own), then hear speech spoken by the target talker. They are then asked to rate the target talker compared to the 'standard' talker. For example, if the 'standard' talker was 100 and the target talker was half as intelligible, this ratio scaling procedure would result in the target talker being rated as 50. Ratings gained from DME depend greatly on how intelligible the 'standard' talker is (Weismer & Laures, 2002). DME is a better procedure for making judgments that are quantitative in nature (e.g., loudness), whereas standard interval rating scales are better for making judgments that are qualitative in nature (e.g., pitch) (Stevens, 1975). More reliable and precise procedures to measure speech intelligibility involve word identification.

Word Identification

In word identification tasks, listeners write down what they hear talkers say in a spontaneous speech sample, or prepared single word or sentence level stimuli. Listeners may also identify the talker's words using a closed-set task. Here, the listeners choose the word they hear spoken from a given set of words. The number of words identified correctly is measured by the match between the words in the intended message and the listener's responses. This number can then be divided by the total number of words spoken and multiplied by 100 to give a percent intelligibility score. Gordon-Brannan, and Hodson (2000) reported that identification scores for word, sentence, and spontaneous speech and subjective ratings correlated positively with each other for four to five year-old children with mild to severe phonological delay/disorder of unknown origin.

Word identification measures of speech intelligibility directly reflect how much of the message is understood by a listener and are the measures used most widely to judge the severity of a speech disorder (Bernthal & Bankson, 1998). Intelligibility scores can also be used to assess the need for treatment and measure treatment progress (Bernthal & Bankson, 1998; Schmidt, 1984). Tests such as the *Assessment of Intelligibility of Dysarthric Speech* (Yorkston & Beukelman, 1981) or the *Children's Speech Intelligibility Measure* (Wilcox & Morris, 1999) provide a percent intelligibility score for research and clinical purposes. They can be used to measure the severity of the speech disorder and change in intelligibility over time in a talker and compare these measures across talkers.

Improving Speech Intelligibility

Because factors that affect speech intelligibility include those related to the talker, environment and the listener, improving speech intelligibility for persons with dysarthria can involve talker, environmental and listener-specific approaches. For the talker, the first step in avoiding a communication breakdown is to use predictable sentence structure and grammar and maintain "clear speech". Instructions given to talkers about how to use clear speech include "speak as clearly as possible, as if [. . .] trying to communicate in a noisy environment", "enunciate consonants more carefully and with greater (vocal) effort" and telling the person to not slur the words together (Picheny, Durlach, & Braida, 1985). The acoustic correlates of using clear speech include factors that are related to the sound source and filter. Characteristics of clear speech include greater intensity and more precise articulation by releasing the stops in words and making vowels more distinct from each other, when speech samples are compared with conversational speech (Picheny, Durlach, & Braida, 1986). Using clear speech has been shown to increase speech intelligibility scores by 17% for persons with hearing loss when listening to a typical talker speak nonsense sentences (Picheny et al., 1985). Use of the clear speech strategy has been encouraged for use by individuals trying to talk with persons with hearing loss. Listeners' subjective reports indicated that the clear speech sentences were more intelligible, but they were tiring to listen to. Griffiths (1990) found that when listening to passages spoken in their second language, subjects comprehended the passage better if it was spoken at a slow rate (100 words per minute (wpm)) compared with a fast rate (200 wpm). Approaches that encourage speaking clearly can be effective for persons with reduced speech intelligibility. The expectation is that in attempting to increase the clarity of the spoken message, the talker will increase his or her effort, which will increase loudness and articulatory precision and improve the voice quality of the speech signal produced.

Due to the chronic nature of dysarthria, to maximize communication success for talkers with disorders such as cerebral palsy, emphasis must also be placed on factors outside of the speaker. These include altering the communication environment and listener behaviour. This can be accomplished by ensuring that the environment is ideal for spoken communication (e.g., being in a quiet place with a high SNR and with few interfering distractions) and by training listeners to encourage talkers to use behaviours that increase speech intelligibility (e.g., use clear speech techniques) and comprehensibility (e.g., reminding the talker to give the listener hints, or cues, about what they are talking about). If clear speech techniques cannot be used effectively so that communication breakdowns continue to occur, both participants of the conversation need to use strategies to either improve the environment (e.g., move closer together), use an effective repair strategy, or both (Berry & Sanders, 1983; Erber, 1996). This highlights the importance of training both talkers and listeners to recognize potential situations where communication may be difficult and to know about strategies to use that will be effective in reducing communication breakdowns and associated frustration.

Speech intelligibility scores can also be increased by the listener becoming more familiar with the speech of the talker. Markam and Hazan (2004) assessed speech intelligibility with typical talkers and listeners. They found that when three words were presented after a precursor sentence spoken by the same person (e.g., "and the next three words are: cat, dog, horse"), higher word intelligibility scores were obtained than when just presented with a single word[6]. Hustad & Cahill (2003) compared speech intelligibility scores for adults with mild to severe dysarthria secondary to cerebral palsy when listeners were presented with sentences spoken by the same talker for 50 consecutive minutes. They found higher intelligibility scores for the last quarter of the words (mean = 66%) when compared to the first quarter (mean = 55%). When McNaughton, Fallon, Tod, Weiner and Neisworth (1994) presented children and adults with words from speech synthesizers, they found that over five sessions, both children and adults improved their word identification scores. This improvement was 17.7% and 14.3% for one device, and 20.85% and 23.35% for another device for children and adults respectively. A possible explanation for these results is that as listeners become more

[6] A ceiling effect was found across subjects, so an analysis that took into account only the bottom quartile was done to look for the effect of familiarity

familiar with the individual talker's voice, they are better able to perceive the acoustic cues in the speech signal, resulting in higher speech intelligibility scores.

Both comprehensibility and speech intelligibility may be improved by the use of effective communication repair strategies. Repair strategies are behaviours that a person does when he or she is not understood and can include altering or increasing both the acoustic (speech) signal as well as providing other cues. Other cues that have been found to be beneficial for individuals with reduced speech intelligibility include using gestures to supplement speech, giving the listener a first letter cue and giving the listener a semantic cue to enhance understanding of the message (Beliveau, Hodge, & Hagler, 1995; Hanson, Yorkston & Beukelman, 2004). These additional cues, beyond those in the acoustic speech signal, help the listener predict and decode the message more easily. Communication breakdowns can also be repaired by using strategies that are specific to the acoustic signal. For example, repeating the message, revising what was said or increasing the loudness and precision of the acoustic signal may help to increase speech intelligibility after a spoken communication breakdown. The above strategies can be thought of as increasing the amount of information about the message that is available to the listener.

Speech Repair Strategies

Strategies and Their Use

When signal clarity is reduced by distorted speech production or adverse environmental factors so listeners cannot identify the signal content, a communication breakdown occurs. If the talker is aware of this breakdown, he or she may try to repair it. Common repair strategies include repeating the utterance and revising the utterance (e.g., rephrasing what was said) (Ansel, McNeil, Hunker, & Bless, 1983; Caissie & Wilson, 1995). The ability to use conversation repair strategies is an early developed skill. This is supported by research that suggests that typically developing children as young as 2 years old are aware of what is required of them if their listener does not understand what they just said and are able to use repair strategies such as repeating or revising their message (Gallagher, 1977; Tomasello, Farrar, & Dines, 1984). Alexander, Wetherby and Prizant (1994) described communication repair strategies as developing alongside intentional communication.

In general, research findings indicate that message repetition is used earlier and more frequently from a developmental standpoint and that strategies for repairing a communication breakdown become more diverse with age (Brinton, Fujiki, Loeb, & Winkler, 1986; Caissie & Wilson, 1995; Gallagher, 1977; Konefal & Fokes, 1984). When repeating the same spoken message, the expectation is that the second repetition will lead to more accurate comprehension by the listener. This preferential use of repetition as a repair strategy has also been shown in children with language impairment, autism spectrum disorder, children with hearing impairments and typically developing children (Brinton et al., 1986; Caissie & Wilson, 1995; Most, 2002; Tomasello et al., 1984; Volden, 2004). Tomasello et al. (1984) asked children "what do you want?" to signal a communication breakdown. They found that children either repeated the entire message, or repeated a portion of their message more than half of the time. In the study by Gallagher (1977), children in Brown's Stages I, II and III were asked "what?" by an adult to signal that they did not understand. The children responded with repetitions of the original utterance (20.6%) and revisions of their utterance (77.3%). Revisions were further broken down into repetitions where the child articulated a sound differently (16 – 45%), elaborated on (24 – 35%), reduced (17 – 33%), or substituted another word for (13 – 30%) something in their original utterance. With more advanced language proficiency, children also became more proficient in responding to the requests by using multiple strategies. It can be concluded that the repair strategy used with the greatest frequency by children at this stage of development is to repeat their utterance, either in whole or in part.

Volden (2004) found that children with autism spectrum disorder and typically developing children often changed the suprasegmental characteristics of the speech signal as measured by subjective increases in loudness, word emphasis, more precise articulation and a slower rate (see previous discussion on "Clear Speech", p. 21) when the examiner said "what?" to signal a communication breakdown. However, these children simply repeated the same words, or part of their message, instead of adding more information, most often in response to the examiner saying "what?". After the child's response to "what?" the examiner signaled the need for a repair a second time by saying, "I don't understand". This time, children responded by adding more information to their original spoken message and by using more gestures (e.g., pointing). However the results for suprasegmental and gestural characteristics were not reported separately by type of verbal repair strategy used.

Brinton et al. (1986), Ciocci and Baran (1998) and Most (2002) let the children know that they did not understand them by saying "huh?", then "what?", then "I don't understand" or "I didn't understand that" at predetermined times throughout a session. Brinton et al. (1986) found that typically developing children used repetition as their repair strategy the most and used this strategy in response to "huh?" more than "what?" and in response to "what?" more than "I didn't understand that". In the study by Most (2002), repair strategies of children with and without hearing impairment were assessed. Children with hearing impairment were divided into two groups; those with good speech intelligibility and those with poor speech intelligibility. Speech intelligibility was determined by unfamiliar listeners rating a speech sample (1 – "completely unintelligible speech" to 5 – "completely intelligible speech"). Children with lower speech ratings (1 or 2) were considered to have poor speech intelligibility and those with higher ratings (4 or 5) were considered to have good speech intelligibility. The children with poor speech intelligibility used the repetition strategy more than children with good speech intelligibility and children without hearing impairment, but repetition was the strategy used most by all three groups.

Caissie and Wilson (1995) studied the communication interactions of children ages 9 – 12 years with hearing impairment before and after intervention to teach different repair strategies in times of communication breakdown. They found that more communication breakdowns occurred when the child with hearing impairment was the listener (i.e., they misperceived a message spoken by a typically hearing peer) compared to when the child with hearing impairment was the talker (i.e., the peers in the group did not understand the speech of the child with hearing impairment). When a listener signaled that he or she did not understand, the most common repair strategy used by both hearing impaired and typical hearing children was to repeat all or part of the original message.

These results on the use of repetition as a repair strategy by children with hearing impairment, however, are inconclusive as other researchers have found that these children use other types of repair strategies more than repetition when asked for clarification in the same way (Ciocci & Baran, 1998; Most, 2002). Givens and Greenfield (1982) found that children with hearing impairment often used nonlinguistic information, such as facial expression, in their repairs when compared to children without hearing impairment. Ciocci and Baran (1998) found that children with hearing impairment revised their message most often and children without hearing impairment were as likely to revise their message as to repeat it. One possible difference between two of these studies is how intelligible the children were. In the study by Ciocci and Baran (1998), the intelligibility of the children with hearing impairment was not reported. However, if the children in this study had highly intelligible speech with exceptional linguistic abilities, the results may support the notion that repetition is used less with more advanced individuals (Brinton et al., 1986; Most, 2002). None of the studies of children's repair strategies assessed the degree to which these repair strategies were effective in increasing speech intelligibility because the repairs were artificially elicited. That is, examiners signaled that a communication breakdown had occurred when in fact, this was not the case.

The most common repair strategy used by adults with dysarthria secondary to cerebral palsy is 'say it again' (Ansel et al., 1983). Ansel et al. (1983) found that when not understood, speakers with dysarthria secondary to cerebral palsy typically repeated all or part of their previous utterance (44%) or repeated all or part of their previous utterance with elaboration (19%). Ansel and her colleagues also found that these speakers did not increase the intensity of the signal, nor did they use linguistically helpful adjustments like giving first letter cues to aid their listener. This study also did not look at the degree to which these repair strategies (e.g., repetition) increased speech intelligibility (B. M. Ansel, personal communication, December 8, 2005).

The nature of the request to clarify when a communication breakdown has occurred influences the response given to repair the breakdown. Requests that are specific, for example, "Did you say the green shirt?" may lead to more effective repair strategies being used than requests that are nonspecific, for example, "huh?" or "what?" (Brinton et al., 1986). Specific requests have also been identified as more favorable to listeners on three pairs of descriptors[7] (Gagne, Stelmacovich & Yovetich, 1991). These specific types of requests suggest that the listener has understood something of what was said and only needs help with a portion of the message. Researchers have found that when listeners ask for clarification by saying "huh?" or "what?", repetitions of the original message are more frequent than when requests were made more specific (e.g., "tell me another way" or "I didn't understand that") (Brinton et al., 1986; Volden, 2002). Nonspecific requests however, are used much more by children when they don't understand (Caissie & Wilson, 1995). Therefore, the likelihood of simply repeating the original message is very high, especially when children are talking to each other. An important area for investigation is to determine the effectiveness of responses that are frequently elicited and used by children in times of a communication breakdown.

Effectiveness of Strategies

Although frequency of use of communication repair strategies is well researched on adults with dysarthria, children with hearing impairment, children with language

[7] Participants viewed skits of a "hearing impaired" individual interacting with another person. The "hearing impaired" individual requested clarification by using specific or non-specific requests in different skits and viewers rated how they would feel if they were interacting with that person (e.g., energetic-tired; refreshed-fatigued; pleased-annoyed).

delay, children with autism spectrum disorder and typically developing children, the degree to which these strategies actually increase speech intelligibility has not been as thoroughly studied. The studies described previously used contrived situations to elicit a repair by creating a communication breakdown at systematic moments in the conversation, regardless of whether the listener understood the spoken message or not. This design does not lend itself to measures of how much more was understood after the repair.

When typical hearing peers revised their utterance, however, children with hearing impairment understood the message more than when the original message was repeated (78% and 58% respectively) (Caissie & Wilson, 1995). However, this was not the case when the children with hearing impairment were the talkers. Their strategies of revision and repetition had comparable success in repairing the communication breakdown (80% and 87% respectively) for their peer listeners with normal hearing. Additional acoustic adjustments, for example, increasing intensity and articulatory precision, were not measured in this study and the authors suggested that the children with hearing impairment may have made these adjustments when they repeated their message a second time. Therefore, it is possible that when listeners experience communication breakdowns because of an impaired auditory system, more effective repair strategies include talkers elaborating or explaining the original message compared to simply repeating the entire, or a portion of the message. When other options are available, for example, when the quality of the speech signal can be improved by the talker (e.g., increasing loudness or articulatory precision), either strategy, that is revising or repeating the message, is effective in repairing the communication breakdown. These data on children with hearing impairment were gained during a group discussion with interfering background noise and distracters (Caissie & Wilson, 1995). This is consistent with how these children would interact on a daily basis rather than in an interaction with a researcher generating contrived communication breakdowns in an ideal listening environment. Regardless, repeating the message was the strategy used most frequently by the talker and usually resulted in an increase in comprehension by the listener. Communication repairs using repetition resulted in 58% or 87% of the messages being understood successfully when the hearing impaired individual was the listener or talker, respectively.

TyeMurray, Purdy, Woodworth and Tyler (1990) presented video images of talkers with no sound to subjects to assess lip-reading ability by only presenting the video signal to subjects. Subjects were divided into six groups, five experimental and one control. Talkers said a sentence and subjects watching were asked to say the sentence aloud. In the experimental groups, if the subject did not repeat the sentence exactly, the talker performed a repair strategy based on their assigned group (repeated, simplified, rephrased, said an important keyword or said two sentences to give more information about the target sentence), then repeated the sentence again. Subjects were asked to say the sentence aloud again after the last repetition. Subjects in the control group also saw a talker saying a sentence and were asked to say the sentence aloud. If they did not repeat the sentence exactly, the talker simply said the sentence again and the subject was asked to say it aloud again. Therefore, when subjects in the control group did not get the sentence correct, they saw it repeated once. Subjects were more accurate in identifying the words in the sentence at the second presentation, compared with the first presentation, and all experimental groups performed better than the control group. The researchers also found that the repair strategies (repetition, simplification, rephrasing, saying an important keyword or saying two sentences to give more information about the target sentence) were equally effective. This suggests that in the visual modality, presenting the same sentence three times (one experimental group) leads to more accurate lip-reading (11% increase in words identified correctly for words misunderstood the first time) than presenting the sentence twice (the control group) (6% increase in words correctly identified), and repeating is equally as effective as simplifying, rephrasing, stating a keyword and saying two sentences. This same effect of a second repetition on increasing understanding has not been found with single words using vision only (Gagne & Wyllie, 1989). In this study, participants viewed video tapes, without sound, of talkers saying single words. Participants were asked to identify the word by lip-reading. If the participant was incorrect, they viewed one of three repair strategies (repetition, synonym or paraphrase). These researchers found that saying a synonym of or paraphrasing the original word was more effective in increasing the percentage of words correctly identified than repeating the original word.

Mechanisms to Account for the Effectiveness of Repetition as a Repair Strategy

As illustrated in the preceding section, repeating the spoken message when it has not been understood is a commonly used strategy. The basis for the positive effect of repeated presentations of the same stimuli on performance success, when participants are engaged in goal-directed behaviours, has been described both behaviourally and physiologically. Repetition priming "refers to enhanced or biased performance with repeated presentation of a stimulus" (Bergerbest, Ghahremani, & Gabrieli, 2004, p. 966).

Behavioural Evidence

Participants' behaviour on goal-directed tasks can be measured in terms of accuracy and response time and, in turn, these measures can provide evidence for the effectiveness of repetition of the stimulus in improving performance on the task. Monsen (1983) found a significant increase in speech intelligibility, from 63% to 70%, when listeners were able to listen to sentences, spoken by 10 talkers with hearing impairment, a second time immediately following the first presentation. These listeners were inexperienced in the sense that they did not have regular contact with the speech of individuals with hearing impairment. This analysis included only six participants in each listening condition, so generalization of the results is limited. Six participants heard the sentences spoken once and wrote down what they heard and six participants heard the sentences spoken twice and then wrote down what they heard. Hodge, Spooner and Wellman (1999) also examined the effect of one versus two presentations of sentences, spoken by 12 children with a wide range of severity of dysarthria, on 36 experienced listeners' word identification scores and found that speech intelligibility scores increased on average by 4% when comparing scores from the first and second presentations. Peng, Spencer and Tomblin (2004) also found that when listeners heard sentences spoken by an individual with a cochlear implant a second time, they were able identify 3.7% more words and rated the talkers as more intelligible on a scale from 1 ("not intelligible at all") to 5 ("totally intelligible"). McNaughton et al. (1994) presented children and adults with words from speech synthesizers. Participants had on average 5.6% higher word identification scores for words that had been presented in previous sessions compared with novel words. These studies suggest that a second presentation of an auditory speech or speech-like signal can increase speech intelligibility scores by 4 to 7%.

At the second presentation, response times for lexical decision-making are usually faster and responses are more accurate when making decisions about the stimulus (Bergerbest et al., 2004[8]; Holcomb, Anderson, Grainger, 2005[9]). Ghahremani (2005) and Rugg (1985) measured reaction time in a lexical decision task that required participants to decide if each stimulus was a word or not. In the study by Ghahremani (2005) individuals responded faster to the second presentation of visual and auditory stimuli compared to the first. In the study by Rugg (1985) individuals responded faster to the second presentation of a visual stimulus compared to the first (an average of 45 milliseconds (ms) faster). Mimura, Verfaellie, and Milberg (1997) studied the effect of presenting the same spoken words and nonwords twice to participants who had to decide if each was a word or not. Participants were consistently faster at identifying the second presentation of words even when there were up to eight intervening stimuli. However, the benefit of repetition was greatest when the repeated presentations were closer together (e.g., immediate repetition). Nonwords also showed an effect of repetition. Reaction times were faster for the second presentation of the stimuli, especially when the second presentation directly followed the first.

Simply repeating what one has just done might be considered the easiest way to attempt to accomplish a goal when one has not been successful the first time. In the case where a talker has been unsuccessful in getting a message across, repeating what has just

[8] Participants had to decide whether an environmental sound was made by an animal or not and responses to initial presentations of the sound were compared to repeated presentations of the sound.
[9] Participants were given a visual stimulus (e.g., printed word) as the prime to a spoken word and asked to decide whether the spoken word was a real word or not.

been said, in the same way, does not require extensive cognitive processing. The talker does not need to think of a different way to say it or use additional physical effort to make the message louder and with more precise articulation. Repetition also can be related to principles of human behaviour. Specifically, a behaviour that results in a desirable outcome is positively reinforced. Therefore, if a communication breakdown occurs and the talker repeats the same message (arguably the easiest thing to do) and this results in the message being understood, it is more likely that this behavior will occur the next time the same problem is encountered (Carlson, Buskist, Enzle, Heth, & 2000). If it is not an effective strategy, and therefore does not result in the problem being resolved, the behaviour is less likely to occur the next time the problem is encountered, that is, at the next communication breakdown. The talker may instead try a different strategy. However, the mental and physical effort associated with other repair strategies also plays a role, such that the talker may continue to repeat the message even if it is not as effective as doing something else if the former takes less effort.

Neurophysiological Evidence

Bergerbest et al. (2004) found reduced activation in the auditory cortex in a functional Magnetic Resonance Imaging (fMRI) study when participants heard the same stimuli later in the experiment in a task that required them to make decisions about whether each stimulus was an animal or non-animal noise. Ghahremani (2005) studied repetition priming in an fMRI study where participants completed a visual lexical decision task. Results indicated reduced activation in the occipito-temporal regions and left prefrontal regions when low frequency words were presented a second time, compared to the initial presentation.

Deacon, Dynowska, Ritter and Grose-Fifer (2004) studied the effect of repetition priming using Event Related Potentials (ERP) measured by electroencephalogram (EEG). The N400 ERP is a negative potential that occurs between 260 and 500 ms after a stimulus is presented and is thought to respond to word processing[10]. Deacon et al. (2004) found that the amplitude of the N400 was attenuated at the second presentation of the stimulus when participants were presented with the same nonword strings of letters twice in a row. Doyle, Rugg and Wells (1996) also found an effect of repetition priming when participants were asked to press a button if the presented string of letters (words and pronounceable nonwords) was not a real English word. When participants had seen the letter string earlier in the experiment, the negative electrical activation in the brain was attenuated compared with the first presentation of the letter string. A difference in electrical activity was also found after 400 ms when formal priming (i.e., when part of the letter string was presented previously – e.g., scan-scandal) was compared with repetition priming. The repetition priming condition showed greater attenuation than the formal priming condition. Therefore, seeing the exact same stimulus a second time creates neurophysiological changes, even when compared with other types of priming. These studies suggest that when a stimulus is repeated, there are changes at the neurophysiological level, in both blood oxygen level and electrical activity, that recognize the stimulus differently and perhaps process it differently at some level during the visual word recognition process (e.g., anywhere from visual form to semantic access). The nature of these differences is characterized by less excitation in some areas of the brain. Analogously, it might be hypothesized that hearing the same word spoken again

[10] This has typically been thought to be responding at the semantic level (Kutas & Hillyard, 1984; Bentin, Kutas & Hillyard, 1993). However, it may also be responding at the orthographic or phonological level (Deacon et al., 2004).

would result in neurophysiological differences in the auditory word recognition process, compared to the first presentation.

Speech Perception Theory

The Fuzzy Logic Model of Speech Perception (FLMP) is based on the premise that many sources of information (e.g., acoustic, visual, semantic, syntactic, etc.) converge to result in the accurate perception of a speech signal (Massaro, 1989). The information is evaluated, the different types of information are integrated and a decision is made about the meaning of the signal (Massaro, 1989; Pickett, 1999). The FLMP model might also explain the beneficial effect of hearing a spoken message repeated on listeners' understanding. At the first presentation, evidence accumulates about the identity of the speech signal based on the phonemes present. If the signal is not understood, a communication repair strategy may be used by the talker (e.g., repetition). At the second presentation of the signal, more acoustic information may be extracted because the listener can focus attention to particular parts of the signal that were not understood the first time. The listener has another opportunity to evaluate and integrate the information after the second presentation of the acoustic signal with that obtained from the first, to lead to a final decision.

It is not uncommon for researchers to include multiple presentations of the speech signal in studies assessing speech intelligibility. However, the effect that this repetition has on speech intelligibility scores has not been fully studied. Gordon-Brannan and Hodson (2000) allowed listeners to listen to spoken sentences, single words and spontaneous speech samples up to three times before entering their final word identification responses. The authors stated that this strategy places less emphasis on short-term memory than having only one presentation of the speech stimulus. In the *Assessment of Intelligibility of Dysarthric Speech* test (Yorkston & Beukelman, 1981), listeners can hear sentences twice but single words only once before they make their final identification responses. This arbitrary, standard procedure controls for the influence of different numbers of presentations on listeners' word identification. The underlying assumption may relate to the FLMP model. Specifically, somehow the listeners are able to extract different or additional information from the sentence the second time they listen to it, possibly leading to higher identification scores.

Memory systems theory.

It is widely accepted that there are different kinds of stores for memory. These can be divided by time. The sensory store for auditory information (a.k.a. echoic memory or auditory sensory memory) is the shortest of these. It has been referred to as "a brief memory system that receives auditory stimuli and preserves them for some amount of time" (Ashcraft, 2002, p. 111). This system temporarily stores auditory information until other mental processes can access it. The duration of this memory store can last up to 10 seconds, depending upon the type of stimuli being heard (Ashcraft, 2002; Baddeley, 1976). For digits or letters, for example, echoic memory begins to decline after about four seconds (Darwin, Turvey, & Crowder, 1972).

Short-term memory is the next step in memory storage. Information can be moved into this storage system if it is attended to or rehearsed. This store can last for up to 60 seconds (Baddeley, 1976; Bostrom & Waldhart, 1988). The last basic store for memory is long-term memory. Whether or not a stimulus is placed into long-term memory or lost may be determined by whether it is rehearsed and can be organized in a person's knowledge.

The effect of repetition priming can be explained within a memory framework. When a stimulus is presented again immediately following an initial presentation, participants respond faster and/or more accurately to the second presentation. This immediate repetition effect has been found for both words and nonwords, but nonwords may show a stronger effect (Mimura et al., 1997). When delayed repetition is used by adding up to eight intervening stimuli between the first and second presentation, the effect of repetition for nonwords is drastically reduced[11]. Repetition priming, therefore may be working at the levels of echoic, short-term and even long-term memory. If a stimulus is presented again immediately after it is first presented, the echoic memory representation will be strengthened, regardless of whether the word is unknown (e.g., a nonword, or a word that is not understood or not mapped onto a known word in the participant's mental dictionary) or known. If the stimulus is known (e.g., a real word that is understood), the stimulus may move into short-term memory or even long-term memory. The next time the same stimulus is presented, the participant responds differently to it. In the current study, it was expected that listeners will type their word identification responses fast enough for the repeated presentations to be within the span of echoic memory (i.e., shorter than 10 seconds).

To date we have located no studies that have empirically tested the hypothesis that repeatedly presenting the acoustic signal of a single word will alter the accuracy of a listener's identification of the word. The current study looked at what effect signal dose,

[11] With no intervening stimuli, the size of the effect is 324 ms; one intervening stimulus, 140 ms; four intervening stimuli, 91 ms; eight intervening stimuli, 110 ms.

defined as number of presentations of the same acoustic signal of a single word, has on speech intelligibility scores measured by accuracy of word identification. In the case of a single word, memory effects are expected to be minimized as an explanation for any changes in word identification scores that are observed.

Measuring Effort

In considering how to measure the effort that listeners put into a communicative interaction, it is useful to review the kinds of measures that have been used to analyze effort, physical and/or mental, expended during goal-directed behaviour in other domains, as well as spoken communication. Measures of effort can be subjective (e.g., rating scales, where an individual provides a measure of their self-perception) or objective (e.g., reaction time/response time, changes in neurological activation, where a physical measure of the individual's behaviour is obtained).

Physical Effort

Physical effort has been measured by using subjective rating scales, including DME, as well as physical measures such as muscle flexion, extension variance and pressure exerted. In an ergonomics intervention study, Laing et al. (2005) asked workers to rate their perceived effort when completing tasks before and after intervention. Workers were asked to indicate this on a 10-point scale (e.g., "How hard or tiring is work at this station on your back"). Ogata et al. (2004) used magnitude estimation (or DME) to assess the mental effort, physical effort and total work[12] that nurses contributed during visits with home health patients. Nurses were given a modulus (50 associated with giving a foot bath) and asked to indicate values in proportion to the modulus for each of the four dimensions after each visit. DME for these measures was identified as a reliable measure

[12] Definition includes aspects of the time, total work, physical effort and mental effort.

of the work of nursing services in an earlier study (Ogata et al., 2000). Participants were asked to rate mental effort, physical effort, total work and the time spent on each home care service when given a profile of hypothetical patients. Re-testing was done 7 to 9 months after initial testing. Correlation coefficients ranged from .82 to .96 for the four above measures of interest. Collins (1998) investigated how consistent patients with spinal conditions were in the effort they contributed to each trial of the task (trunk-extension-flexion at 60° per second on a Cybex device). They used the average points of variance, or how consistent the individuals were, as a measure of effort, or motivation, to perform the task.

Measures of physical effort when talking have been reported for individuals with spinal cord injury (Reiger, 2002). In this study, workload, or effort, was measured using speech-breathing (e.g., mean volume for pre-speech inspiration), speech-production (e.g., average number of syllables per breath group) and kinematic (e.g., rib cage contribution to lung volume exchange) variables. In a study by McHenry, Whatman and Pou (2002), listeners rated on a 5-point scale the effort (e.g., "sounds effortful") expended by an individual with spastic dysarthria while speaking before and after Botox injection to muscles of the vocal folds. Speech recordings were judged to be less effortful following Botox injection when compared to pre-injection. Solomon and Robin (2005) assessed effort in adults with Parkinson's Disease (PD) and neurologically normal adults. Participants were asked to use DME to rate their perceptions of effort in activities of daily living and while speaking. They were given a modulus of 100, representing "no particular effort". They also measured percent effort, by asking participants to squeeze air-filled bulbs with a given amount of percent effort (10 to 100%) and constant effort, by asking participant to squeeze with a constant effort and measuring decay over time, for both the tongue and hand grip. Researchers found that DME scores for participants with PD were higher than neurologically normal participants (higher ratings of daily and speaking effort) and the pressure dropped faster when asked to produce constant effort. These results were expected given the neurological basis of PD.

Mental Effort

Mental effort can be examined in terms of the mental energy that is being required to attain a certain goal. Response time, or the time elapsed from presenting the stimulus to when the participant responds, is a measure of how quickly information can be processed. Number of errors made is also a measure of mental effort. Tasks that are harder to complete typically result in longer reaction times and more errors. For example, reaction times have been used as a physical measure of mental processing in lexical decision tasks where participants are presented with strings of letters (typically half real words and half nonwords) and asked to decide if the letter string is a real word or not. Participants respond more slowly and make more errors when the nonwords are phonologically identical to real words, but differ orthographically (e.g., brane) when compared to the condition where the nonwords are pronounceable, i.e., follow the phonotactic characteristics of English, but do not sound like a real word (e.g., brap). The same effect (i.e., slower response times and more errors) is seen when comparing this latter condition with conditions where the nonwords are very distinct from real words, i.e., are not pronounceable (e.g., frts) (Stone & Van Orden, 1993). These results suggest that more mental effort is required to process the stimuli as the task requires more processing, i.e., becomes more difficult.

In terms of listening, measures of the mental effort that listeners expend can be quantified objectively as well as subjectively. For example, how long it takes listeners to decide what word they heard (i.e., reaction time/response time) provides an objective measure of their mental processing and therefore may serve as a correlate of listener effort. Larsby et al. (2005) found that reaction times were longer when subjects had to make a decision based on a spoken word in background noise compared to when there was no background noise. Participants with hearing impairment also had longer reaction times than those with normal hearing in both noise and quiet conditions. Larsby (personal communication, November 8, 2005) indicated that in this study, measures of reaction time frequently correlated negatively with measures of speech intelligibility.

Mackersie, Neuman and Levitt (1999) examined reaction time and percent words correct as listeners identified a spoken word using a closed-set task. Words were presented in varying levels of background noise (from -3 to +12 dB SNR). The authors found that both reaction time, as measured from the beginning of the item being presented to the response, and percent words correct changed with differing levels of SNR. Specifically, their subjects exhibited longer reaction times and fewer correct responses with lower SNRs. Pratt (1981) measured the percentage of words selected correctly in a 6-item multiple choice format and reaction time when listeners listened to and identified single words presented auditorily. Audio recordings were collected using different microphones to see the relative benefit of each. Results indicated that reaction time was more sensitive to differences between microphones than was percentage of words identified correctly.

Munro and Derwing (1995b) found that listeners' processing times for deciding whether a spoken sentence was true or false were longer (an average of 62 ms slower) when talkers were nonnative speakers of English compared to native talkers. Reaction time was measured from the end of the spoken sentence to the listener's button press to indicate true/false. This difference in response time was found even though response times were only used if the listeners verified the sentence correctly (correct in their true/false choice) and wrote the sentence down correctly. The researchers also found that ratings of how difficult the talkers were to understand (1 – "not difficult to understand at all" to 9 – "very difficult to understand") were related to reaction times. Specifically, speech from talkers identified as difficult to understand (7, 8, or 9 on the rating scale) resulted in listeners having longer reaction times than did other talkers.

Munro and Derwing (1995a) asked listeners to rate the understandability of talkers speaking English as a Second Language and native speakers of English. The rating scale used was 1 ("extremely easy to understand") to 9 ("impossible to understand"). In general, native speakers were rated as easier to understand than non-native speakers and scores were positively skewed. For 83% of the listeners, there was a significant negative correlation between their subjective ratings of understandability and their percentage of words identified correctly for spontaneous speech sample[13]. The authors suggest that the non-perfect relationship between percentage of words identified correctly (intelligibility) and the ratings of ease of understandability may be due to differences in processing difficulty. The listeners may have been able to understand a specific talker and therefore, had high intelligibility scores, but may have found it very difficult, and therefore, rated the talker as more difficult to understand. At a later date, the

[13] For these 15 listeners, however, there was a large range in correlations (- .44 to - . 90).

same listeners were asked to rate the amount of accent the talkers had from 1 ("no foreign accent") to 9 ("very strong foreign accent"). A significant positive correlation between accent scores and word identification scores was found for only 28% of listeners.

Preminger and Van Tasell (1995) reported correlations between ratings of intelligibility (defined as clarity) and ratings of effort ranging from .93 to .95 when listeners were asked to rate 20 passages spoken by one talker. Stimuli were assigned a rating between 0 and 100 for each dimension[14]. Whitehill and Wong (2006) had 20 listeners identify words spoken in 99 sentences, three spoken by each talker, an adult with dysarthria (total number of talkers = 33). After hearing each talker, listeners indicated the effort required to understand that talker on a 10 cm visual analog scale from "no effort required" to "maximum effort required". Each talkers score was averaged across the 20 listeners, resulting in a large spread in word identification scores and effort ratings. The correlation between word identification scores and effort ratings was - .95.

Hicks and Tharpe (2002) asked children between 5 and 11 years old with and without hearing impairment to complete two tasks: repeat words presented in varying levels of SNR and press a button when a light came on. They found that children with hearing impairment had longer reaction times to the light than did children with normal hearing when performing both tasks at the same time, suggesting that the task of repeating words required more mental processing for children with an impaired auditory system. Rakerd, Seitz and Whearty (1996) gave older adults with hearing impairment and younger adults with normal hearing a dual task to complete. Participants had to keep digits in memory as they listened to either a spoken passage that they would later have to

[14] A high correlation between judgments of clarity and sentence intelligibility scores has been found previously.

answer questions about, or noise. Subjects forgot more digits when listening to the speech condition compared to the noise condition, supporting the effectiveness of the dual task (i.e., when more effort is required elsewhere, fewer digits can be remembered). The older adults recalled fewer digits and were more affected by the speech condition compared to the noise condition than were younger adults. In a dual processing task, Downs and Crum (1978) asked adults to say the second half of a spondee word after being presented with the first half and to respond to the presentation of a light by pressing a button. When the auditory signal was presented in quiet, participants responded to the light faster than when there was intervening background noise (+ 6 dB SNR). Participants were also asked to rate the effort required to complete the task on a scale from 1 ("very easy") to 7 ("very hard"). Results were not significantly different between the noisy and quiet conditions (Downs & Crum, 1978). While reaction time differed between noise and quiet in the primary task when a second task was added, listeners did not reflect this difference in their ratings, suggesting that this type of rating scale may not be a sensitive measure of effort when the task is relatively easy.

Rating how much effort listeners think it takes to understand the speech signal provides a subjective evaluation of listener behaviour. The amount of effort listeners perceive it takes to perform a task has typically been measured using ordinal scales that have accompanying descriptors (Feltz, McClure, & O'Hare, 2002; Larsby et al., 2005). Numbers are displayed and a description is given (e.g., 1 – "no effort required" to 5 – "maximal effort" for Feltz et al., 2002; 0 – "none at all" to 10 – "extremely great" for Larsby et al., 2005) and participants are instructed to circle the appropriate number. A preliminary study of listeners' subjective evaluation of effort was analyzed and compared to speech intelligibility scores by Feltz et al. (2002). Listeners were asked to rate how much effort it took to understand words spoken by six young children with typical speech production. The correlation between listener effort and word identification measures of speech intelligibility was -.62 ($p = .13$), such that more effort was required to understand children with less intelligible speech. McHenry et al. (2002) asked listeners to rate their own perceived listener burden (e.g., “hard to listen to”) on a 5-point scale when listening to the speech of an individual with spastic dysarthria. Listeners reported less burden for speech recordings of the talker made following Botox injection when compared to pre-injection.

Larsby et al. (2005) also studied the amount of effort that listeners perceived when trying to make decisions based on a spoken word. They found that listeners reported that it took more effort when listening to words with interfering background noise. Their subjects with hearing impairment reported that listening took more effort compared to their subjects with normal hearing, in both noise and quiet conditions. Preminger and Van Tasell (1995) asked listeners to rate on a scale from 0 to 100 the effort it took to understand distorted speech and to estimate the percent of words they understood. They found that measures of effort and words understood were highly correlated (e.g., $r = -.93$ and $r = -.95$ in several conditions that varied the frequency content of the stimuli), such that more effort was required when fewer words were understood. The notion that more effort is required when trying to understand a distorted speech signal is further supported by researchers who found that interfering background noise increased the effort required to perform the task (Crandell & Smaldino, 2002; Downs, 1982). Hicks and Tharpe (2002) asked children with normal hearing and hearing impairment to rate how much effort it took to repeat words presented in SNRs of quiet, + 20 dB, + 15 dB and + 10 dB, on a scale from 1 ("not hard at all") to 5 ("very hard"). They found a trend for higher ratings with poorer SNRs, but this was not significant. This result suggests that this kind of rating scale may not be sensitive enough to pick up differences in effort at favourable SNRs. Humes et al. (1999) asked listeners with hearing impairment to rate the "ease of listening" when listening to speech in quiet and in noise using no hearing aids and two different kinds of hearing aids. Listeners indicated that it was easier to listen in the quiet conditions when using hearing aids compared to no hearing aids, signifying that an enhanced signal can decrease the amount of effort required.

These studies provide evidence for the detrimental effects of a distorted speech signal produced by the talker, interfering background noise or competing tasks, or reduced auditory acuity, on processing the speech signal. Furthermore, they demonstrate this effect on measures of accuracy, response time and perceived effort for understanding speech at SNRs that are relatively favourable (e.g., + 10 dB).

Southwood (1996) asked listeners to judge the 'bizarreness' and 'naturalness' of sentences spoken by adults with amyotrophic lateral sclerosis (ALS), some of whom had dysarthria. The purpose of her study was to determine whether these two phenomena, bizarreness and naturalness, are better measured using interval scaling or DME. In each of four sessions, listeners were asked to rate either bizarreness or naturalness using an interval scale or using DME. For the interval scale section, listeners were asked to write a number between 1 and 7 for each speech sample. For example, on the naturalness scale, if the speech sample sounded unnatural, participants were to write "1" and if it sounded highly natural, participants were to write "7". For the DME, listeners heard a sentence recorded from a speaker with dysarthria who was approximately 50% intelligible, based on scores on spoken sentences. Listeners assigned a number to this standard stimulus (modulus) and then rated the talkers, based on this stimulus. Results for interval scaling and DME scores were plotted against each other for both bizarreness and naturalness ratings. The results demonstrated that both bizarreness and naturalness are curvilinear phenomena, suggesting that these dimensions may be prothetic. Therefore, DME may be more appropriate for making these types of judgments than interval scaling (see Stevens, 1975).

Results from the previously mentioned studies suggest that DME may also be a more appropriate measure of listener behaviour (e.g., perceived effort) than interval scaling. Since DME can reveal if the phenomenon of perceived effort, when attempting to understand spoken words, has a linear or nonlinear relationship with word identification scores, it appeared advisable to use it rather than interval scaling, which assumes a linear relationship.

Ratings of effort can be evaluated based on how reliable listeners are when identifying the same stimulus. When using intraclass correlation coefficients (ICC), Bloom and Fisher (1982) suggested that scores above .80 are good and that scores above .70 are satisfactory. In Southwood's (1996) study of bizzareness and naturalness, individual ICCs on DME ranged from .79 to .84 (mean = .92 for bizzareness and .84 for naturalness). These listeners heard and rated 60 sentences in total, 15 of these sentences were rated twice. Southwood and Weismer (1993) had listeners rate bizzareness, acceptability, naturalness and normalcy of dysarthric speech from ALS talkers during automatic speech, imitated sentences and read paragraphs. Individual listener's ICCs ranged from .57 to .99. Mean ICCs across listeners showed that rating paragraphs resulted in the highest ICCs, all above .77. Mean ICC ratings of sentences and automatic speech ranged from .69 to .83 with ratings of acceptability and naturalness resulting in the lowest ICCs (all mean scores were below .75).

Eadie and Doyle (2004) asked listeners to rate the voice pleasantness and acceptability[15] of tracheoesophageal speakers using both DME and interval scaling. Listeners heard the same sentence spoken by 20 different talkers and rated the two voice characteristics. They heard five of the speakers a second time for a measure of intra-rater reliability. Pearson correlation coefficients on the DME ranged from .48 to .99 for voice pleasantness (mean = .77) and from .57 to 1.00 for voice acceptability (mean = .84). Whitehill, Lee and Chun (2002) had listeners rate hypernasality using DME while listening to connected speech samples. Reliability was calculated using Pearson product moment coefficients. When listeners were given a modulus to compare their responses to, mean reliability was .67. When listeners were not given a modulus, mean reliability was .95. Based on these studies, it can be expected that the mean reliability for DMEs should be between .67 and .95 and range between .48 and .99.

Summary

Reduced speech intelligibility can negatively affect communication as it makes interactions with others more difficult and effortful and can lead to frustration for the talker and listener and decreased talker participation in communicative interactions. The

[15] Pleasantness was defined as how pleasing the listener found the speaker's voice; acceptability was defined as how acceptable the listener found speaker's voice.

severity of a distorted speech signal, such as occurs in the speech of children with dysarthria, has serious implications on how much a listener can understand of what is said, how long it takes to process the message and how much mental processing effort is required to understand it. Young children with moderate to severe dysarthria are at a high risk of encountering communication situations where there is a breakdown and their message is not understood (Hodge & Wellman, 1999). Therefore it was of interest to examine the effectiveness of a commonly used repair strategy, specifically repetition, on increasing how accurately these children's spoken words can be understood. Signal dose refers to the number of times the word is presented to the listener. There was also potential for the effect of signal dose to differ for children with more or less intelligible speech. For the purpose of this study, severity of a child's speech disorder due to spastic dysarthria (ranging from mild to severe) was classified based on recordings of spoken sentences, judged in ideal listening conditions by unfamiliar adults. The stimuli presented to listeners in the current study were audio recordings of single words in an SNR that is within the typical range for a classroom and the child talker's everyday life (+ 10 dB SNR). It was considered important to determine the effects of variables such as disorder severity and signal dose in this level of noise, as opposed to an "ideal" listening environment to give the best advice to children with reduced speech intelligibility and their families and teachers about using repetition as a repair strategy.

Three measures of listener behaviour that included two objective measures (accuracy of word identification, i.e., speech intelligibility scores, and response time) and one subjective measure (perceived effort) were obtained. It was anticipated that information about the degree of effort that listeners exert and report when listening to children with varying severities of dysarthria might help listeners better prepare for communication opportunities with these children.

Based on the literature reviewed to date, no published studies were found to answer two important questions raised by the preceding literature review. Specifically;

1) What is the effect of increasing signal dose (number of presentations) on listeners' ability to identify single words produced by young children who vary in severity of dysarthria, in a less than ideal listening environment? This has implications for communication repair training choices and education of children with reduced speech intelligibility.

2) How strong are the relationships between subjective and objective measures of listeners' mental processing in identifying these children's words? This has implications for educating listeners about how much effort might be required to understand children with intelligibility deficits.

Purpose

The purpose of this study was to evaluate the effects of talker severity and signal dose (repeated presentations) on listeners' abilities to understand words spoken by children with dysarthria. To increase generalizability of the findings, these measures were obtained using a SNR (e.g., + 10 dB) that is within the typical range of a classroom environment. No literature was located that reported speech intelligibility measures for children with varying severities of dysarthria in a less than ideal listening environment. Furthermore, this study provides new information about the relationship between speech intelligibility scores and listener response times; speech intelligibility scores and measures of perceived listener effort; and listener response times and measures of perceived listening effort.

Research Questions and Hypotheses

Accuracy: Word Identification Scores

This study posed the following questions with regard to the dependent variable of word identification score:

1) What is the effect of severity of childhood dysarthria on listeners' single word identification scores when stimuli are balanced for phonetic content, word frequency and phonological neighbourhood and presented in an SNR that is typical of a classroom environment?
   - Hypothesis: Based on findings from previous literature, there will be a significant effect of severity of speech disorder on word identification scores such that listening to children with more severe dysarthria will result in lower speech intelligibility scores.
2) What is the effect of repeating the identical signal on listeners' word identification scores when stimuli are presented in an SNR that is typical of a classroom environment?
   - Hypothesis: Based on previous literature, there will be a significant effect of dose such that more presentations, or an increase in dose, will lead to higher speech identification scores (potentially 4 – 7 % based on connected speech) due to repetition priming and

increased opportunities for the listener to attend to different parts of the speech signal.

3) Does 'dose' or number of presentations interact with severity of childhood dysarthria on listeners' word identification scores?

    - Hypothesis: There will be a significant interaction between severity of childhood dysarthria and signal dose, i.e., the effect of more repetitions on word identification scores will differ depending on talker severity. This hypothesis is non-directional. Repetitions may benefit talkers with less severe dysarthria (e.g., there is more acoustically accurate information to work with), or they may benefit talkers with more severe dysarthria (e.g., single words have been shown to the most understandable form of speech for these individuals).

Response Time

This study posed the following questions with regard to the dependent variable of response time:

1) What is the effect of severity of childhood dysarthria on listeners' response times when stimuli are presented in an SNR that is typical of a classroom environment?

    - Hypothesis: Based on previous literature on signal clarity, listening to children with more severe dysarthria, as measured by identifying words spoken in sentences in ideal listening conditions, will result in longer response times due to increased processing time required.

2) What is the effect of repeating the identical signal on listeners' response times when stimuli are presented in an SNR that is typical of a classroom environment?

    - Hypothesis: Based on previous literature on repetition priming, there will be a significant effect of dose such that more repetitions, or an increase in dose, will lead to faster response times due to listeners being more confident in their response and already having an idea of what the word might be when listening to the same signal a second or third time.

3) Does 'dose', or number of presentations, interact with severity of childhood dysarthria on listeners' response times?

    - Hypothesis: There will be a significant interaction between severity of childhood dysarthria and signal dose, i.e., the effect of more presentations on listeners' response times will differ depending on severity of the speech disorder. This hypothesis is non-directional. Repetitions may result in listeners responding faster with talkers with less severe dysarthria (e.g., listeners will be more confident in their responses), or they may result in listeners responding faster with talkers with more severe dysarthria (e.g., they may have little information to make a decision and make one quickly).

Effort Ratings

This study posed the following question with regard to the dependent variable of listeners' perceived effort:

What is the effect of severity of childhood dysarthria on perception of listening effort measured using a direct magnitude estimation (DME) task when stimuli are presented in an SNR that is typical of a classroom environment?

- Hypothesis: Based on previous literature on rating scales and speech disorders, listening to children with more severe dysarthria will result in higher ratings of perceived effort needed to understand the child's words.

Is DME a reliable method to measure perceived effort?

- Hypothesis: DME will be a reliable procedure for measuring perceived effort as demonstrated by listeners achieving a mean intra-rater reliability coefficient of a minimum of .70 (Bloom & Fisher, 1982).

Relationship between Objective and Subjective Measures of Listener Behaviour

This study posed the following questions with regard to the three dependent measures of listener behaviour investigated in this study.

How well do listeners' single word identification scores predict an objective measure of their behaviour (i.e., response time)? How well do listeners' single word identification scores and response times each predict a subjective measure of their behaviour (i.e., DME ratings)?

- Hypothesis: Based on previous literature, for the three measures of listener behaviour, single word identification scores will predict response times (negative relationship) such that listeners who have higher word identification scores will have shorter response times. Single word identification scores will predict subjective effort ratings (negative relationship) such that listeners who have lower word identification scores will assign higher effort ratings. Response times will predict subjective effort ratings (positive relationship) such that faster listeners will assign lower effort ratings (less difficult). Of these three predictions, it was expected that the strongest relationship would be between the two objective measures: single word identification scores and listener response times.

Listeners' Beliefs about Listening Task 1

This study posed the following three questions related to listeners' perceptions of the word identification task:

Q1) What kinds of things made it easy to understand these children's words?

Q2) What kinds of things made it difficult to understand their words?

Q3) When a word was difficult to understand, what kinds of things did you do to try and understand the word?

- There were no specific hypotheses related to these questions. However, it was expected that listeners' descriptions would provide insight into their experiences and relate to the quality and clarity of the signals they were judging.

METHOD

Design and Analysis

This study used several designs to answer the research questions and test the associated hypotheses. Separate two-factor analyses of variance, with repeated measures on each factor, were conducted for each of the dependent variables 1) word identification scores and 2) response times. A multivariate analysis of variance was not conducted for these two dependent variables because of the negative relationship expected and confirmed between them ($r = -.42$, $p = .016$). Several authors (Field, n.d.; Max & Onghena, 1999) have identified potential problems when repeated measures are used. Specifically, if there is a correlation between levels of the repeated factor, sphericity can not be assumed. In this case, corrections need to be made to ensure that the appropriate degrees of freedom are used when calculating the F statistic. Greenhouse-Geisser, Huynh-Feldt and the Multivariate Test statistics provide these corrections and are available in SPSS ver. 13 (2004). Application of these tests using SPSS ver. 13 to the two 2-way ANOVAs and the 1-way ANOVA in this study revealed no differences in the *p*-values reported. The Greenhouse-Geisser statistic is reported in the Results section when sphericity could not be assumed.

1. A 3 X 4 Within-Subjects Quasi-Experimental Design was used to test the hypotheses about the main effects of talker severity and number of presentations and the interaction between these two factors on word identification scores. The categorical variable is talker severity. This variable has four levels with four children contributing audio recordings, one to each level. The independent variable is number

of presentations. This variable has three levels, i.e., how many times the signal was presented (first, second or third time). The dependent variable is word identification score and the unit is number of words identified correctly (out of a possible 25). The word stimuli from the four talker severity levels were pooled and randomized for presentation to reduce the effect of familiarization on listeners' scores.

2. A 3 X 4 Within-Subjects Quasi-Experimental Design was used to test the hypotheses about the main effects of talker severity and number of presentations and the interaction between these two factors on listeners' response times. The categorical variable is talker severity. The independent variable is number of presentations. The dependent variable is response time, measured in seconds. This was defined as the time lapse from the beginning of the word being presented to the time the listener began to enter their response. This time was corrected for the duration between onset of the .wav stimulus file and onset of the word and "corrected" to adjust for individual differences in typing speed. These corrections are further described in the procedure.

3. A one-way Within-Subjects Quasi-Experimental Design was used to test the hypothesis about the main effect of speaker severity on listeners' measures of perceived effort. The categorical variable is talker severity. The dependent variable is a rating of listening effort. This measure was collected during a second task by having listeners create a Direct Magnitude Estimation (DME) of how much effort it took to understand a spoken word. This second listening task was completed after all word identification scores had been collected for all children. Listeners were presented with a modulus. This was a word spoken by a child with dysarthria who had an

intelligibility score of 54% on The *Test of Children's Speech (TOCS)* sentence intelligibility measure (Hodge, 1996) and who was not a talker for the word identification task. This modulus was selected from available subject recordings as this child was closest to the mid-range of talker severities (38 – 86% on *TOCS* sentence intelligibility measure) used in the study. Listeners were told that the reference for this modulus was 100. Listeners then heard 48 words. This included 10 words spoken by each of the four children whose recordings were used in the word identification task (these are a subset of the 25 words used in the first task), plus a second presentation of two words, selected randomly from each child's set of 10 words, to provide a measure of intra-rater reliability on this task. These 48 word stimuli were randomized for each listener. After listening to each word, listeners were asked to make a rating response. The modulus was presented again after every 10 stimuli, following the procedures of Southwood (1996). Responses were to be made as a ratio compared to this modulus. The mean of each listener's DME responses for each child's recordings served as the dependent variable for listening effort.

4. Intra-rater reliability of the DME ratings was evaluated using an intraclass correlation coefficient (ICC). Two of the 10 words from each child were presented twice to each listener for rating. Across the four children this provided a total of eight words rated twice for each judge. The ICC calculation measures the consistency or agreement of two given values (first response and second response). A score of 1.00 indicates that the two variables are in perfect agreement. An ICC score based on the consistency of the eight words for each listener was calculated.

5. The relationships between the two objective measures of listener behaviour, word identification scores and response times, and the extent to which each of these measures predicted the subjective measure of listener behaviour, effort ratings, were evaluated using linear regression. This provided new information about listener behaviours and how these relate to each other. Word identification scores was theoretically the more important behavioural variable and was used to predict response time. Global measures of word identification scores, response times and ratings of subjective effort for the first presentation only were used as the respective dependent variables. For word identification scores and DME effort ratings, the mean score across the four children for each listener served as the respective global measure for these two variables. For response time, the median score across the four children for each listener served as the global measure. Procedures for calculating these global scores were based on cautions identified by Lorch & Myers (1990) for within-subjects designs. The linear regression analysis used to predict response times from word identification scores was based on listeners' responses to the full set of 100 words from all talkers (i.e., mean word identification scores and mean response times for each listener for the first presentation of all 100 stimuli). The linear regression analyses used to predict DME effort ratings from word identification scores and from response times were based on the subset of 40 words form all talkers used in the DME task (i.e., mean word identification scores, mean response times and mean DME effort ratings for each listener for the 40 words).
6. Listeners' beliefs about the nature of the task were obtained using three open-ended questions related to the ease or difficulty of certain words and strategies the listener

used to try and increase their understanding. A theme analysis was conducted to identify the nature and relative frequency of their responses to each question.

Participants

Thirty-three listeners were recruited for this study. To reduce possible effects of language level or hearing acuity, listeners had some level of post-secondary education and normal hearing as evidenced by self-report of passing a hearing screening within the past 12 months or passing a hearing screening on the day of participation (American Speech and Hearing Association, 1985). Listeners were also monolingual English speakers[16] with minimal experience in listening to distorted speech (e.g., not a teacher of English as a Second Language or experienced listening to children with speech disorders[17]). Listeners were recruited based on their past or current enrollment in professional programs that could lead to employment in settings where contact with young children with dysarthria is expected. These included students in the Faculty of Education and in the Departments of Occupational and Physical Therapy. This number of subjects was based on a sample size calculation using an estimated standard deviation of 10% from previous literature (Hodge, 1996) and a minimal mean improvement of 10% between scores in the original and repeated presentation conditions. The standard deviation of 10% is based on data for 15 groups of three listeners where each group of

[16] Three listeners indicated that they took French in school; however they did not consider themselves bilingual.
[17] Listeners were asked if they had ever taught children with speech disorders, worked or volunteered at functions with many children with speech disorders or had any family members with speech disorders. One listener stated that she had limited experience working with deaf children in the past, two stated that they worked occasionally with a child/children with autism and two listeners indicated that they had experience teaching in a classroom that had a child with a speech disorder in it. No listeners indicated having family members with speech disorders or working/volunteering at functions with many children with speech disorders.

three listeners judged a given child's *Test of Children's Speech (TOCS)* single word intelligibility measure in an open-set word identification task. The 15 children whose word tests were judged all had dysarthria and ranged in speech disorder severity from mild to severe. The mean standard deviation of the 15 groups of three listeners was 3.8% (SD = 2.1%). An estimated standard deviation for the sample size calculation was set at 3.8% plus 3 standard deviations (3 x 2.1), which equals approximately 10%. A minimum of a 10% difference was selected based on the magnitude of changes reported in the literature reviewed that reported significant differences in identification scores for repeated presentation of stimuli (4% - 7%) as well as what might be considered a clinically significant difference in word identification intelligibility scores (Yorkston & Beukelman, 1981). Using these estimated values for a standard deviation of 10%, a critical difference of 10%, a power level of .8 and a p value of .01, a minimum of 33 subjects was needed. For the main effect and interaction analyses, a p value of .01 was selected as a "familywise" error size because of the number of analyses being conducted (.05 / 3 = .017). Appendix A illustrates the power calculation used to determine the minimum number of listener subjects needed.

Data Collection

Selection of Stimuli

Selection of the Four Child Talkers (Severity Level Factor)

Four children with spastic dysarthria associated with a diagnosis of spastic cerebral palsy were selected to represent four different severity levels of speech disorder. The speech of these children had been recorded previously as part of a larger research study in the Speech Analysis Laboratory in the Department of Speech Pathology and Audiology at the University of Alberta. Severity level was determined by examining several measures. Intelligibility scores (percent correct word identification) of these children's spoken sentences on the *TOCS* sentence intelligibility measure in ideal listening conditions from past listeners, provided an initial severity rating. The desired four levels of severity of speech disorder based on the *TOCS* sentence intelligibility scores were; mild (85-100%), mild-moderate (65-75%), moderate-severe (45-55%) and severe (<35%). These ranges were identified so that there would be at least a 10% difference in intelligibility scores for the four children. The four best candidates were chosen from the pool of children's recordings available. Each child selected fit into one of these levels of severity except for the mild-moderate category. This child had a score of 76% on the sentences. This was considered acceptable as it was only 1% above the top of the range for mild-moderate and there was still a 10% difference between this child's score and the score of the child selected for the mild severity level.

Percent Consonants Correct (PCC), Percent Vowels Correct (PVC) and Percent Phonemes Correct (PPC) were calculated from a recorded spontaneous speech sample for each child using procedures outlined in Shriberg (1986). These provided a second source of information about relative severity. Speech samples were recorded to video cassette tape and transcribed using both video and audio signals. Appendix B provides a description of Shriberg's conventions for obtaining PCC, PVC and PPC and obtaining severity ratings from the PCC. The ratings are mild (PCC = 85-100%), mild-moderate (PCC = 65-85%), moderate-severe (PCC = 50-65%) and severe (PCC < 50%). The PCC measure accurately predicts severity of speech disorder, as measured on a 7-point rating scale by speech-language pathologists, in 90% of cases (Shriberg & Kwiatkowski, 1982). However, it is noted that these results are for children with a developmental phonological disorder from a variety of etiologies, not specifically dysarthria. As no guidelines for severity ratings that include PVC have been published by Shriberg and colleagues, the PVC measures for the four children were compared with normative data provided by Austin and Shriberg (1997). The PPC value combines both the PCC and PVC to give an overall measure of how accurate the child is at producing all sounds. The PPC value was also compared with normative data provided by Austin and Shriberg (1997). Table 3 provides descriptive information about the child talkers' characteristics and their respective sentence intelligibility scores, PPC, PVC and PPC measures.

As is evident in Table 1, the severity classification of the *TOCS* sentence intelligibility scores and the PCC scores did not mirror each other. While the sentence scores differentiated all four children according to the desired severity levels, PCC scores were very similar for Child 1 and 2 and for Child 3 and 4. Applying the PCC relative severity ratings to the PCC scores, Child 1 would be classified as moderate-severe, Child 2 as mild-moderate, Child 3 as moderate-severe and Child 4 as moderate-severe. PVC scores and standard deviations from the mean reflected the same ordering of severity as the *TOCS* sentence intelligibility measure for all children except Child 2 and 3. The standard deviations ordered Child 2 as more severe than Child 3 due to Child 2 being one year older. PPC scores and standard deviations from the mean did reflect the same ordering of severity as the *TOCS* sentence intelligibility measure; however the ranges did not differentiate the children as well. For example, PPC scores grouped Child 1, 2 and 3 within 10% of each other (63-73%).

As speech intelligibility is influenced by factors other than phoneme accuracy, (e.g., voice and resonance features, prosodic features), the severity levels based on the *TOCS* sentence intelligibility measure were used, with the knowledge that the range and number of severity levels were not maintained for the four children based on PCC, PVC and PPC scores. The range of speech severity represented by the *TOCS* sentence intelligibility measure, with the added detrimental effect of background noise, allowed listeners to improve their identification scores over repetitions without a ceiling effect.

The words spoken by these children had already been recorded as part of a larger research study in the Speech Analysis Laboratory in the Department of Speech Pathology and Audiology at the University of Alberta. Speech samples were recorded to cassette tape and then digitized at a sampling rate of 22 kHz with 16-Bit quantization using CSpeech Version 4.0 (Milenkovic & Read, 1992). Files were saved as digital audio files (.wav files).

Table 1. *Characteristics of Child Talkers*

Child Code	Sex	Age in Years	Neurological Diagnosis	Type of Dysarthria	% words correct[a]	PCC[b]	PVC[d]	PPC[f]	Total # of words[g]
1	M	4	Right Spastic Hemiparesis	Spastic	86.5% Mild	64.60% Mod-Severe[c] 177/274	83.42% 5.03 SD[e] 156/187	72.23% 2.47 SD[e] 333/461	174
2	M	5	Spastic Diplegia	Spastic	76% Mild-Mod	67.41% Mild-Mod 213/316	73.17% 12.02 SD 180/246	69.93% 4.39 SD 393/562	205
3	F	4	Left Spastic Hemiplegia	Spastic	55% Mod-Sev	58.28% Mod-Severe 169/290	72.06% 10.15 SD 147/204	63.97% 5.22 SD 316/494	199
4	F	4	Spastic Quadriplegia	Spastic	38% Severe	57.80% Mod-Severe 163/282	43.66% 22.50 SD 93/213	51.72% 8.07 SD 256/495	183

Note. PCC, PVC and PPC were calculated from a spontaneous speech sample following Shriberg's (1986) conventions as Described in Appendix B. For Child 1, 2 and 3, this included 90 first occurrence words. For Child 4, this included 70 utterances with at least one non-questionable word. Fractions represent total correct consonants/vowels/phonemes over total number of consonants/vowels/phonemes used in the sample.
[a]Mean from three listeners identifying words from *TOCS* sentence intelligibility measure (80 words total). [b]Percent Consonants Correct. [c]Severity classification based on Shriberg & Kwiatkowski (1982). [d]Percent Vowels Correct. [e]SD indicates the number of standard deviations that the child's score is below the age norms for all sounds, for children following normal speech acquisition and males and females separately, reported by Shriberg and Austin (1997). [f]Percent Phonemes Correct. [g]Total number of individual words used in the PCC, PVC and PPC analyses.

Selection of the Word Stimuli for Listening Tasks 1 and 2

The *TOCS* single word intelligibility measure has three versions, each containing 78 words. The *TOCS* uses a phonetic contrast approach to intelligibility assessment, targeting phonetic contrasts known to be difficult for children with dysarthria. Contrasts include consonant voicing (e.g., bat vs. pat), place (e.g., seat vs. sheet) and manner (e.g., pan vs. fan); vowel place (e.g., heat vs. hoot), height (e.g., heat vs. hat), length (e.g., beat vs. bit), manner (e.g., lawn vs. line), point-central (e.g., hot vs. hut) and null-rhotic (e.g., jaw vs. jar); and syllable shape (e.g., heat vs. _eat), number (e.g., bee_ vs. beanie) and stress (e.g., a knee vs. Annie). Administration of these word tests consists of two parts: recording child productions and playing back to listeners for judging. Child word productions are elicited by a visual picture and audio model. The child's productions are audio recorded and then played to listeners using one of two response options. Listeners use either the open-set format where they type the word they heard the child say or the closed-set format, where they are asked to choose from a selection of four options (target word, minimal pair contrast, blank space for entering a separate response or can't identify).

The children selected for this study had recorded words from at least two different word tests. Twenty-five words were selected for each child such that the four different lists would not repeat any words and would be approximately equivalent in terms of phonetic balancing, word frequency and phonological neighbourhood. This balancing procedure is further described in the next section.

Listening Task 1: Word Identification

The set of 25 words selected for each child differed across the four children to reduce any effect of presenting the same words multiple times. Appendix C contains the list of words selected for each child. Individual words were chosen based on creating subsets of phonetically balanced word lists from the pool of words recorded from each child's *TOCS* single word intelligibility measure productions. These lists were also required to be approximately equal in phonetic and written frequency and phonological neighbourhood. Each of the four word lists has an equal distribution of syllable types and approximately equal distribution of individual consonants and vowels in the different word positions (see Benki, 2003). Appendix D contains a description of the phonetic balancing results for the four word lists.

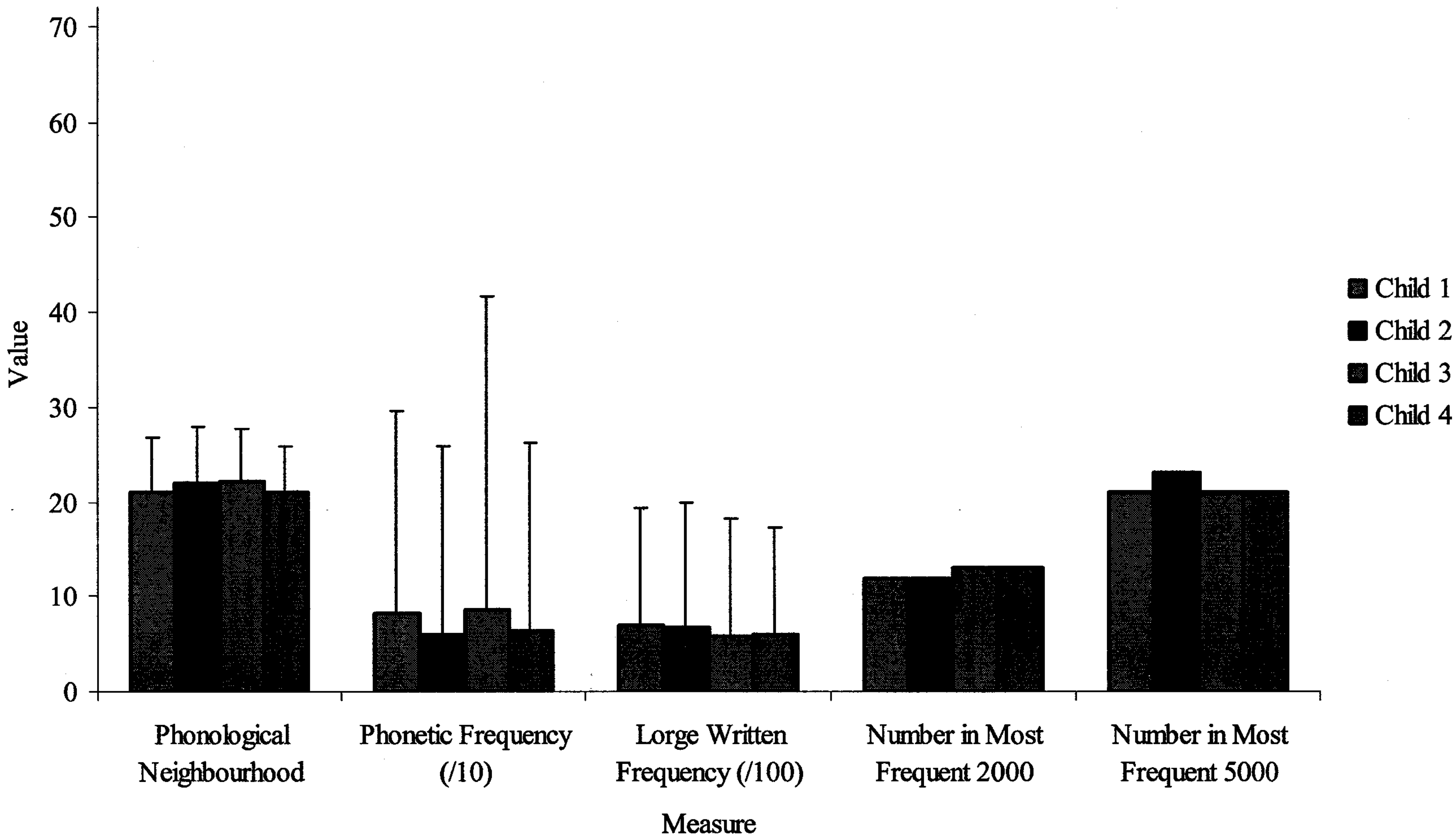
Measures of spoken phonological neighbourhood and word frequency for each child's list of word stimuli are reported in Appendix C. Phonological neighbourhood and frequency values were calculated for each word and averaged for each child. A word's phonological neighbourhood is the number of words in the English language that differ in only one phoneme from the target word. These values were obtained from a database not yet publicly available (C. Westbury, personal communication, April, 2006). The mean phonological neighbourhood for the four levels of talker severity ranged from 21.00 to 22.20 (SD = 4.94 to 5.83). This indicates that many of the selected words were similar to many other English words. Words with more than eight phonological neighbours have been previously described as having a dense phonological neighbourhood (Grainger et al., 2005). For this study, with the mean phonological neighbourhood above 20, these word lists are considered to have very dense phonological neighbourhoods[18].

In this study, spoken word, or phonetic frequency for each word was obtained from a database of 1,000,000 written words (C. Westbury, personal communication, April, 2006 from Baayen, Piepenbrock, & Gulikers, 1995). A word with a high phonetic frequency indicates that the word is spoken more than a word with a low phonetic frequency. These measures were obtained from a database not yet publicly available. The mean phonetic frequency for the four levels of talker severity ranged from 59.04 to 85.60 (SD = 199.44 to 330.31). Written word frequency was obtained from the Thorndike and Lorge lists (1944). These lists report the number of times each word appeared in a 4,500,000 written word sample. The mean written frequency for the four levels of talker severity ranged from 567.60 to 684.72 (SD = 1141.40 to 1328.31). Between twelve and thirteen of the 25 words were in the most frequent 2000 and between 21 and 23 were in the most frequent 5000. Overall, this indicates that most of these words were highly familiar (Balota & Chumbley, 1985). Figure 2 compares the phonological neighbourhood and frequency analysis results for the word lists for each of the four children.

Two practice items, which were not the same as any of the words on the four stimulus lists, spoken by a child who was not one of the four children, were used to acquaint the listeners with the task. This child had spastic dysarthria and her score, averaged across three listeners, on the *TOCS* sentence intelligibility measure was 74%.

[18] Only one word (tube, spoken by child 3) had a phonological neighbourhood below 10, it had a phonological neighbourhood of 3.

Figure 1. Phonological neighbourhood and frequency equivalence means across child talkers for Listening Task 1.



Note. Standard deviations shown as error bars

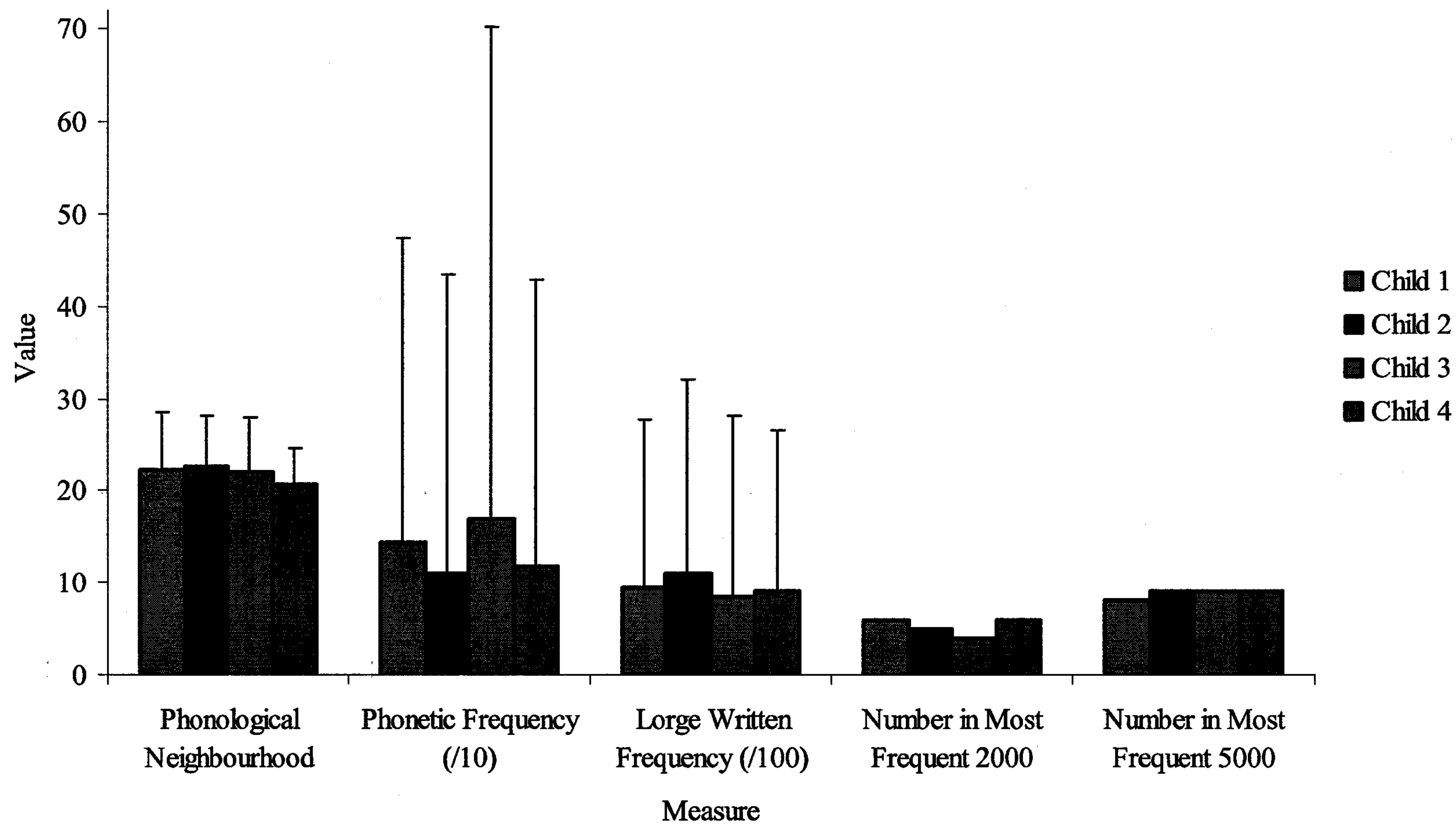
Listening Task 2: Effort Rating Using DME

Ten words from each of the four children were selected for the test items for Listening Task 2. These words are subset of the 25 words used in Listening Task 1. These subsets were required to be approximately equal with regards to phonetic balancing, word frequency and phonological neighbourhood. These values were calculated as they were in Listening Task 1. Appendix E contains the words used in Listening Task 2 along with phonological neighbourhood and word frequency balancing information, organized by each child. Appendix F contains the phonetic balancing information for the words in Listening Task 2. The mean phonological neighbourhood for the four levels of talker severity ranged from 20.60 to 22.70 (SD = 3.95 to 6.42). The mean phonetic frequency for the four levels of talker severity ranged from 109.50 to 170.10 (SD = 311.40 to 533.11). The mean written frequency for the four levels of talker severity ranged from 853.90 to 1097.89 (SD = 1734.35 to 2110.41). Between four and six of the 10 words were in the most frequent 2000 and between eight and nine were in the most frequent 5000. Overall, the words chosen for Listening Task 2 were those that had higher frequency scores in the set of 25 words used in Listening Task 1. Overall, words used in Listening Task 2 had a dense phonological neighbourhood and were highly familiar (Balota & Chumbley, 1985; Grainger et al., 2005). Figure 3 shows the results of the word list phonological neighbourhood and frequency balancing analyses for the 10 words selected for each of the four children.

Four words, one from each of the four child talkers whose recordings were rated, were used as practice items. These words were also unique from the test words used in both tasks 1 and 2. One word recorded from a child with dysarthria who was about 54%

intelligible on the *TOCS* sentence intelligibility measure (near mid-range of the range of severity used in this study) was used as the modulus for the DME rating task. This word was not the same as any of the words used in either of the two listening tasks.

Figure 2. Phonological neighbourhood and frequency equivalence means across child talkers for Listening Task 2.



Note. Standard deviations shown as error bars.

Preparation of Word Stimuli

Each stimulus word existed as a .wav file (22 kHz sampling rate and 16 bit quantization). A file with a 2.0s sample of background noise consisting of broad-spectrum multitalker babble with a relatively flat intensity was created from an existing larger .wav file. The volume of the audio files was equated using GoldWave v5.12 (GoldWave Inc., 2005). All 100 test files, the 6 practice files, the modulus file for Listening Task 2 and the noise segment were processed as follows. The volume was maximized to the full dynamic range to make the files "as loud as possible without distorting or clipping the waveform" and the new files were saved for further preparation (GoldWave Incorporated, 2005). The noise segment was manipulated by decreasing the strength of the entire 2 second segment by 10 dB. Each word was mixed with the manipulated background noise file using the multichannel audio software component of Adobe Audition 1.5 (Adobe Systems Incorporated, 2004). This created a file with the word at + 10 dB SNR. The two files were then saved as one .wav file for playback to listeners. The volume level for playback to listeners was set for all listeners at about 60 to 65 dB A.

Procedure

The recruitment poster appears in Appendix G. When potential listeners contacted the researcher, appointment times were arranged. Listeners were instructed to meet at Corbett Hall at the University of Alberta for participation. Participation was completed in a professional, acoustically treated sound booth. Housed in the booth were a computer screen, keyboard and mouse and two ElectroVoice S-40 compact monitor speakers. The computer hard drive was located outside of the sound booth and connected to a Technics

Stereo Integrated Amplifier (model SU-V460). Upon entering the listening booth, listeners were given a brief explanation of the study and then the information letter which they were asked to read. A copy of the information letter is included in Appendix H. They were also given a copy of the consent form, shown in Appendix I. Listeners were informed that they could have a copy of either if they wished. Listeners were asked about which faculty they were enrolled in, their year of study, birth date, past hearing screenings and history of hearing loss. A hearing screening was given to listeners who had not passed a hearing screening within the last year (ASHA, 1985). Listeners were also asked about their own language background (e.g., monolingual, bilingual), prior experience with English as a Second Language and experience with children with speech disorders.

Listeners then completed a short typing task. The purpose of this task was to obtain a measure of each listener's typing ability, defined as the mean time it took them to type a single letter in ideal conditions (seconds per letter). Listeners were asked to type the sentence "the quick brown fox jumps over the lazy dog." as quickly and accurately as possible. They were also given a written copy of the sentence so they did not have to rely on memory. The time in seconds it took from typing the initial character in the sentence to typing the period at the end was recorded. If the listener indicated that their attempt was not an accurate reflection of their typing speed, or more than three errors were made (less than 90% accuracy), they completed the task again and the second trial was used. This time was then divided by 35 (the total number of letters in the sentence) and the resulting number was the mean seconds per letter score for the listener. The mean seconds per letter score was then subtracted from the response times obtained during

Listening Task 1 for that particular listener. As these response times were measured from the start of the word being presented to when the first letter was entered by the listener, subtracting the mean time it took that particular listener to type a single letter provided a means to adjust for individual differences in typing speed.

Listening Task 1

To introduce the task, listeners were given both verbal and written instructions. These instructions are provided in Appendix J. Listeners were also given a list of 200 words to read over to familiarize themselves with the kinds of words they would be asked to identify. This list contained the 100 test words plus 100 additional words. This list of priming words appears in Appendix K.

Each listener listened to all 100 words (25 spoken by each of the four children). Words were presented using computer software that allowed listeners to type in what they heard. This software was created using Authorware ver. 6 (Macromedia, 2001). Listeners first practiced the task by hearing two words spoken by a fifth child. They heard each word three consecutive times, typing in what they heard after each presentation. The test words were presented one word at a time and each word was presented three consecutive times. Listeners were instructed to type in what they heard the child say after each presentation. This resulted in each listener typing in a response a total of 300 times for the children's stimulus words. Listeners had as much time as they need to respond. The next audio stimulus was not played until the listener pressed 'enter'. The software pooled the four children's test words and then created a unique randomized presentation order for each listener to control for order effects. The computer software recorded and saved the responses entered by the listener in text format. Appendix L shows the presentation screen for Listening Task 1.

The text file created by the computer program listed the word spoken, the child number, the presentation number, response typed in by the listener, the accuracy (exact letter by letter match only) and response time (in seconds). Responses were scanned for homonyms with actual word spoken (e.g., “die” for “dye”) and perceived typing errors (e.g., “sheett” for “sheet”) and instances of these were counted as accurate. Data were then sorted by child, by presentation number and then by word such that for Child 1, the first presentations for each word were grouped together alphabetically. A template was created to count the number of words correct (word identification score) for each child at each presentation number for each listener.

Computer response times were measured as the latency between the beginning of the audio file and the onset of the first keystroke of the listener’s response. This was recorded as listeners typed in a response to each presentation of each stimulus word. Two corrections were required to obtain the final response time scores.

a. A correction was required to subtract the time from the beginning of the audio file (noise) to the onset of the word. Table 2 gives an example of this correction.

Table 2. *Example Response Time for One Listener Identifying Words Spoken by Child 2 Corrected for the Difference Between the Audio File and the Word Starting*

Word	Program Response Time[a]	Difference[b]	Final Response Time[c]
back	3.06	0.35	2.72
bag	2.70	0.43	2.27
chip	3.02	0.46	2.56
comb	3.56	0.64	2.93
dye	5.89	0.56	5.33
feet	2.94	0.40	2.54
fill	3.42	0.48	2.95
hen	2.36	0.65	1.71
jay	11.11	0.55	10.56
leap	10.86	0.55	10.31
lip	2.23	0.56	1.67
low	3.70	0.63	3.08
match	3.38	0.13	3.24
nap	2.73	0.50	2.24
peas	5.45	0.51	4.95
pop	2.86	0.54	2.32
pot	2.73	0.56	2.18
pout	3.36	0.36	3.00
shut	5.27	0.42	4.85
sick	3.00	0.24	2.76
thin	3.33	0.64	2.69
toad	3.19	0.45	2.73
tongue	3.89	0.63	3.26
tube	4.39	0.55	3.84
whoa	2.94	0.56	2.38

[a]Time from the audio file beginning to the first letter being typed by the listener (seconds). [b]Difference between the audio file beginning and the beginning of the word (seconds). [c]Program Response Time minus Difference (seconds).

b. A second correction was required to account for the variability in participants' typing abilities in the response time measurement, using the mean number of seconds per letter score described previously. For example, if it took 6.51 seconds to complete the sentence, a score of .19 seconds per letter was calculated. This number, .19 seconds, was subtracted from each of the response time scores to give a score adjusted for the unwanted effect of variability across listeners in typing speed. Table 3 illustrates an example of this adjustment. Response time scores were collected for each signal presented. The median response time of each of the listeners' responses to the 25 words for each child in each presentation condition served as the dependent variable for response time. Median scores were used as they are less sensitive to outliers (Borowsky & Masson, 1996).

Table 3. *Example Response Time Corrected for a Listener's Typing Speed*

	Presentation Number					
	1		2		3	
	Median		Median		Median	
Talker Severity	U[a]	C[b]	U[a]	C[b]	U[a]	C[b]
1	2.40	2.21	2.08	1.89	1.83	1.64
2	2.67	2.48	2.52	2.33	2.22	2.03
3	2.60	2.41	2.27	2.08	1.76	1.57
4	3.36	3.17	2.63	2.44	2.26	2.07

Note. Each response time for each word was subtracted by .19 seconds. The median of these 25 responses are reported.
[a]Uncorrected response time. [b]Corrected response time.

The template described above included both of the corrections required for response time and calculated the median response time score for each listener for each child for each presentation condition. This response time score, corrected for both the noise lead time before the onset of the word and the listener's typing speed was used in all analyses involving response time.

Listeners were given a break after 34 items and after 66 items. After the first break, listeners were reminded that they could type in nonwords if they could not identify a real word match.

Listening Task 2

Measures of perceived listener effort were obtained in a second task after listeners completed the word identification task. A separate program was created using Authorware ver. 6 (Macromedia, 2001) for this task. To introduce the task, listeners were given both verbal and written instructions. These instructions are given in Appendix M. Listeners heard one word spoken from the modulus talker for the DME rating and told that the number associated with the effort required to understand that modulus was 100. They then practiced the DME task on four items, one from each child talker. Next, they heard 48 words, 10 from each child, with two words selected randomly from each child's set of 10 words played a second time to provide an estimate of intra-rater agreement on the DME task. A unique randomization of the 48 test words was presented to each listener. Listeners were asked to identify, by typing in a number, how much effort it took to understand each. The program saved these responses in a text format. Listeners were reminded about the modulus after rating 10 words, 20 words, 30 words and 40 words. For this reminder, the word for the modulus was played again and the number 100 (how much effort was required to understand) was associated with this word. Appendix N shows the presentation screen for Listening Task 2.

The text file created by the computer program listed the word spoken, the child number and the number response given for the 40 items and for the reliability items separately. Data was sorted by child and then by word and a mean score for each child for each listener was created. The initial response to the eight intra-rater reliability items was entered beside the second response for use in the calculation of ICC to estimate intra-rater reliability.

Additional word identification scores and response times from Listening Task 1 were calculated using a template that only included the first presentation of the 10 words from each child that were used in Listening Task 2. These mean and median scores were used in regression analyses involving the DME effort ratings. Scores were calculated in the same way as they were for Listening Task 1.

Open-Ended Questions about Listeners' Beliefs

Following completion of Listening Task 2, participants were asked to respond to three open-ended questions to provide insight into their listening experiences. They were asked to think back to the first task and then respond to the following questions: "We expect that some of the words that you have heard were easy to understand and that some were quite difficult to understand. What kinds of things made it easy to understand these children's words? What kinds of things made it difficult to understand their words? When a word was difficult to understand, what kinds of things did you do to try and understand the word?". The researcher transcribed the oral responses of each listener.

Following completion of the listening tasks, listeners were debriefed verbally and their questions were addressed. At the completion of the listening session they were provided with a $10 honorarium for their participation.

Analysis Summary

A two-way ANOVA (4 X 3) with repeated measures on both factors was used to determine if there were significant main effects for talker severity and presentation number and/or a significant interaction of these on the listeners' word identification scores. This procedure was repeated for the dependent variable of response time. Pair wise comparisons using the Least Significant Difference statistic with significance corrected for "familywise" error size were used to determine the nature of any significant differences found. For talker severity, the significance level of .008 was used (.05 / 6 comparisons). For presentation number, the significance level of .017 was used (.05 / 3 comparisons).

A one-way repeated measures ANOVA was used to examine the effect of talker severity on perceived effort ratings obtained in Listening Task 2. Pair wise comparisons were used to determine the nature of any significant differences observed. An intraclass correlation coefficient was calculated for each listener to estimate the intra-rater reliability for the listeners' DME ratings on the first and second presentation of the eight repeated items.

Effect sizes were calculated to determine the magnitude of observed significant differences. The two conditions with the smallest difference, that was statistically significant, for each variable were selected for the effect size calculation (Cohen, Welkowitz, & Ewen, 1991). For repeated presentations where the smallest difference was between presentation 1 and 2 or 2 and 3, an additional effect size was calculated for the difference between presentation 1 and 3. This was done as the advice given about the use of repetition as a repair strategy could be based on the effectiveness of two presentations compared with three presentations. Effect sizes were classified as trivial (<.20), small (.20), medium (.50) and large (.80) based on Cohen (1988). The following formula was used to determine the effect sizes.

$$\text{Effect size} = (\mu_1 - \mu_2)/SD_{pooled}$$

$$SD_{pooled} = \sqrt{(SD_1)^2 + (SD_2)^2}$$

Linear regression was used to examine the relationships between word identification scores and response times obtained in Listening Task 1 and between accuracy scores and effort ratings and between response times and effort ratings obtained in Listening Tasks 1 and 2, respectively. Global measures of word identification scores, response times and DME effort ratings for the first presentation of the stimulus were used as the dependent variable. The calculation of these global measures was described in the Design and Analysis section. For analyses involving DME effort ratings, word identification scores and response times included only the 10 words for each child used in Listening Task 2.

A theme analysis was conducted on the responses to each of the three open-ended questions about listeners' experiences by the student researcher. A second person (supervisor) reviewed the theme analysis. The final themes reflected the consensus of both student and supervisor.

RESULTS

Results are presented in the following order: Listening Task 1, Listening Task 2, investigation of the relationships between the objective and subjective measures of listener behaviour, theme analyses of the three open-ended questions posed to listeners' about Listening Task 1. For Listening Task 1, the effects of talker severity level and presentation number are reported first for accuracy scores and then for response times. Results on accuracy scores include the dependent measure of word identification scores as well as an analysis of phonetic accuracy based on listener's orthographic transcriptions. For Listening Task 2, results are presented first for the effect of talker severity on DME effort ratings and then for the estimation of intra-rater reliability for the DME task.

Effects of Talker Severity and Presentation Number on Accuracy

Words Identified Correctly

Group data for the 33 listeners' word identification scores for each level of talker severity and each presentation number are shown in Table 4. The mean word identification scores are also shown in Figure 4 for each level of talker severity and presentation number. Data are presented as means of the raw scores, out of a possible 25. For severity level, averaged across presentation number, word identification scores ranged from 2.42 (SD = 1.18) for Child 4 to 11.27 (SD = 1.79) for Child 2. For presentation number, averaged across severity level, word identification scores ranged from 8.18 (SD = 3.89) for the first presentation to 8.88 (SD = 4.02) for the third presentation. Mean accuracy scores increased slightly for each child across the three presentations. The greatest increase was for Child 2 (1.24 words) and the smallest increase was for Child 1 (.37 words).

A 4 X 3 ANOVA with repeated measures on both factors was used to examine the main effects of talker severity and presentation number as well as the possible interaction between these two factors. Sphericity could not be assumed for presentation number. The Greenhouse-Geisser F test was used in this case. This test is used for within subjects designs where the assumption of equal variances is not met (Marasculio & Serlin, 1988). As hypothesized, there was a significant main effect of talker severity (F (3, 96) = 266.73, $p < .001$) and a significant main effect of presentation number (F (1.65, 52.85) = 13.40, $p < .001$). The interaction between talker severity and presentation number was not significant (F (6, 192) = 2.02, $p = .064$).

Pairwise comparisons were conducted to determine the nature of the main effect of talker severity. There was a significant difference between Child 1 and 4 ($p < .001$), Child 2 and 3 ($p = .001$), Child 2 and 4 ($p < .001$) and Child 3 and 4 ($p < .001$). Child 2 was the most intelligible, followed by Child 1, Child 3 and Child 4. To gain insight into the magnitude of differences in word identification scores where statistically significant results were obtained, effect sizes were calculated for the smallest difference found to be significantly different using Cohen et al.'s (1991) procedure. This procedure is described in more detail in the Methods section. For the effect of talker severity, the smallest difference found to be significant was between Child 2 and 3 (a mean difference of 1.38). The effect size for this difference is .48, which is considered to be medium (Cohen, 1988). The effect size calculation is shown in Appendix O.

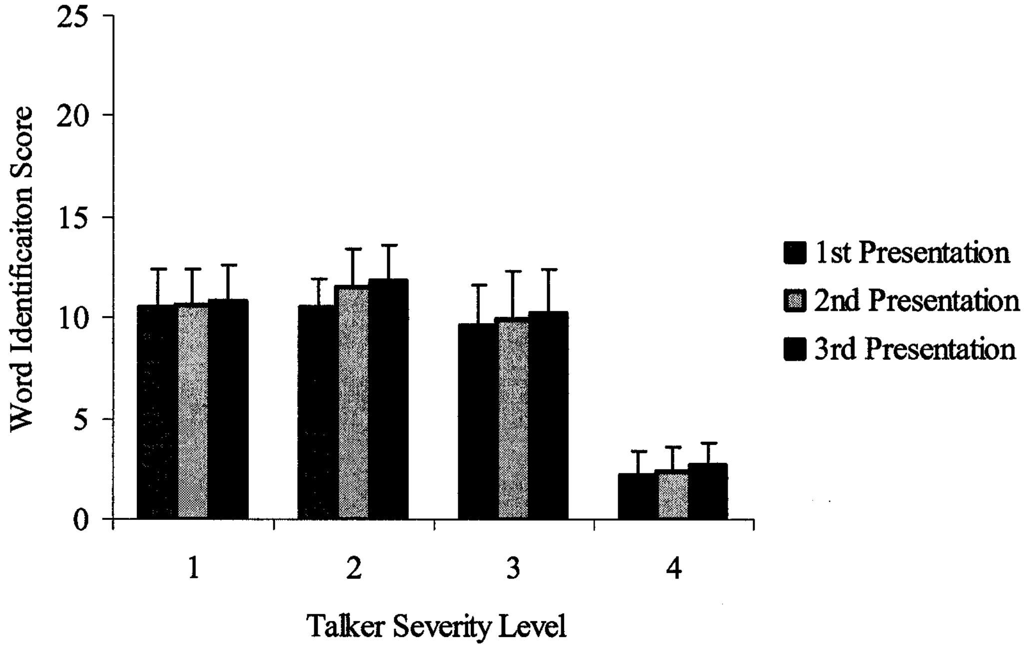
Pairwise comparisons were also conducted to determine the nature of the main effect of presentation number. There was a significant difference between presentation 1 and 2 ($p = .012$), presentation 1 and 3 ($p < .001$), and presentation 2 and 3 ($p = .013$) such that word identification scores increased with more repetitions of the word. For the effect of presentation number, the smallest difference found to be significant was between presentation 2 and 3 (a mean difference of .27). The effect size for this difference is .05, which is considered to be trivial (Cohen, 1988). The effect size for the difference between presentation 1 and 3 (a mean difference of .70) is .13, which is also considered to be trivial (Cohen, 1988). The effect size calculations are shown in Appendix O.

Table 4. *Word Identification Scores for Talker Severity and Presentation Number*

	Presentation Number							
	1		2		3			
Talker Severity	Mean	(SD)	Mean	(SD)	Mean	(SD)	**Mean**	**(SD)**
1	10.48	(1.92)	10.58	(1.86)	10.85	(1.75)	**10.64**	**(1.83)**
2	10.52	(1.37)	11.55	(1.87)	11.76	(1.87)	**11.27**	**(1.79)**
3	9.58	(2.05)	9.88	(2.47)	10.21	(2.18)	**9.89**	**(2.23)**
4	2.15	(1.23)	2.42	(1.17)	2.70	(1.10)	**2.42**	**(1.18)**
Mean (SD)	**8.18**	**(3.89)**	**8.61**	**(4.09)**	**8.88**	**(4.02)**		

Note. Values are means for all listeners out of 25. Bold Means and Standard Deviations are derived from each listener's mean scores, collapsed across one variable (e.g., for the three different presentations, 132 scores were used, for the four talker severity levels, 99 scores were used).
Note. Statistically significant differences for accuracy were found for talker severity ($p \leq .008$) (1 vs. 4, 2 vs. 3, 2 vs. 4 and 3 vs. 4); and for presentation number ($p \leq .017$) (1 vs. 2, 1 vs. 3, 2 vs. 3).

Figure 4. Listener word identification scores for talker severity and presentation number.



Note. Standard deviations shown as error bars.

Measures of Phonetic Accuracy

While the main effect of talker severity on word identification scores was as predicted, the actual ordering of word identification scores by level of talker severity was not as predicted. Child 1 was expected to have the highest accuracy scores, followed by Child 2 and 3, and Child 4 was expected to have the lowest scores. However, no significant differences were found between Child 1 and 2 or Child 1 and 3 and Child 2's mean word identification score was slightly higher than Child 1's mean score. A finer grained analysis using phoneme and syllable shape accuracy was conducted on listeners' responses to determine if this could better differentiate the four children. Listener scores, based on their orthographic response for each word, were determined for each of the following variables: initial consonants correct (25 possible), vowels correct (25 possible), final consonants correct (21 possible) and syllable shapes correct (25 possible). Table 5 reports the mean scores for all listeners at each talker severity level and presentation number for each of these four dependent variables. Ordering of Child 1, 2 and 3 varied by dependent variable, while Child 4 maintained the lowest score regardless of the variable.

Initial Consonants Correct

The mean initial consonants correct scores are shown in Figure 5 for each level of talker severity and presentation number. For talker severity level, averaged across presentation number, initial consonants correct scores ranged from 14.27 (SD = 2.01) for Child 4 to 18.63 (SD = 1.46) for Child 2. For presentation number, averaged across talker severity level, initial consonants correct scores ranged from 16.04 (SD = 2.22) for the first presentation to 16.58 (SD = 2.24) for the third presentation. Mean initial consonants correct scores increased across presentation number in all cases, with the exception of Child 2 between presentations 2 and 3. The greatest increase was for Child 2 (.73 initial consonants) and the smallest increase was for Child 1 (.28 initial consonants). The effect size calculation is shown in Appendix O.

A 4 X 3 ANOVA with repeated measures on both factors was used to examine the main effects of talker severity and presentation number as well as the possible interaction between these two factors. Sphericity could not be assumed for the interaction between talker severity and presentation number. The Greenhouse-Geisser F test was used in this case. There was a significant main effect of talker severity (F (3, 96) = 53.60, $p < .001$) and a significant main effect of presentation number (F (2, 64) = 21.54, $p < .001$). The interaction between talker severity and presentation number was not significant (F (4.63, 148.11) = 1.35, p = .249).

Pairwise comparisons were conducted to determine the nature of the main effect of talker severity. There was a significant difference between Child 1 and 2 ($p < .001$), Child 1 and 4 ($p < .001$), Child 2 and 3 ($p < .001$), Child 2 and 4 ($p < .001$), and Child 3 and 4 ($p < .001$) such that Child 2 had the highest scores, followed by Child 1 and 3, and Child 4 had the lowest scores. For the effect of talker severity, the smallest difference found to be significant was between child 3 and 4 (a mean difference of 1.78). The effect size for this difference is .65, which is considered to be medium-large (Cohen, 1988). The effect size calculation is shown in Appendix O.

Pairwise comparisons were conducted to determine the nature of the main effect of presentation number. There was a significant difference between presentation 1 and 2 ($p < .001$), and presentation 1 and 3 ($p < .001$) such that scores were lowest for the first presentation. For the effect of presentation number, the smallest difference found to be significant was between presentation 1 and 2 (a mean difference of .51). The effect size for this difference is .16, which is considered to be trivial (Cohen, 1988). The effect size for the difference between presentation 1 and 3 (a mean difference of .54) is .17, which is also considered to be trivial (Cohen, 1988). The effect size calculations are shown in Appendix O.

Table 5. *Phonetic Accuracy Scores for Talker Severity and Presentation Number*

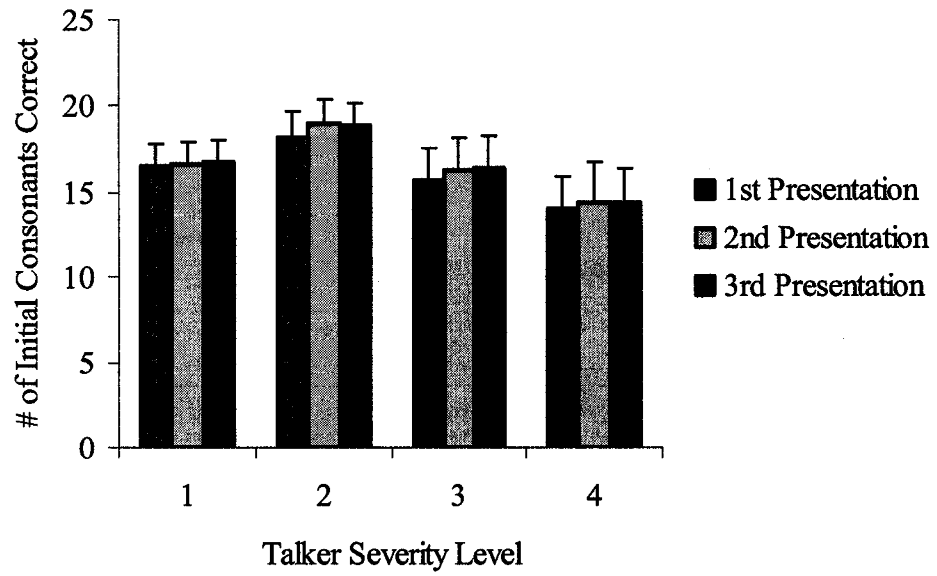
Measure	Talker Severity	1 Mean	1 (SD)	2 Mean	2 (SD)	3 Mean	3 (SD)	**Mean (SD)**
		Presentation Number						
Initial Consonant[a]	1	16.48	(1.30)	16.61	(1.32)	16.76	(1.25)	**16.62 (1.28)**
	2	18.09	(1.55)	18.97	(1.38)	18.82	(1.33)	**18.63 (1.46)**
	3	15.61	(1.87)	16.21	(1.85)	16.33	(1.85)	**16.05 (1.86)**
	4	13.97	(1.86)	14.42	(2.24)	14.42	(1.94)	**14.27 (2.01)**
	Mean (SD)	**16.04**	**(2.22)**	**16.55**	**(2.36)**	**16.58**	**(2.24)**	
Vowel[b]	1	19.48	(1.33)	19.39	(1.46)	19.88	(1.54)	**19.59 (1.44)**
	2	16.21	(1.56)	16.73	(1.31)	16.82	(1.31)	**16.59 (1.41)**
	3	17.58	(2.39)	17.39	(2.47)	17.73	(2.11)	**17.57 (2.31)**
	4	11.48	(1.86)	11.97	(1.59)	12.21	(1.63)	**11.89 (1.71)**
	Mean (SD)	**16.19**	**(3.47)**	**16.37**	**(3.24)**	**16.66**	**(3.26)**	
Final Consonant[c]	1	14.97	(1.74)	15.24	(1.62)	15.30	(1.67)	**15.17 (1.67)**
	2	15.09	(1.63)	15.91	(1.23)	15.91	(1.40)	**15.64 (1.47)**
	3	14.27	(2.00)	14.79	(1.52)	15.00	(1.54)	**14.69 (1.71)**
	4	5.12	(1.39)	6.03	(1.47)	6.24	(1.46)	**5.80 (1.50)**
	Mean (SD)	**12.36**	**(4.53)**	**12.99**	**(4.31)**	**13.11**	**(4.27)**	
Syllable Shape[d]	1	21.39	(1.56)	21.30	(1.51)	21.30	(1.19)	**21.33 (1.41)**
	2	18.64	(1.65)	18.82	(1.67)	19.15	(1.77)	**18.87 (1.69)**
	3	19.36	(2.43)	19.15	(2.82)	18.88	(2.79)	**19.13 (2.67)**
	4	13.00	(2.00)	12.94	(1.77)	12.94	(1.78)	**12.96 (1.83)**
	Mean (SD)	**18.10**	**(3.67)**	**18.05**	**(3.69)**	**18.07**	**(3.68)**	

Note. Bold Means and Standard Deviations are derived from each listener's mean scores, collapsed across one variable (e.g., for the three different presentations, 132 scores were used, for the four talker severity levels, 99 scores were used).

Note. Statistically significant differences for initial consonants correct scores were found for talker severity ($p \leq .008$) (1 vs. 2, 1 vs. 4, 2 vs. 3, 2 vs. 4 and 3 vs. 4); and for presentation number ($p \leq .017$) (1 vs. 2 and 1 vs. 3). Statistically significant differences for vowels correct scores were found for talker severity ($p \leq .008$) (1 vs. 2, 1 vs. 3, 1 vs. 4, 2 vs. 4) and 3 vs. 4; and for presentation number ($p \leq .017$) (1 vs. 3 and 2 vs. 3). Statistically significant differences for final consonants correct scores were found for talker severity ($p \leq .008$) (1 vs. 4, 2 vs. 3, 2 vs. 4 and 3 vs. 4); and for presentation number ($p \leq .017$) (1 vs. 2 and 1 vs. 3). Statistically significant differences for syllable shapes correct scores were found for talker severity ($p \leq .008$) (1 vs. 2, 1 vs. 3, 1 vs. 4, 2 vs. 4 and 3 vs. 4).

[a]Mean number of initial consonants correct (25 total) [b]Mean number of vowels correct (25 total) [c]Mean number of final consonants correct (21 total) [d]Mean number of syllable shapes correct (25 total).

Figure 5. Listeners' initial consonants correct scores for talker severity and presentation number.



Note. Standard Deviations shown as error bars.

Vowels Correct

The mean vowels correct scores are shown in Figure 6 for each level of talker severity and presentation number. For talker severity level, averaged across presentation number, vowels correct scores ranged from 11.89 (SD = 1.71) for Child 4 to 19.59 (SD = 1.44) for Child 1. For presentation number, averaged across talker severity level, vowels correct scores ranged from 16.19 (SD = 3.47) for the first presentation to 16.66 (SD = 3.26) for the third presentation. Mean vowels correct scores increased across presentation number in all cases with the exception of Child 1 and 3 between presentations 1 and 2. The greatest increase was for Child 4 (.73 vowels) and the lowest increase was for child 3 (.15 vowels).

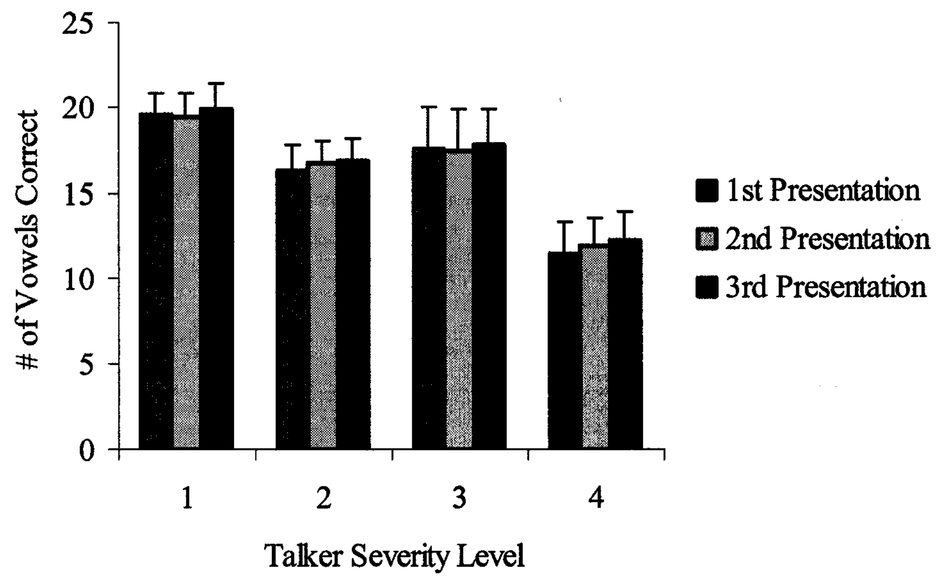
A 4 X 3 ANOVA with repeated measures on both factors was used to examine the main effects of talker severity and presentation number as well as the possible interaction between these two factors. Sphericity could not be assumed for talker severity or presentation number. The Greenhouse-Geisser *F* test was used in these cases. There was a significant main effect of talker severity (F (2.35, 75.20) = 168.67, $p < .001$) and a significant main effect of presentation number (F (1.53, 49.08) = 6.09, $p = .008$). The interaction between talker severity and presentation number was not significant (F (6, 192) = 1.51, $p = .196$).

Pairwise comparisons were conducted to determine the nature of the main effect of talker severity. There was a significant difference between Child 1 and 2 ($p < .001$), Child 1 and 3 ($p < .001$), Child 1 and 4 ($p < .001$), Child 2 and 4 ($p < .001$), and Child 3 and 4 ($p < .001$) such that Child 1 had the highest scores, followed by Child 2, 3, and 4. For the effect of talker severity, the smallest difference found to be significant was between Child 1 and 3 (a mean difference of 2.02). The effect size for this difference is .74, which is considered to be medium-large (Cohen, 1988). The effect size calculation is shown in Appendix O.

Pairwise comparisons were conducted to determine the nature of the main effect of presentation number. There was a significant difference between presentation 1 and 3 ($p = .008$), and presentation 2 and 3 ($p = .008$) such that scores improved after the second presentation of the word. For the effect of presentation number, the smallest difference found to be significant was between presentation 2 and 3 (a mean difference of .29). The effect size for this difference is .06, which is considered to be trivial (Cohen, 1988). The effect size for the difference between presentation 1 and 3 (a mean difference of .47) is

.10, which is also considered to be trivial (Cohen, 1988). The effect size calculations are shown in Appendix O.

Figure 6. Listeners' vowels correct scores for talker severity and presentation number.



Note. Standard Deviations shown as error bars.

Final Consonants Correct

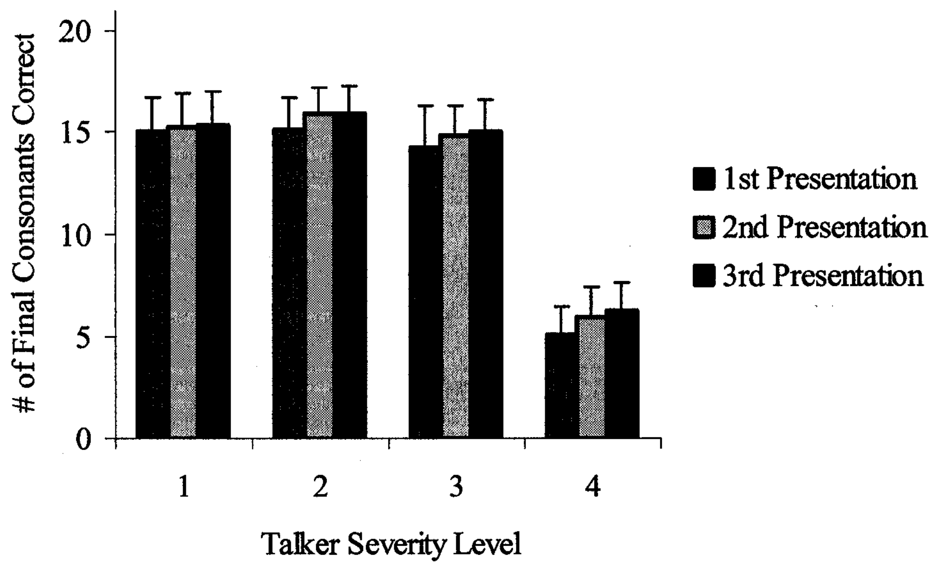
The mean final consonants correct scores are shown in Figure 7 for each level of talker severity and presentation number. For talker severity level, averaged across presentation number, final consonants correct scores ranged from 5.80 (SD = 1.50) for Child 4 to 15.64 (SD = 1.47) for Child 2. For presentation number, averaged across talker severity level, final consonants correct scores ranged from 12.36 (SD = 4.53) for the first presentation to 13.11 (SD = 4.27) for the third presentation. Mean final consonants correct scores increased across presentations for all cases with the exception of Child 2 between presentations 2 and 3. The greatest increase was for Child 4 (1.12 final consonants) and the lowest increase was for Child 1 (.33 final consonants).

A 4 X 3 ANOVA with repeated measures on both factors was used to examine the main effects of talker severity and presentation number as well as the possible interaction between these two factors. Sphericity could not be assumed for presentation number. The Greenhouse-Geisser *F* test was used in this case. There was a significant main effect of talker severity (F (3, 96) = 436.43, $p < .001$) and a significant main effect of presentation number (F (1.40, 44.88) = 19.69, $p < .001$). The interaction between talker severity and presentation number was not significant (F (6, 192) = 1.34, $p = .241$).

Pairwise comparisons were conducted to determine the nature of the main effect of talker severity. There was a significant difference between Child 1 and 4 ($p < .001$), Child 2 and 3 ($p = .005$), Child 2 and 4 ($p < .001$), and Child 3 and 4 ($p < .001$). Child 1 and 2 had the highest scores, followed by Child 3, and then Child 4. For the effect of talker severity, the smallest difference found to be significant was between Child 2 and 3 (a mean difference of .95). The effect size for this difference is .42, which is considered to be small-medium (Cohen, 1988). The effect size calculations are shown in Appendix O.

Pairwise comparisons were conducted to determine the nature of the main effect of presentation number. There was a significant difference between presentation 1 and 2 ($p < .001$), and presentation 1 and 3 ($p < .001$) such that scores improved with more repetitions of the word. For the effect of presentation number, the smallest difference found to be significant was between presentation 1 and 2 (a mean difference of .63). The effect size for this difference is .10, which is considered to be trivial (Cohen, 1988). The effect size for the difference between presentation 1 and 3 (a mean difference of .75) is .12, which is also considered trivial (Cohen, 1988). The effect size calculations are shown in Appendix O.

Figure 7. Listeners' final consonants correct scores for talker severity and presentation number.



Note. Standard Deviations shown as error bars.

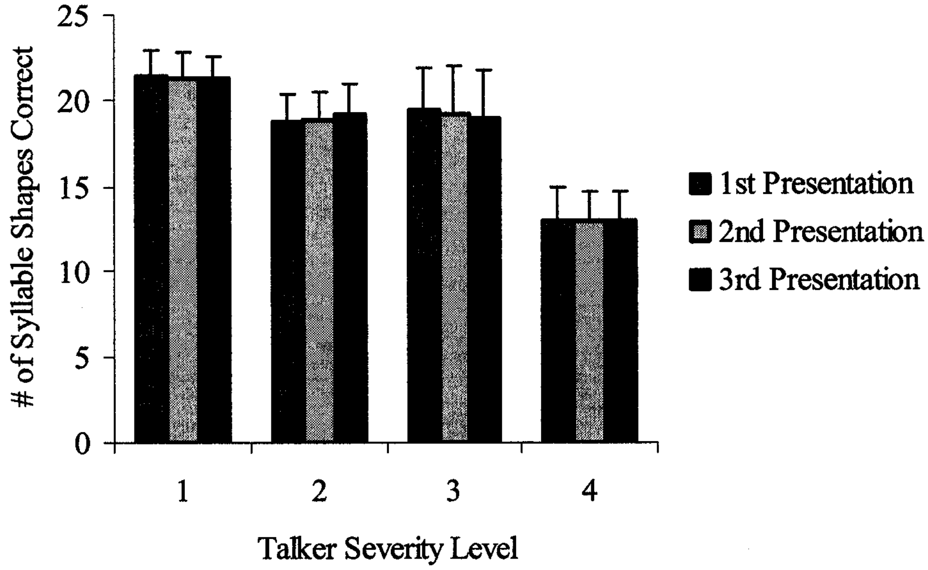
Syllable Shapes Correct

The mean syllable shapes correct scores are shown in Figure 8 for each level of talker severity and presentation number. For talker severity level, averaged across presentation number, syllable shapes correct scores ranged from 12.96 (SD = 1.83) for Child 4 to 21.33 (SD = 1.44) for Child 1. For presentation number, averaged across talker severity level, syllable shapes correct scores ranged from 18.05 (SD = 3.69) for the second presentation to 18.10 (SD = 3.67) for the first presentation. Mean syllable shapes correct scores increased across presentation number for Child 2 only (.51 syllables). For Child 1, 3 and 4, syllable shapes correct scores decreased from the first to the third presentation (range of decrease from .06 to .48 syllables).

A 4 X 3 ANOVA with repeated measures on both factors was used to examine the main effects of talker severity and presentation number as well as the possible interaction between these two factors. Sphericity could not be assumed for talker severity, presentation number, nor for the interaction. The Greenhouse-Geisser F test was used. There was a significant main effect of talker severity (F (2.15, 68.83) = 164.03, $p < .001$). The main effect of presentation number was not significant (F (1.21, 38.63) = 0.50, p = 0.867). The interaction between talker severity and presentation number was not significant (F (4.46, 142.80) = 0.94, $p = .452$). The mean syllable shapes correct scores are shown in Figure 8 for each level of talker severity and presentation number.

Pairwise comparisons were conducted to determine the nature of the main effect of talker severity. There was a significant difference between Child 1 and 2 ($p < .001$), Child 1 and 3 ($p < .001$), 1 and 4 ($p < .001$), Child 2 and 4 ($p < .001$), and Child 3 and 4 ($p < .001$) such that Child 1 had the highest scores, followed by Child 2, 3 and 4. For the effect of talker severity, the smallest difference found to be significant was between Child 1 and 3 (a mean difference of 2.20). The effect size for this difference is .73, which is considered to be medium-large (Cohen, 1988). The effect size calculation is shown in Appendix O.

Figure 8. Listeners' syllable shapes correct scores for talker severity and presentation number.



Note. Standard Deviations shown as error bars.

Effect of Talker Severity and Presentation Number on Response Time

Group means and standard deviations of listeners' median response times for the 33 listeners for each level of talker severity and each presentation number are reported in Table 6 and shown in Figure 9. Data presented represent the number of seconds it took listeners to respond to the word as measured from the time lapse between the beginning of the word and the listener entering the first letter, corrected for the individual listener's average typing speed per letter. For severity level, averaged across presentation number, response times ranged from 1.97 seconds (SD = 0.49) for Child 1 to 2.45 seconds (SD = 0.65) for Child 4. For presentation number, averaged across severity level, response times ranged from 2.55 seconds (SD = 0.49) for the first presentation to 1.85 seconds (SD = 0.54) for the third presentation. Across children response times decreased by between .64 and .79 seconds from the first to the third presentation.

A 4 X 3 ANOVA with repeated measures on both factors was computed to test the hypotheses that words spoken by more severe children would take longer to identify and that response times would decrease with subsequent repetitions of the word. It also tested the interaction between these two variables. Sphericity could not be assumed for talker severity and presentation number. The Greenhouse-Geisser F test was used in these cases. There was a significant main effect of talker severity (F (2.24, 71.53) = 43.94, $p <$ .001) and a significant main effect of presentation number (F (1.40, 44.76) = 116.53, $p <$ .001). The interaction between child and presentation was not significant (F (6, 192) = 1.30, $p = .259$).

Pairwise comparisons were conducted to determine the nature of the main effect of talker severity. There was a significant difference between Child 1 and 2 ($p < .001$), 1 and 4 ($p < .001$), Child 2 and 3 ($p < .001$), Child 2 and 4 ($p < .001$) and Child 3 and 4 ($p < .001$) such that response times were fastest for Child 1 and 3, followed by Child 2 and then Child 4. For the effect of talker severity, the smallest difference found to be significant was between Child 2 and 3 (a mean difference of .17 seconds). The effect size for this difference is .23 which is considered to be small (Cohen, 1988). The effect size calculation is shown in Appendix O.

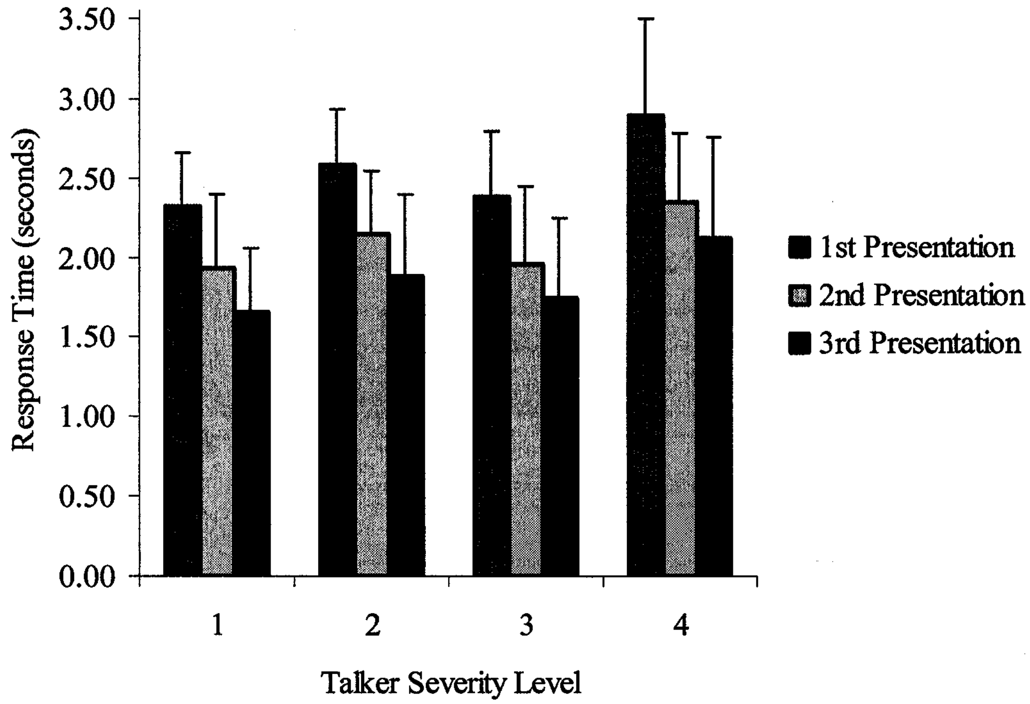
Pairwise comparisons were conducted to determine the nature of the main effect of presentation number. There was a significant difference between presentation 1 and 2 ($p = .014$), presentation 1 and 3 ($p < .001$), and presentation 2 and 3 ($p = .013$) such that accuracy increased with more repetitions of the word. For the effect of presentation number, the smallest difference found to be significant was between presentation 2 and 3 (a mean difference of .24 seconds). The effect size for this difference is .33 which is considered to be small-medium (Cohen, 1988). The effect size for the difference between presentation 1 and 3 (a mean difference of .70 seconds) is .96, which is considered to be large (Cohen, 1988). The effect size calculations are shown in Appendix O.

Table 6. *Mean of Listeners' Response Times for Talker Severity and Presentation Number*

	Presentation Number							
	1		2		3			
Talker Severity	Mean	(SD)	Mean	(SD)	Mean	(SD)	**Mean**	**(SD)**
1	2.33	(0.34)	1.94	(0.46)	1.65	(0.40)	**1.97**	**(0.49)**
2	2.58	(0.35)	2.14	(0.41)	1.88	(0.52)	**2.20**	**(0.51)**
3	2.39	(0.41)	1.95	(0.49)	1.75	(0.49)	**2.03**	**(0.53)**
4	2.90	(0.61)	2.34	(0.44)	2.11	(0.64)	**2.45**	**(0.65)**
Mean (SD)	**2.55**	**(0.49)**	**2.09**	**(0.48)**	**1.85**	**(0.54)**		

Note. Mean of Median Response Time Scores are reported in seconds. Bold Means and Standard Deviations were derived from each listener's median score, collapsed across one variable (e.g., for the three different presentations, 132 scores were used, for the four talker severity levels, 99 scores were used).
Note. Statistically significant differences for response time were found for talker severity ($p \leq .008$) (1 vs. 2, 1 vs. 4, 2 vs. 3, 2 vs. 4 and 3 vs. 4); and for presentation number ($p \leq .017$) (1 vs. 2, 1 vs. 3 and 2 vs. 3).

Figure 9. Listeners' median response times for talker severity and presentation number.



Note. Standard deviations shown as error bars.

Effect of Talker Severity on Effort Ratings

Mean listener effort ratings for each listener's direct magnitude estimation (DME) ratings were computed for the 10 words rated for each child. Group means and standard deviations for listener effort by talker severity are reported in Table 7 and shown in Figure 10. Mean ratings ranged from 81.65 (SD = 30.40) for Child 1 to 163.52 (SD = 80.86) for Child 4.

A one-way repeated measures ANOVA with four levels was used to test the hypothesis that listeners would rate more severe children as harder to understand (e.g., Child 4 would be the hardest to understand, followed by child 3, 2 and 1). Sphericity could not be assumed. The Greenhouse-Geisser F test was used. There was a significant main effect of child (F (1.28, 40.98) = 32.32, $p < .001$).

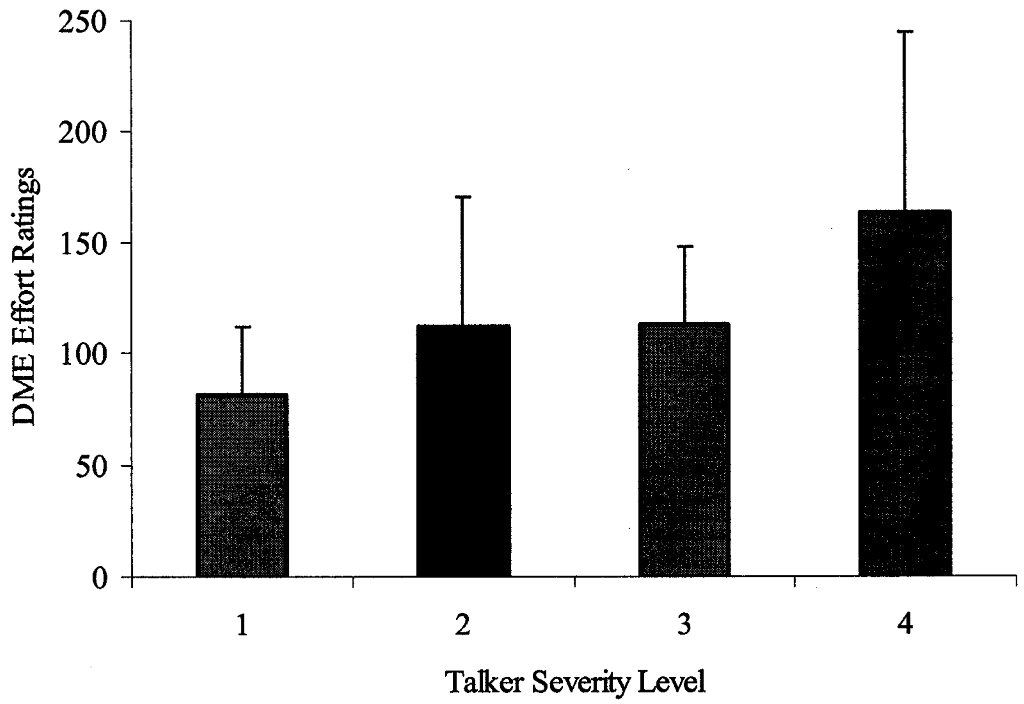
Pairwise comparisons were calculated to determine the nature of the main effect of talker severity. There was a significant difference between Child 1 and 2 ($p = .002$), Child 1 and 3 ($p < .001$), Child 1 and 4 ($p < .001$), Child 2 and 4 ($p < .001$) and Child 3 and 4 ($p < .001$) such that less effort was reported for child 1, followed by Child 2 and 3 and the most effort was reported for Child 4. The smallest difference found to be significant was between Child 1 and 2 (a mean difference of 30.51). The effect size for this difference is .47 which is considered to be small-medium (Cohen, 1988). The effect size calculation is shown in Appendix O.

Table 7. *Group Means of Listeners' DME Ratings for Talker Severity*

Talker Severity	Mean	(SD)
1	81.65	(30.40)
2	112.16	(57.72)
3	112.50	(35.20)
4	163.52	(80.86)
Mean	**117.46**	**(61.73)**

Note. Bold Mean and Standard Deviation were derived from each listener's mean score. Therefore, a total of 132 scores were used.
Note. Statistically significant differences for DME ($p \leq .008$) were found for 1 vs. 2, 1 vs. 3, 1 vs. 4, 2 vs. 4 and 3 vs. 4.

Figure 10. Listeners' DME ratings for talker severity.

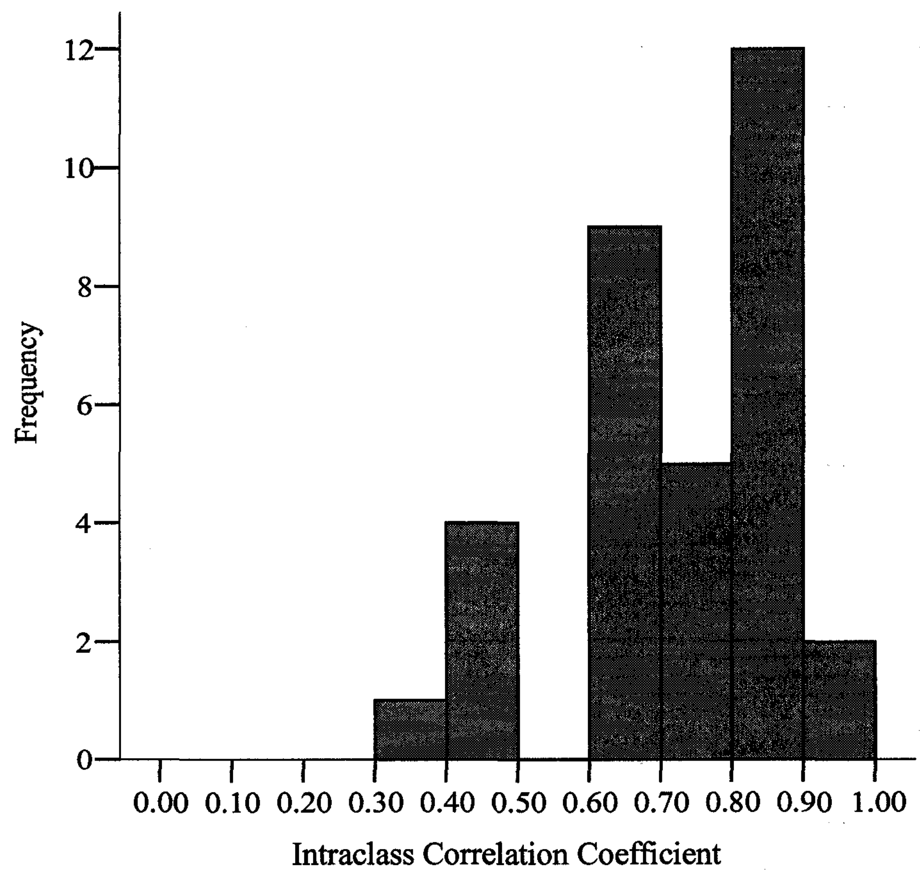


Note. Standard deviations shown as error bars.

Reliability of Listeners' DME Effort Ratings

Intraclass correlation coefficients (ICCs) were used to estimate the reliability of listeners' direct magnitude estimates of listening effort. One coefficient was computed for each listener for the eight words that were rated twice. For the 33 listeners, the mean ICC was .72, the standard deviation was .15 and the range was .61 (Min = .37, Max = .98). Nineteen listeners had ICCs above .70, nine had ICCs between .60 and .70, and five had ICCs below .50. Figure 11 shows the frequency distribution of all 33 listeners' reliability scores.

Figure 11. Frequency distribution of intraclass correlation coefficient for reliability of DME effort ratings for 33 listeners.



Relationships between Objective and Subjective Measures of Listening Effort

The relationships between objective and subjective measures of listening effort were evaluated using regression analyses. The relationship between listeners' word identification scores and median response times was investigated by regressing listeners' word identification scores, averaged across child, on their median response times, averaged across child, for the first presentation only. These results are shown in Figure 12. The relationship ($r = -.42$) between accuracy and response time was significant ($t = 2.55, p = .016$) with response time increasing as accuracy decreased.

The relationship between listeners' word identification scores and DME effort ratings was investigated by regressing listeners' word identification scores, averaged across child for the first presentation only of the 10 words used in the DME, on their DME scores, averaged across child. These results are shown in Figure 13. This relationship ($r = -.21$), while in the predicted direction, was not significant ($t = 1.21, p = .234$).

The relationship between listeners' response times and DME effort ratings was investigated by regressing listeners' response times, averaged across child for the first presentation only of the 10 words used in the DME, on their DME scores, averaged across child. These results are shown in Figure 14. This relationship ($r = -.06$) was not in the predicted direction and was not significant ($t = 0.35, p = .727$).

Figure 12. Relationship between word identification scores (predictor variable) and response time (predicted variable) for the 33 listeners.

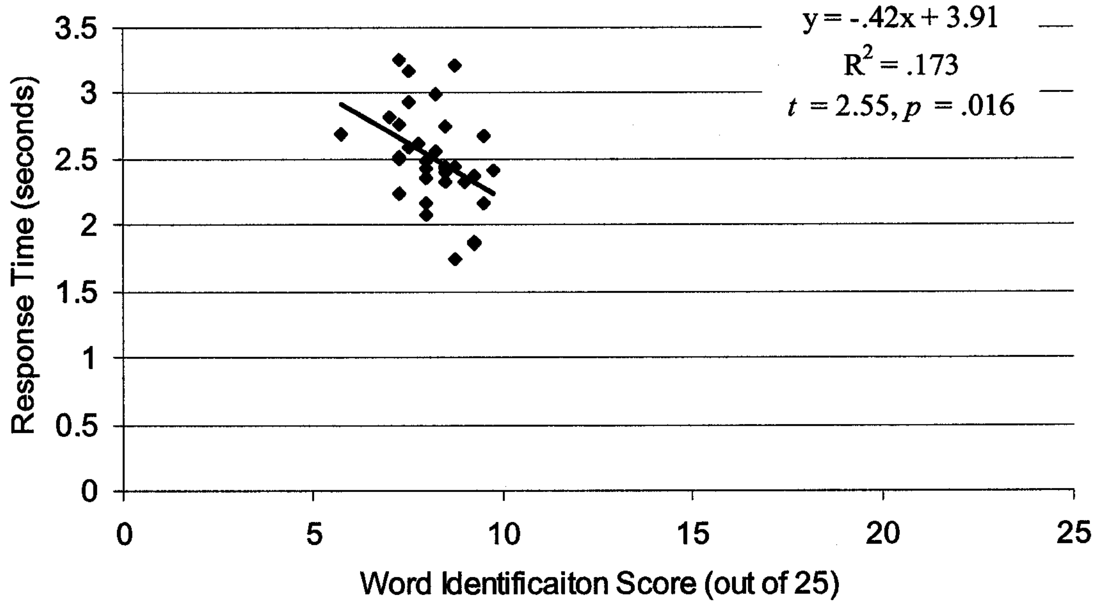


Figure 13. Relationship between word identification scores (predictor variable) and DME effort ratings (predicted variable) for the 33 listeners.

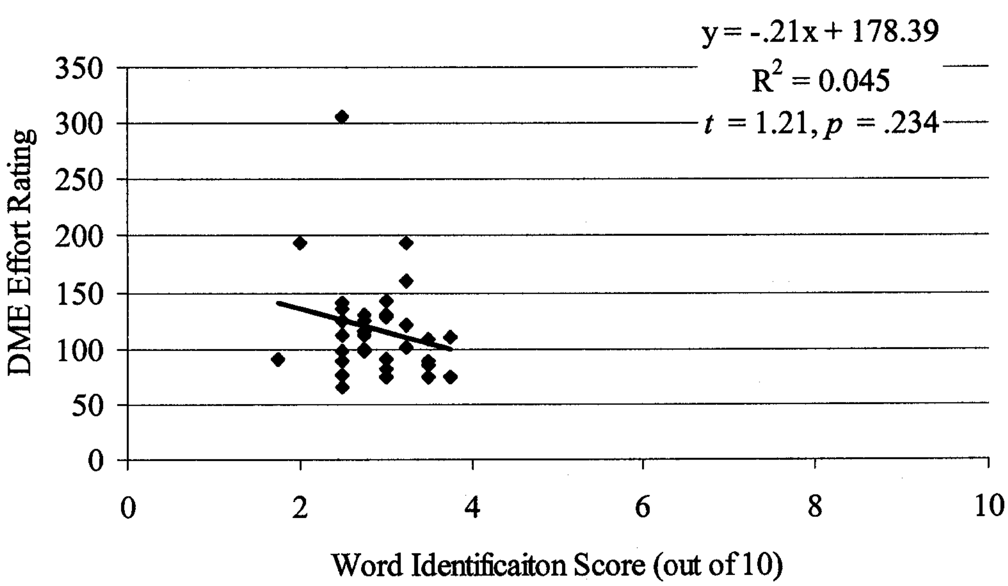
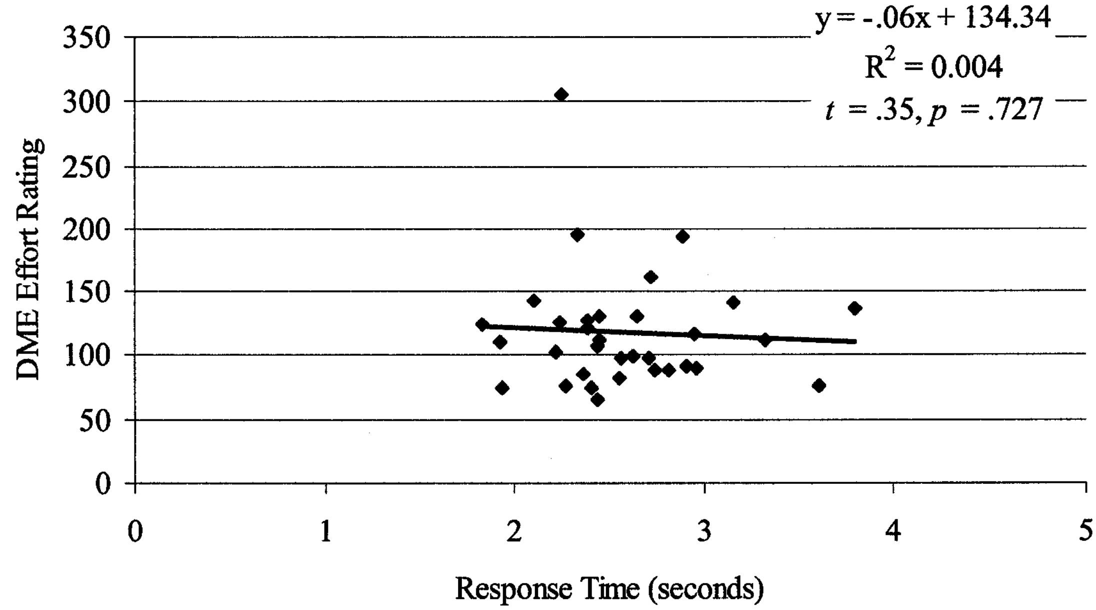


Figure 14. Relationship between response time (predictor variable) and DME effort ratings (predicted variable) for the 33 listeners.



Listeners' Beliefs about Listening Task 1

Qualitative analyses were conducted for each of the three questions that listeners were asked at the end of their listening session regarding their experiences on Listening Task 1. For each question, the responses were grouped by common themes. For the first two questions that asked about what made words easier or more difficult to understand, three themes were identified: Articulation/Voice Characteristics, Word Characteristics and Other. Listener's responses are summarized in Table 8 for question 1, "What kinds of things made it easy to understand these children's words?". Listeners frequently identified increased speaking effort and clarity as important factors in making the word easy to understand. Listener's responses are summarized in Table 9 for question 2, "What kinds of things made it difficult to understand these children's words?". Listeners frequently identified quite/mumbling/whispered speech and the background noise as factors contributing to making the word difficult to understand.

For the third question, "When a word was difficult to understand, what kinds of things did you do to try and understand the word?", seven themes were identified from listeners' responses: Tried to Find out Unknown from Known Using Phonetic Information, Repetition, Body/Postural Adjustments, Specific Characteristics of the Word, Listened for Particular Parts, Increased Mental Effort and Other. Listener's responses are summarized in Table 10. Listeners frequently reported saying the word to themselves again, making postural adjustments and thinking about what real word the stimulus sounded closest to in order to identify it.

Table 8. *Theme Analysis for Listeners' Beliefs - Question 1: What kinds of things made it easy to understand these children's words?*

Response	Number of Respondents[a]
Articulation/Voice Characteristics	53 (total)
A lot of speaking effort/clear on last sound	16
A lot of speaking effort/clear on first sound	7
Clear/strong pronunciation	7
Slower/drawn out word	6
Louder	6
Obstruents (e.g., /k/)	3
High pitch	2
Better combination of letters [sounds] for mouth movement	1
Longer words	1
More repetitions	1
Said without pauses between sounds	1
Tone of voice	1
Word Characteristics	4 (total)
Common Word	2
Distinct words (few phonological neighbours)	2
Other	2 (total)
Words where the background noise interfered less	1
Time/Increased exposure to speech of the children	1

[a]Total number of listeners who gave that response (may total more than 33 if listeners gave multiple responses)

Table 9. *Theme Analysis for Listeners' Beliefs - Question 2: What kinds of things made it difficult to understand these children's words?*

Response	Number of Respondents[a]
Articulation/Voice Characteristics	48 (total)
Quiet/mumbling/whispered	16
Short words or quickly said words	8
Didn't say the end of the word/trailed off at end	7
Poor pronunciation/missed sounds	5
Hard to distinguish sounds (e.g., f, p, d, b, vowels)	3
Deep/scratchy/croaky voice	2
No vowels/couldn't hear a vowel	2
Lower pitch	1
Stuttered	1
Sounds were separated by a break	1
Younger sounding voice	1
Consonant Clusters (e.g., ts)	1
Word Characteristics	7 (total)
Many similar sounding words	4
Couldn't find a word to match	2
Started with a vowel	1
Other	12 (total)
Background noise	10
If second guessed first answer	1
Couldn't figure out at least one letter right away	1

[a]Total number of listeners who gave that response (may total more than 33 if listeners gave multiple responses)

Table 10. *Theme Analysis for Listeners' Beliefs - Question 3: When a word was difficult to understand, what kinds of things did you do to try and understand the word?*

Response	Number of Respondents[a]
Tried to Find out Unknown from Known Using Phonetic Information	21 (total)
What sounds is it close to?	10
Tried to figure out the rest from 1 or 2 sounds known	3
Wrote down the letters, then tried to match a word	2
Picked 2 sounds that it might be, then listened again	2
Tried to find any word that might match	1
Broke it down into the individual sounds heard	1
Tried to rhyme what I heard to a word	1
Figured out middle from sounds known	1
Repetition	12 (total)
Repeated in my mind	8
Said it aloud	2
Compared previous guesses with what I heard the next time	1
Just entered something to be able to hear it again	1
Body/Postural Adjustments	8 (total)
Leaned closer	4
Closed my eyes the next time	2
Turned ear sideways	1
Listened with other ear the next time	1
Specific to Characteristics of the Word	7 (total)
Is it an appropriate word for a child?	5
Tried to use context (e.g., people talking in a restaurant)	2
Listened for Particular Parts	6 (total)
Listened hard to first and last letter	3
Focused on the last sound the next time	2
Listened for key vowels and consonants	1
Increased Mental Effort	
Listened harder/focused more the next time	4
Other	3 (total)
Remembered the priming list	2
Guessed	1

[a]Total number of listeners who gave that response (may total more than 33 if listeners gave multiple responses)

DISCUSSION

The purpose of this study was to investigate the effect of signal dose (repeated presentations) on listeners' abilities to understand words spoken by children who varied in the level of severity of their dysarthria. The effects of these factors (number of presentations and level of talker severity) were investigated for two dependent measures: word identification scores and response times. To increase generalizability of the findings, these measures were obtained using a signal-to-noise ratio (SNR) within the range of a typical classroom environment (+10 dB). As predicted, significant main effects of talker severity and presentation number were found for both word identification scores (i.e., the most severe talker had significantly lower word identification scores and additional presentations had significantly higher word identification scores) and response times (i.e., the most severe talker had significantly longer response times and repeated presentations had shorter response times). However, a significant interaction of talker severity and presentation number was not found for either word identification scores or response times. In addition, post-hoc testing conducted to determine the nature of the main effect of talker severity on these dependent variables revealed several unexpected findings regarding the range and magnitude of differences in scores across the four levels of severity.

This study also found a significant effect of talker severity on direct magnitude estimates (DME) of listening effort. As predicted, the most severe talker was assigned the highest effort ratings and the least severe talker was assigned the lowest effort ratings. As expected, regression of word identification scores on response times revealed a significant (moderately negative) relationship. Unexpectedly, significant relationships between word identification scores and listening effort ratings and between response times and listening effort ratings were not found. However, testing of these relationships was limited by the design of the study. Qualitative analysis of listeners' perceptions of their experiences in Listening Task 1 provided insights into how they coped with trying to understand the words and what kinds of things made some words easy or difficult to understand. The findings from the quantitative and qualitative analyses are discussed in the following sections.

Effect of Talker Severity on Measures of Listener Behaviour

Effect of Talker Severity on Word Identification Scores

The factor of talker severity level was based on the four children's scores on a sentence intelligibility measure, where each child's score was at least 10% different from the scores of the adjacent children. These scores represented the mean of three listeners' responses obtained in an ideal listening environment where listeners judged all sentences spoken by one child in a block (segregated speaker condition). Children's sentence intelligibility scores ranged from 38% to 86.5%. In Listening Task 1, the dependent variable, word identification scores, was based on listeners' identification of 25 single words spoken by each child, balanced for phonetic content, neighbourhood density and word frequency across the four children and presented in a typical SNR, with the 100 words in random order. As noted, while the main effect of talker severity was significant, the specific results were not as predicted. On Listening Task 1, Child 1's word identification scores did not differ from those of Child 2 or 3, the greatest difference among these three children's word identification scores was 5.5% (between Child 2 and

3) and the range of scores was compressed to the moderately severe to profound region of severity (i.e., Child 2 – 45.1% to Child 4 – 9.7%). Possible explanations for these findings follow.

Reduction in Size of Differences in Word Identification Scores on Listening Task 1 across Child 1, 2 and 3

As noted, Listening Task 1 word identification scores for Child 1 did not differ from Child 2 and 3 and the range of scores for these three talkers was only 5.5% (Child 2 – 45.1% to Child 3 – 39.6%) compared to the *TOCS* sentence intelligibility scores for these children, which differed by at least 10%, with a range of 31.5% (Child 1 – 86.5% to Child 3 –55.0%). This result must be accounted for at least in part by the combined effect of the 25 stimulus words selected for the children in Listening Task 1, the mixed talker design and the added noise. In Listening Task 1, listeners heard single syllable words randomized across the four children in + 10 dB SNR. The *TOCS* sentence intelligibility scores were obtained using listeners who heard all the sentences spoken by one child (segregated talker condition), in ideal (i.e., quiet) listening conditions. The size of each of these possible influences can not be determined directly from this study because they were not treated as independent variables, i.e., utterance length (sentence versus single word), talker presentation condition (segregated versus mixed) and SNR level (ideal versus typical classroom environment). However, an estimate of their potential influence is possible based on existing information available about these children's word identification scores on the *TOCS* word intelligibility measure, obtained under the same conditions as their word identification scores on the *TOCS* sentence intelligibility measure. Like the *TOCS* sentence intelligibility measure used to classify the children by severity of speech disorder, three listeners judged each child's word recordings for the *TOCS* word intelligibility measure in ideal listening conditions with listeners identifying words spoken by a single child (i.e., segregated condition). Word identification scores were calculated for the subset of 25 words, which served as the stimuli for each child in Listening Task 1, based on previous listeners' judgments of the full *TOCS* word intelligibility measure for comparison[19]. The results of these calculations revealed that the *TOCS* word intelligibility scores calculated for the subset of 25 words for each child were very similar: Child 1 had a score of 68%, Child 2 a score of 67% and Child 3 a score of 67%[20]. While Child 1, 2 and 3 had different levels of speech disorder severity based on *TOCS* sentence intelligibility scores, this was not reflected in their *TOCS* single word intelligibility scores for the subset of 25 words, balanced for phonetic context, neighbourhood density and word frequency and judged by listeners in quiet and a segregated talker condition. Only two of four levels used to classify talker severity on the *TOCS* sentence intelligibility measure are represented by the *TOCS* single word intelligibility scores for the subset of 25 words: mild-moderate (Child 1, 2 and 3) and severe (Child 4).

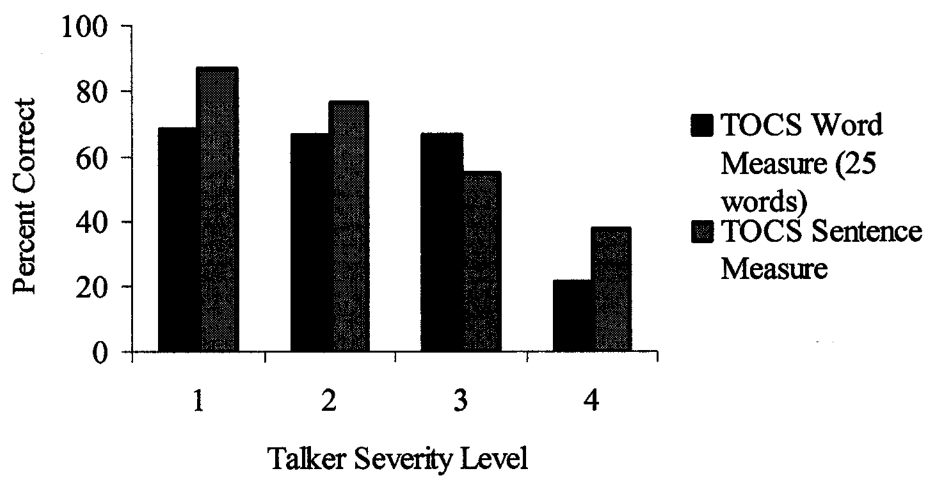
Gordon-Brannan and Hodson (2000) reported that word identification scores for imitated sentences were about 10% higher than word identification scores for imitated words, regardless of severity of speech disorder for children with a range of phonological delays/disorders of unknown origin. For the child talkers in the present study, in comparable listening conditions, the *TOCS* single word intelligibility scores for the subset

[19] Scores were derived from responses from up to six listeners for each child, with three listeners identifying each word. Listeners heard 78 words spoken by the same child and wrote down what they thought the child was saying. Scores are an average of the three responses to each word (correct or incorrect).

[20] Child 4 had a score of 21%.

of 25 words were on average 8% lower than their *TOCS* sentence intelligibility scores. However, this difference varied greatly depending on the child. For example, Child 1's word scores were 18.5% lower than his sentence scores and child 3's word scores were 18% higher than her sentence scores. *TOCS* word intelligibility scores for the subsets of 25 words and the sentence intelligibility measure obtained in the same listening conditions are shown in Figure 15. These findings suggest that the relationship between sentence and single word intelligibility scores is nonlinear, at least across these four children with dysarthria. Further analysis of the speech of Child 3 is recommended to attempt to determine why her single word and sentence intelligibility measures show a reversed pattern (higher for single words, lower for sentences), compared to the other three children and the pattern reported by Gordon-Brannan and Hodson (2000).

Figure 15. Word identification scores for the *TOCS* word intelligibility (subset of 25 words) and sentence measures obtained in ideal listening conditions.



As noted, the word identification scores for Child 1, 2 and 3 were almost identical in ideal SNR and segregated talker conditions (range of 1%) for the subset of 25 words used in this study. Therefore, it appears that the influence of utterance length and type (sentence versus monosyllabic word) can account for the reduction in differences in the word identification scores across Child 1, 2 and 3 observed in Listening Task 1. The added noise (+ 10 dB SNR) and randomized presentation of stimuli with mixed talkers in Listening Task 1 does not appear responsible for the similarity in word identification scores for Child 1, 2 and 3 on this task. Rather, as well as reducing scores overall, the presence of background noise and mixing the talkers in Listening Task 1 appeared to create slightly more spread in the word identification scores for the sets of 25 word stimuli (range of 5.5%) for Child 1, 2 and 3. Two of the children who had identical *TOCS* word intelligibility scores for the subset 25 words (Child 2 and 3) in an ideal SNR and segregated talker condition obtained scores that were significantly different from each other in Listening Task 1. Three levels of talker severity are represented by the word identification scores obtained in Listening Task 1; moderate-severe (Child 2), severe (Child 1 and 3) and profound (Child 4), based on the ranges of percent words identified correctly used to classify the severity levels for the *TOCS* sentence intelligibility measure. The medium effect size found for the difference of 5.5 % words identified correctly between Child 2 and 3 suggests that this is a meaningful difference in scores between these children (Cohen, 1988) under the conditions of Listening Task 1. Smaller differences existed, but were not found to be significant.

Examination of finer-grained measures of phonetic accuracy, derived from listeners' orthographic responses to Listening Task 1 were conducted to determine if these would reveal greater differences in scores across Child 1, 2 and 3 than the word identification scores. For purposes of the Discussion, the results reported in Table 5 and Figures 5, 6, 7 and 8 in the Results section have been combined into the categories of PPC-ID (combination of initial and final consonants identified correctly – total possible raw score of 46); PVC-ID (percent vowels identified correctly – total possible raw score of 25); and PPC-ID (percent phonemes identified correctly - total possible raw score of 71). These percent scores are reported in Table 11.

Table 11. *Phonetic Accuracy Scores for Each Child Based on Listener's Orthographic Transcription in Listening Task 1*

Talker Severity	PCC-ID	PVC-ID	PPC-ID
Child 1	31.79 (69.1%)	19.59 (78.8%)	51.38 (72.4%)
Child 2	34.27 (74.5%)	16.59 (66.4%)	50.86 (71.6%)
Child 3	30.74 (66.8%)	17.57 (70.3%)	48.31 (68.0%)
Child 4	20.07 (43.6%)	11.89 (47.6%)	31.96 (45.0%)

Note. Raw scores (percent) averaged across presentation number.
PCC-ID max = 46, PVC-ID max = 25, PPC-ID max = 76.

Using Shriberg's severity classification for PCC to classify severity based on PCC-ID[21], Child 1, 2 and 3 are classified as mild-moderate, and Child 4 is classified as severe. These are less severe ratings overall than those based on the word identification scores for Listening Task 1 but still show the same pattern of ordering and similarity in scores for Child 2 (74.5%), Child 1 (69.1%) and Child 3 (66.8 %), with a similar range

[21] PCC scores use narrow transcription. The PCC-ID (as well as PVC-ID and PPC-ID) scores were calculated based on listeners' orthographic transcription.

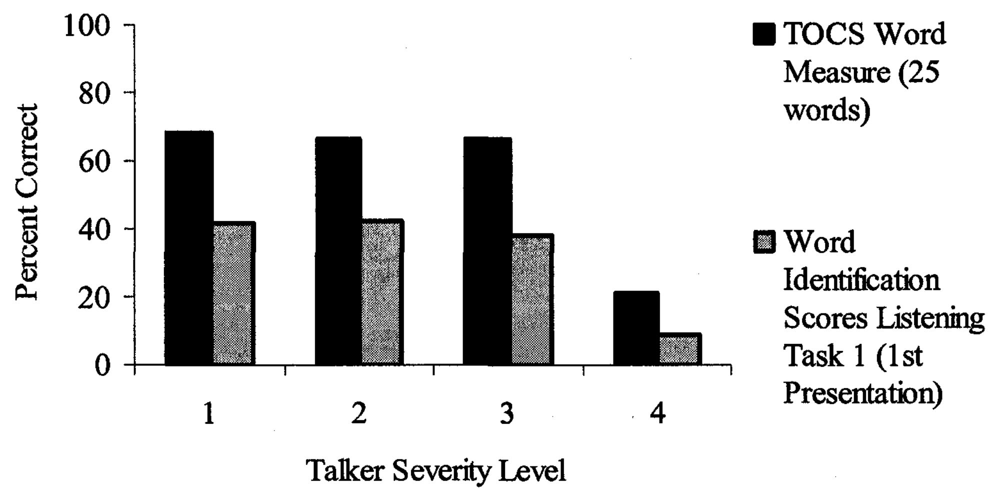
(about 8%). An even smaller range across these three children's scores is evident for PPC-ID (range of about 4%) and a slightly larger range is evident for PVC-ID (about 12%). In summary, while these finer-grained measures of phonetic accuracy showed higher scores and associated less severe classifications of speech disorder than the word identification scores for Listening Task 1, they did not provide a wider distribution of scores for Child 1, 2 and 3 nor an increase in the number of levels of severity classification.

Overall Reduction in Scores

As noted previously, the four children's word identification scores on Listening Task 1 were, on average, 31% lower than their *TOCS* sentence intelligibility scores, which were used to classify talker severity. Child 1 showed the largest difference (45%) and child 3 showed the smallest difference (17%). These differences reflect, at least in part, the combined effect of the 25 monosyllabic words used as stimuli, the mixed talker design and the added noise. The influence of utterance length and type (sentences versus monosyllabic words) on these children's word identification scores was discussed in the previous section. Discussion of the potential influences of the + 10 dB SNR and mixed talker design used in Listening Task 1 follow.

As noted previously, comparison of the differences between the *TOCS* word intelligibility scores for the subset of 25 words in quiet using a segregated talker design and the word identification scores obtained during the first presentation of words in Listening Task 1 for the four children revealed that scores were on average 22% lower for Listening Task 1. Child 3 showed the largest difference (27%) and Child 4 showed the smallest difference (11%). The *TOCS* word intelligibility scores were calculated from three listeners identifying a single presentation of each word. These three listeners heard a single talker and words were presented in quiet. The word identification scores from Listening Task 1 were obtained from 33 listeners identifying the first presentation of each word in a mixed talker design and with words presented in + 10 dB SNR. As shown in Figure 16, despite the difference in procedures between how the *TOCS* word intelligibility scores and the word identification scores in Listening Task 1 were obtained, the scores follow a similar pattern. This overall large reduction in scores can be attributed to the combined effect of background noise and the mixed talker design. Of note is that Child 1, 2 and 3 (moderate-severe/severe speech disorders), each showed a similar reduction in scores (between 22% and 27%), suggesting that this is a reliable difference at this severity level.

Figure 16. Comparison of word identification scores for the *TOCS* word intelligibility measure (subset of 25 words) and for Listening Task 1 (first presentation).



Past research has shown that a listener's success in identifying spoken words depends upon familiarity with the talkers' speech. When identifying words spoken by only one talker, listeners can be as much as 10-15% more accurate than when identifying words spoken by a multiple talkers, randomized across talkers (Mullennix, Pisoni, & Martin, 1989; Sommers, Kirk, & Pisoni, 1997; Sommers, Nygaard, & Pisoni, 1994; Verbrugge, Strange, Shankweiler, & Edman, 1976). Verbrugge et al. (1976) recorded vowels spoken by men, women and children in the syllable structure /pVp/. Listeners were asked to identify the syllable spoken. In the mixed talker task, listeners heard two sets of 45 syllables, spoken by a total of 15 talkers. In the segregated talker condition, listeners heard 45 syllables spoken by a single talker. Each listener heard a man, woman and child each say the 45 syllables in three separate blocks. In the mixed talker condition, listeners made an average of 17% errors while in the segregated talker condition, listeners made an average of 9.5% errors (a difference of 7.5%). The increase in errors in the mixed talker condition is attributed to listeners not being able to become familiar with one talker and their speaking characteristics. Accurate identification of ambiguous vowels (e.g., e, æ, 9,7,4) showed the most improvement in the single talker condition. The benefit of listening to a single talker has also been found with single words. Sommers et al. (1997) presented listeners with 100 words of varied syllable structures spoken by 10 different talkers in the multiple talker condition and one talker in the single talker condition. Words were presented in three different levels of background noise: quiet, - 5 dB and + 5 dB. Listeners were asked to type in the word they heard. Regardless of the SNR the words were presented in, listeners were better able to identify words presented in the single talker condition compared with the multiple talker condition.

Actual differences were not reported. However, an inspection of the graphs revealed an average benefit of about 11% for the single talker condition (range from 4% to 17.5%). Furthermore, when Hustad & Cahill (2003) compared speech intelligibility scores for listeners identifying words spoken by one adult with mild to severe dysarthria, listeners were 11% more accurate in the last quarter of the task compared to the first quarter, an improvement associated with becoming more familiar with the talker.

To evaluate the effect of the background noise, the SNR was calculated for the original audio files before the background noise was added. This was done to determine the mean reduction in SNR that resulted in the current study when background noise was added at an intensity of 10 dB less than the signal. SNR was calculated for this ideal listening condition using Adobe Audition 1.5 (Adobe Systems Incorporated, 2004). The intensities of the target word and background noise for the ten words for each child used in Listening Task 2 were measured. The 50 millisecond (ms) segment of the word with the highest intensity was selected and the Maximum RMS Power[22] within this segment was measured. A second 50 ms section of the background (e.g., part of the audio file where the child was not talking) was selected and the Maximum RMS Power within this segment was recorded. For 15 of the words, a 40 ms segment was used as the duration of the file before the signal began or after it finished was exactly 50 ms or less.

SNR values for the ideal listening condition (without background noise added) were calculated by subtracting the background intensity from the signal intensity and averaging the absolute values of these SNRs. Average SNRs ranged from 27 dB (SD = 5.6) for Child 3 to 44 dB (SD = 3.1) for Child 1. The SNRs used in this study, therefore,

[22] This is based on a full-scale sine wave being 0 dB. Intensity values are therefore negative values.

were 17 to 34 dB lower than the original audio files. Cooper and Cutts (1971) found that accuracy scores decreased an average 3.57% per dB HL that the noise was increased for normal hearing subjects when identifying single words spoken by a typical talker. Listeners' word identification scores in Listening Task 1 were not influenced to this degree by the reduced SNR. However, it is noted that, the SNR for the stimulus words was reduced from a maximum of 44 dB before the addition of noise to 10 dB after the addition of noise. SNRs used in Cooper and Cutts (1971) ranged from + 12 dB to 0 dB for the normal hearing listeners and from "quiet"[23] to + 4 dB for the hearing impaired listeners.

No comparable studies were identified for direct comparison of the effect of added background noise on speech intelligibility scores for children with dysarthria. However, for a normal talker and normal hearing listener, an SNR of + 10 dB does not have much of an effect on word identification scores. From this, it might be expected that word identification scores in Listening Task 1 for these children with dysarthria would be similar to the *TOCS* word intelligibility scores for the same 25 words when in fact scores dropped an average of 22%. For these children, with already compromised intelligibility, the combined effect of the background noise and the mixed talker condition had negative effects on word identification scores. However, the individual effects of background noise and the mixed talker condition are unknown as these were not treated as independent variables.

In summary, the findings suggest that the overall reduction in word identification scores on Listening Task 1 compared to the *TOCS* sentence intelligibility scores used to

[23] Words were presented at 40 dB relative to each hearing impaired listener's Speech Reception Threshold. The presence or absence of background noise or recording noise in the audio signals was not reported.

classify the children by severity level can be attributed to the differences between the two tasks (e.g., sentence versus single word stimuli; mixed versus segregated talker design and background noise versus quiet). Findings also suggest that the amount of reduction in scores due to the nature of the utterances used (sentences versus monosyllabic words) varied across the four children, while the amount of reduction in scores due to the combination of the mixed talker design and added noise was similar across children.

Use of Different Measures for Classifying Severity of Speech Disorder

The findings discussed in the previous sections support the observation made in the Introduction that measures of intelligibility are relative, and by extension, that measures for classifying severity of speech disorder are also relative. For a given talker, these measures can vary depending on aspects of the speaking task, the listening environment and the listening task. Although it has been documented that measures of intelligibility for sentences, conversational speech samples and words are related to one another, this study has demonstrated different ordering of children by severity depending upon the measure used. For the four children whose recordings were used in this study, word identification scores based on imitated sentences (*TOCS* sentence intelligibility measure) had a wider range of scores that were more evenly distributed across the range, compared with word identification scores based on the subset of 25 words from the *TOCS* single word intelligibility measure and when compared with measures of speech disorder severity based on PCC, PVC and PPC determined from a spontaneous speech sample.

The talker severity classification was based on the *TOCS* sentence intelligibility scores, with the children ordered 1, 2, 3 and 4 from least to most severe. When using PCC from a conversational speech sample, Child 2 was the least severe and there was no difference in severity classification for Child 1, 3 and 4 using Shriberg's criteria. When using PVC and PPC from a conversational speech sample, the children were ordered similarly to the *TOCS* sentence intelligibility scores with a difference between all children of at least one standard deviation. The ranges of PCC, PVC and PPC scores from conversational speech samples were reduced (PCC – 9.61%, PVC – 39.76%, PPC – 20.51%) compared to the range observed with the *TOCS* sentence intelligibility scores (48.50%). Finally, past results from *TOCS* listeners identifying the 25 words used in this study without background noise and when listening to a single talker were equivalent, for Child 1, 2 and 3. Child 4 was the only one whose scores were ordered consistently.

Finer grained measures of phonetic accuracy (percent initial consonants, percent final consonants, percent vowels and percent syllables shapes correct) derived from the listeners' responses on the word identification task for Listening Task 1 were subsequently obtained and analyzed to determine if these showed an increase in range of scores and differences in scores between children. Results for initial consonants, vowels, final consonants and syllable shapes are discussed with regard to their relationship to the overall word identification scores and *TOCS* sentence intelligibility scores. In general, these results support the clustering observed among Child 1, 2 and 3 and the ordering of the word identification scores across the four children in Listening Task 1.

For initial consonants correct, Child 2 had the highest score, followed by Child 1 and 3 and then Child 4. There was no difference in scores between Child 1 and 3. This supports the overall word identification scores obtained in Listening Task 1 where Child 2 had the highest score, suggesting that if the listener was able to get the initial consonant, they were likely to get the word right. The lack of difference observed between Child 1 and 3 on initial consonants correct scores also supports the word identification score results for Listening Task 1, i.e., no difference observed between these two children.

For vowels correct, Child 1 had the highest score, followed by Child 2 and 3 and then Child 4, which is the same ordering as the *TOCS* sentence intelligibility scores. There was no difference in vowels correct scores between Child 2 and 3. Child 1 had the highest score for vowels correct and it is possible that this is what made his sentences easier to understand than his words. Listeners may have been better able to fill in consonants when given the context of a sentence and accurate vowel production.

For final consonants correct, Child 2 had the highest score, followed by Child 1 and 3 and then Child 4. Child 1's score did not differ from Child 2 and 3's scores. This ordering is parallel to the word identification scores for Listening Task 1. This parallel ordering of severity suggests that when the listener knows the final consonant, he or she is better able to identify the word correctly.

For syllable shapes correct, Child 1 had the highest score, followed by Child 2 and 3 and then Child 4. Child 2's score did not differ from Child 3's score. This is a similar pattern found in the *TOCS* sentence intelligibility scores, with Child 1 having the highest scores for both the *TOCS* sentences intelligibility measure and syllable shapes for single words in Listening Task 1. It is possible that in sentences, Child 1 was most intelligible because when given the context of a sentence, accurate vowels and accurate syllable shapes, the listener is better able to correctly identify the distorted consonants.

For all of the measures of phonetic accuracy for Listening Task 1 and the spontaneous sample, Child 4 had the lowest scores. This supports the *TOCS* sentence intelligibility score classification of most severe and the distribution of word identification scores on Listening Task 1.

As discussed and as shown in Table 1, a reduced range of scores for Child 1, 2 and 3 is evident in the Percent Consonants Correct (PCC), Percent Vowels Correct (PVC) and Percent Phonemes Correct (PPC) scores calculated from the conversational samples of these children. Results of the PCC analysis classify Child 2 as mild-moderate and Child 1, 3 and 4 as moderate-severe. The PVC analysis yielded a broader range of scores (44 – 83%) but scores were very similar for Child 2 and 3 (73% versus 72%). The PPC analysis revealed a maximum difference of only 8.3% in scores across Child 1, 2 and 3 and a range of 20.5% across the four children. Of interest is that when PPC, PVC and PPC scores obtained from conversational speech samples of these children are compared with similar measures derived from listeners' orthographic responses to the monosyllabic words presented in Listening Task 1. As shown in Table 11, the patterns of score distributions are very similar. The order of PCC and PCC-ID scores by child are the same (Child 2 has the highest) and the clustering of scores for Child 2, 1 and 3 is similar (range of 9.6 % for the conversational speech sample and 7.7% for Listening Task 1). For PVC and PVC-ID scores, the ordering of Child 2 and 3's scores change, with Child 2 being lower in Listening Task 1 and higher in spontaneous speech. However, the range across Child 1, 2 and 3 is similar (11.3% in spontaneous sample versus 12.4% in Listening Task 1). The PPC and PPC-ID scores also show the same ordering of children from high to low in the spontaneous sample and Listening Task 1 and a small range for Child 1, 2 and 3 (8.3% in spontaneous speech sample and 4.4% in Listening Task 1.) An unexpected result of this comparison is the observation that all children's PCC scores are slightly lower in the spontaneous speech sample than their PCC-IC scores obtained from the listeners' orthographic transcription of the monosyllabic words presented in Listening Task 1. However, there are significant differences between the methods of data collection for PCC and PCC-IC. Specifically, PCC scores are from a narrow phonetic transcription of a spontaneous sample whereas the PCC-IC scores are from listeners identifying sounds in the mixed talker design with added noise. This may reflect the simple syllable shapes of the monosyllabic words used in Listening Task 1 and possibly a greater clarity of production by the children on a single word imitation task compared with a conversational speech sample. As expected, PVC scores in the spontaneous sample are slightly higher than the PVC-IC scores. Overall, this results in the PPC/PPC-IC scores being very similar for the spontaneous sample and Listening Task1 with the greatest difference observed for Child 4 (51.7% versus 45.0% respectively).

These results indicate that of all the available indices of talker severity for the child talkers in this study (*TOCS* sentence intelligibility scores, *TOCS* word intelligibility scores for the sets of 25 words, PCC, PVC and PPC on spontaneous speech samples, word identification scores for the sets of 25 word stimuli presented in Listening Task 1 and PCC-IC, PVC-IC and PPC-IC scores based on listeners' orthographic responses to the words presented in Listening Task 1), sentence intelligibility scores provided the widest range in scores (38.0 to 86.5%) and the most even distribution of scores across the range (at least a 10% difference among all scores for the four children). PVC scores showed the next widest range (43.7 to 83.4) in the spontaneous sample condition but the difference between Child 2 and 3 was 1% and the range for Child 1, 2 and 3 was 11.3%). The smallest range was evident in the PCC scores with only a 9.6% difference between the highest and lowest scores. The sentence intelligibility scores identified four levels of talker severity ranging from mild to severe. The PCC scores from the spontaneous sample identified only two levels of talker severity: mild-moderate and moderate-severe. No published information was found that assigned severity ratings to PVC and PPC scores. However the range of scores on the spontaneous sample for PVC and PPC was small. As noted, using the PCC-ID scores from listeners' orthographic transcription in Listening Task 1, two levels of severity are apparent using Shriberg's guidelines (mild-moderate and severe). It would be desirable to have speech disorder severity classification criteria available for PVC and PPC. In the case of children with dysarthria, at least in the current study, PCC based on a spontaneous speech sample did not appear sensitive to the range of their speech disorder severity. Based on the spontaneous sample, PPC scores showed the same ordering of children's scores as the *TOCS* sentence intelligibility measure and PVC scores showed the widest range of scores across the children.

For the *TOCS* word intelligibility scores, where words were presented in an ideal listening condition and listeners heard only one talker, two levels of severity can be identified: mild-moderate and severe. Word identification scores in Listening Task 1, where listeners heard the words presented in background noise and in a mixed talker design, identified three levels of talker severity: moderate–severe, severe and profound.

Implications of the Comparisons of the Severity Measures Available for the Four Children with Spastic Dysarthria.

As described previously, not only do the number of levels of talker severity and range of scores change depending upon the measure used, but the children's ordering of talker severity also changes. From high to low word identification scores, children were ordered 1, 2, 3 and 4 when using *TOCS* sentence intelligibility scores, PVC and PPC scores. For the two levels identified with PCC scores, Child 2 had the highest scores, followed by similar scores for Child 1, 3 and 4. For the *TOCS* word intelligibility scores for the 25 words, the two levels identified Child 1, 2 and 3 as the most accurate and Child 4 as less accurate. For Listening Task 1, children were ordered in three different levels: Child 2, 3 and 4[24]. Findings indicate that for the specific 25 words used in this study, balanced for phonetic content, neighbourhood density and word frequency, the range in levels of talker severity is reduced and shifted to the more severe region of the severity continuum.

In summary, it is evident that for these four children, the relative performance and spread on word identification scores and finer grained phonetic analyses varied depending on the task. In this experiment, a range of speech disorder severity from mild to severe was desired to test hypotheses about the interaction between talker severity and presentation number. The 25 single word stimuli for each talker did not provide this same range of talker severity and therefore prevented a full testing of the hypothesis of the interaction because the mild and moderate levels of severity were not represented. In view of this, the lack of a significant severity-presentation number interaction must be interpreted with caution as this finding is based on a listening task where listeners' responses indicated that the children's speech severity levels were in the moderate severe to profound levels. If the range of severity observed in response to Listening Task 1 had included the mild region, a significant interaction may have been observed for word identification scores and response times.

---

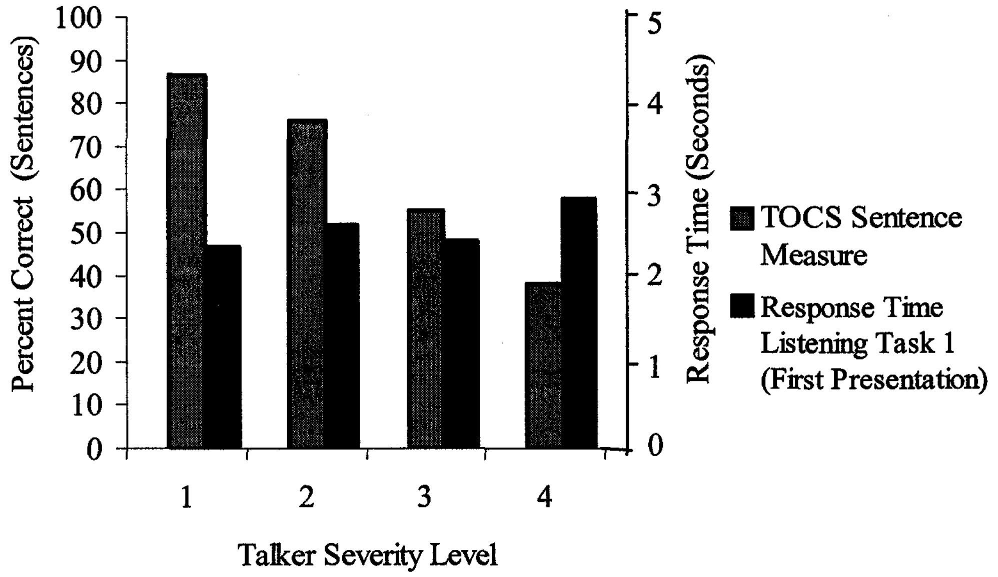
[24] Child 1 was ordered between child 2 and 3.

An implication of these findings for future research is that caution must be taken when classifying talker severity if a full range and even distribution among levels is desired. In hindsight, selecting child talkers by severity level using the *TOCS* word intelligibility scores for the subset of 25 words obtained in the ideal conditions would have likely resulted in a wider spread in the results obtained for the word identification scores in Listening Task 1. Based on the observed decrease in scores in the mixed talker and + 10 dB SNR condition, to maintain a mild-moderate score, a child would need to have a score close to 100% on this word intelligibility measure. It is unlikely that a young child with dysarthria could be identified with this high a score. Mean word identification scores on the *TOCS* word intelligibility measure for six to seven year-old children with no history of speech delay/disorder are in the range of 80 - 97% (95% confidence interval) (M. Hodge, personal communication, January, 2007). Therefore, to test hypotheses about the interaction of speech disorder severity and presentation number in a mixed talker design in + 10 dB SNR, across the full range of severity, using children seven years and younger without speech disorders appears feasible. The results obtained would be applicable only to severity levels based on the measure(s) used to classify severity. Ideally, it would be desirable to find talkers who represented the same range and distribution of speech disorder severity on a variety of measures based on connected speech and single words. In addition, it is recommended that the effect of the background noise be controlled, i.e., add noise as an independent variable and have a quiet versus noise condition so that its influence can be dissociated from that of the mixed talker design. This means that the researchers would have a very good idea in advance of what the word identification scores would be in the quiet condition for the first presentation of the stimuli, but if the most important question is whether a severity by presentation number or severity by presentation number by noise interaction is present for word identification scores and response times, this control on levels of speech disorder severity would allow testing of these hypotheses across the full range of severity.

Effect of Talker Severity on Response Time

As predicted, the main effect of talker severity on response time was significant. It was expected that Child 1 would have the fastest response times, followed by Child 2, 3 and 4. The ordering of talker severity for three of the four children (1, 2 and 4) was as expected with words spoken by Child 1 being responded to the fastest and words spoken by Child 4 being responded to the slowest. Response times to words spoken by Child 3 were faster than expected given the talker severity levels identified by the *TOCS* sentence intelligibility scores. Child 3's response times were no different than Child 1's. Listeners may have felt more confident in knowing the words spoken by Child 3, when in fact they were unable to identify many of the words. Figure 17 shows the *TOCS* sentence intelligibility scores and response times to the 25 words in Listening Task 1.

Figure 17. TOCS sentence intelligibility scores and response times for Listening Task 1.



If word identification scores and response times are thought to be related measures of listeners' behaviour when trying to understand the stimulus words presented, in view of the results from the word identification scores and derived PCC-ID scores from Listening Task 1, it might be expected that Child 2 would have the shortest response times (not the case) and that Child 1 and 3 would not differ (the case). Instead, Child 1 and 3 had response times similar to each other and shorter than Child 2. In examining all the severity measures reported in Tables 1, 4 and 5, there is no case where Child 1 and 3 are very similar and significantly better (less severe) than Child 2. Like all the other severity measures discussed, with the exception of the *TOCS* sentence intelligibility measure, response times for Listening Task 1 did not distribute evenly across the children (range of .23 seconds across Child 1, 3 and 2; and .48 seconds between Child 1 and 4). For the effect of talker severity, the smallest difference in response time found to be significant was between Child 2 and 3. Significant differences were found for five comparisons (1-2; 3-2, 1-4, 3-4, 2-4), versus four comparisons for word identification scores for Listening Task 1 (1-4; 2-3; 2-4; 3-4). The effect size for the smallest significant difference for response time was .23 and the effect size for largest difference in response time found between children was .59 (Child 1 versus 4) compared to word identification scores where the effect size for the smallest significant difference was .48 and the largest difference found between children was 4.13 (Child 2 versus 4). Effect size calculations for these largest differences are shown in Appendix O. Response times appeared sensitive to smaller differences between children, i.e., significant difference found between Child 1 and 2 for response time but not for word identification scores. Across the range of severity of children in the study, larger effect sizes were found for word identification scores when the children with the largest and smallest scores were compared on Listening Task 1.

Mean response times across children in Listening Task 1 were between 2 and 3 seconds. Larsby et al. (2005) identified mean response times between 1.265 and 1.400 seconds for young, normal hearing listeners identifying whether a spoken stimulus was a word or not by pressing a button. In a study by Downs (1982), listeners with hearing impairment were asked to repeat words presented auditorily. Mean response times between .104 and .130 seconds were calculated for listeners with and without hearing aids, respectively. These scores were the result of subtracting true mean response times from each participant's mean baseline response time. Mean baseline response times were calculated by having listeners press a button when they saw a light. These baseline response times ranged from .260 to .467 seconds. In the study by Downs (1982), listeners spoke, rather than typed, the word they thought they heard and word identification scores were higher than those in the present study (means of 70 to 50%, with and without hearing aids, respectively). The response times in Listening Task 1 are longer than those reported in previous studies, however, the nature of the tasks are different. For example, studies measuring listeners' response times to word stimuli often involve asking listeners to identify whether the spoken stimulus was a real word or not. This decision is easier and would be expected to take less time than deciding what word the stimulus was, especially when the signal is distorted or difficult to understand.

In summary, the pattern of results obtained for the effect of talker severity on response times differs from that obtained from the word identification scores for Child 1, 2 and 3 and for the ordering of talker severity using the *TOCS* sentence intelligibility measure. It is not evident why Child 3's response time should be similar to that of Child 1 and significantly higher than that of Child 2. As observed previously, additional analysis of Child 3's speech patterns appears warranted given several unexpected findings concerning her pattern of results on the measures of severity reported. Additional investigation of the stimuli used for Child 1 and 2 in Listening Task 1 is also recommended to attempt to determine why their word identification scores did not differ significantly but their response times did. As expected, response times for Child 4 were longest, consistent with the most severe scores observed for this child in all the measures of accuracy available (word identification scores and phonetic accuracy). No studies were found that investigated response times for identifying words spoken by children with dysarthria. Studies found had listeners make a decision about whether a presented stimulus was a word or not, or listeners repeated the words instead of typing them. The relatively long response times observed in this study (2 – 3 seconds) compared to those in the literature cited (less than 1.5 seconds) would be expected given the task, i.e., actual identification of words spoken by children with a speech disorder and a typed instead of spoken response. This finding supports the observations made in the Introduction that when conversing with children with dysarthria, listeners need to be prepared to have longer response times (i.e., to take more time listening) if they want to try and understand these children's speech.

Effect of Talker Severity on Effort Ratings

With regard to the dependent variable of listeners' direct magnitude estimate (DME) ratings of perceived effort, the overall results were as expected. The child classified as least severe based on the *TOCS* sentence intelligibility measure had the lowest ratings of effort and the child classified as the most severe had the highest ratings. The two children in the middle did not show significantly different effort ratings. The ordering of severity based on these DME ratings of effort is similar to the *TOCS* sentence intelligibility scores. Figure 18 shows the children's *TOCS* sentence intelligibility scores and their mean DME effort ratings. However, while listeners rated Child 1 as requiring significantly less effort than Child 2, this was not reflected in their word identification scores in Listening Task 1. The word identification scores for Listening Task 1 and the effort ratings for Listening Task 2 are presented in Figure 19.

Figure 18. TOCS sentence intelligibility scores and DME effort ratings for the 10 words in Listening Task 2.

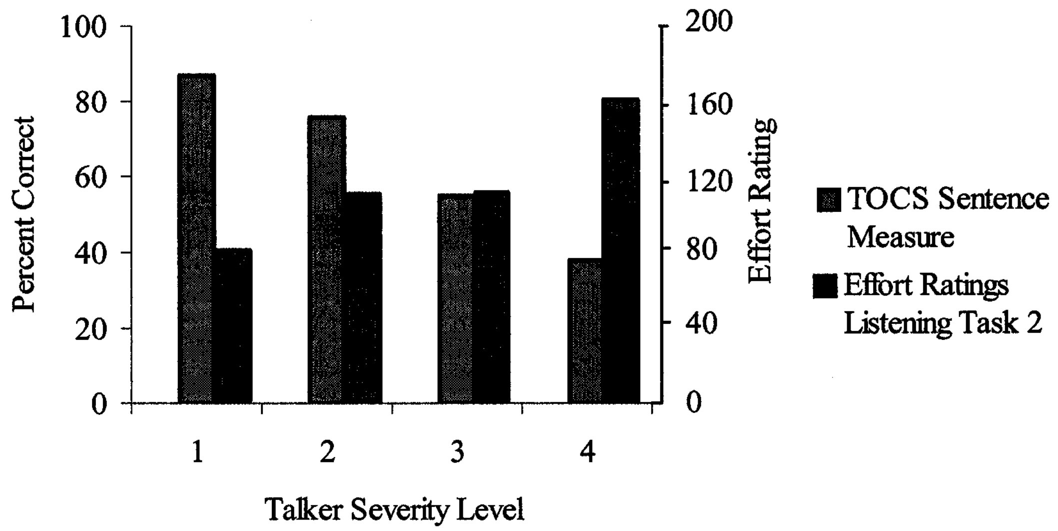
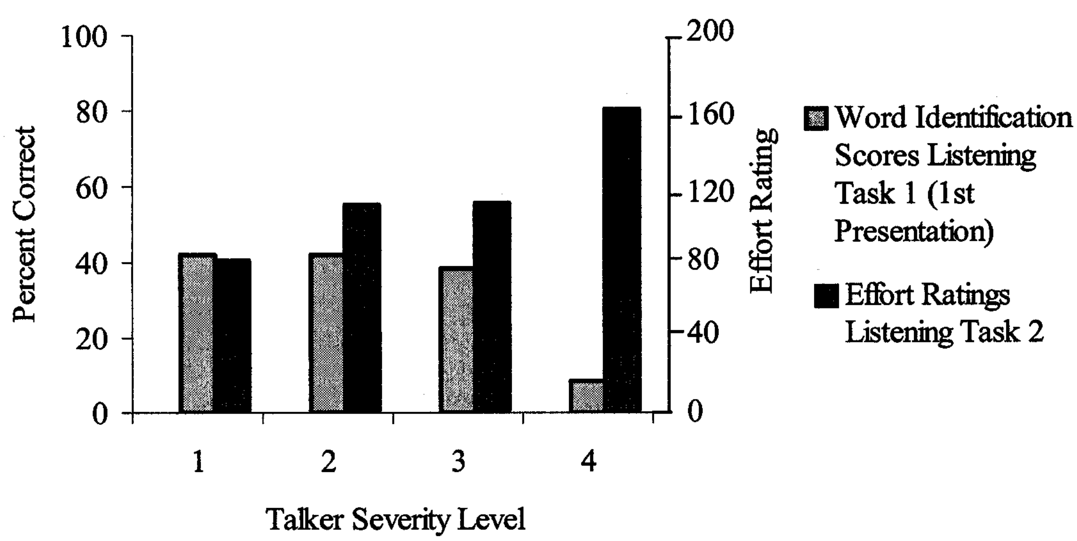


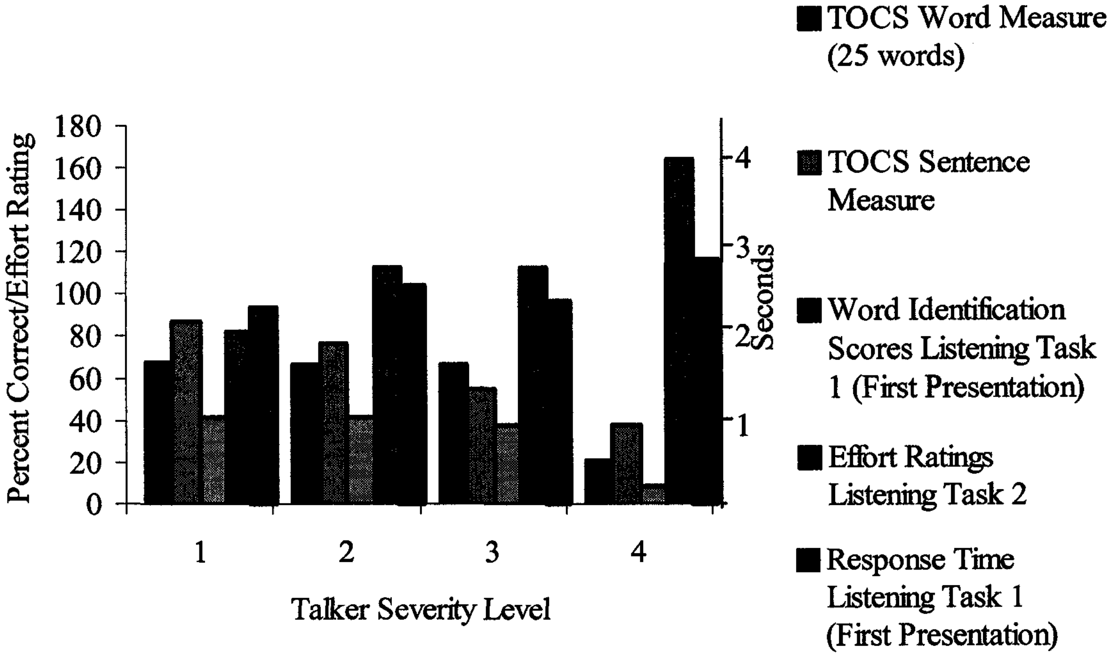
Figure 19. Word identification scores from Listening Task 1 (25 words) and Effort Ratings for the 10 words used in Listening Task 2.



Something about the words spoken by Child 1 made listeners feel as though little effort was required when in fact they were frequently incorrect in their actual identification of the words. The other measures where Child 1 had the best scores were the number of vowels correct and the number of syllable shapes correct for listeners' orthographic responses on Listening Task 1. When listeners perceive the vowel and syllable shape as being clear, they may feel as though the word is easy to understand. In contrast, when the vowel and syllable shape are not clear, listeners may feel as though they need to work harder to understand the word. The similarity in effort ratings for Child 2 and 3 supports the findings from the phonetic analyses of vowels and syllable shapes correct based on listeners' responses in Listening Task 1, where scores for Child 2 and 3 did not differ and were ordered between Child 1 and 4. For Child 4 the results from listeners' ratings of effort on Listening Task 2 reflect their word identification scores and response times on Listening Task 1. Child 4 had the lowest word identification scores, slowest response times and highest ratings of effort, all as expected given her severe classification on the *TOCS* sentence intelligibility measure. That is, the classification of Child 4 as most severe held constant regardless of the measure used.

For the readers benefit, Figure 20 shows the patterns of scores for the *TOCS* word intelligibility measure for the subset of 25 words, *TOCS* sentence intelligibility measure (both obtained in quite using a segregated speaker design) and the dependent measures obtained in Listening Tasks 1 and 2 (word identification scores, DME effort ratings and response times).

Figure 20. TOCS intelligibility scores and measures of listener behaviour from Listening Tasks 1 and 2.



In summary, the ordering of DME effort ratings reflected that of the *TOCS* sentence intelligibility measure and the phonetic accuracy of vowels and syllable shapes obtained from responses to Listening Task 1. Child 1 had the lowest ratings of effort and the highest vowel and syllable shape correct scores, Child 4 had the highest ratings of effort and the lowest vowel and syllable shape correct and Child 2 and 3 had scores in the middle. However, unlike the *TOCS* sentence intelligibility measure, but similar to the vowels and syllable shapes correct scores in Listening Task 1, the effort ratings for Child 2 and 3 did not differ from each other. Of interest is that Child 1 had significantly lower effort ratings than Child 2 despite Child 2 having slightly higher word identification scores in Listening Task 1. In this study, measures of vowels and syllable shapes correct and effort ratings were ranked similarly to each other, but differently from the ranking of their actual word identification scores or response times for these same single syllable word productions. This suggests that listeners are using something about the single words other than, or in addition to, the effort involved in the task of identifying a word. Based on the listeners' responses to questions 1 and 2 about what kinds of things made it easier or more difficult to understand these children's words, possibilities include vowel accuracy, syllable shape information, voice quality, coarticulation cues, word duration, word familiarity, phonological neighbourhood density and the relative impact of background noise on sound segment identity. It also suggests that listeners use more than response time alone in generating their effort ratings.

Reliability of Effort Ratings

Fifty-eight percent of listeners were able to reliably rate their effort with "good" reliability (i.e., attain ICCs of .7 or higher). An additional 27% achieved slightly lower than "good" reliability (.60 - .70). Past researchers have reported ICCs for intra-rater reliability on direct magnitude estimations of bizzareness, naturalness, pleasantness, acceptability and hypernasality of speech samples ranging between .48 and .95 (Eadie & Doyle, 2004; Southwood, 1996; Southwood & Weismer, 1993; Whitehill et al., 2002). Mean reliability measures using ICCs from these studies ranged from .67 to .84. The mean ICC value from the present study was .72, within the range for past research on DME. The *TOCS* sentence intelligibility scores for the child talker used as the modulus for the DME task and assigned a rating of 100 was 58%. The mean rating for Child 1 who had a *TOCS* sentence intelligibility score of 86.5 % was 81.6 and the mean rating for Child 4 who had a *TOCS* sentence intelligibility score of 38% was 163.5 indicating that listeners were using the modulus to make their ratings in a valid manner. Child 2 (*TOCS* sentence intelligibility score of 76%) and Child 3 (*TOCS* sentence intelligibility score of 55%) had almost identical mean effort ratings (112.2 and 112.5, respectively) suggesting that in this severity range listeners judged these child talkers similarly and as requiring slightly more effort than the modulus.

Compared to Listening Task 2, previous studies also used more stimuli for intra-rater reliability and listeners rated either spoken sentences or connected speech samples. For example, in Southwood (1996), listeners heard 60 sentences in total and 15 of these were rated twice. In the present study, listeners heard 40 words and eight of these were rated twice. The talkers used in this study were also very difficult to understand. All talkers had word identification scores at or below 45% when the words were presented in background noise and in a mixed talker condition. They would be classified as moderate-severe to severe based on the original classification (e.g., 45-55% moderate-severe, < 35% severe). Based on mean ICCs for the 33 listeners, it appears that DME is a reliable measure of listener effort when rating words spoken by children with severe dysarthria and with as few as eight repeated stimuli. When the one-way ANOVA was run again excluding the five listeners with poor reliability (i.e., less than .6) the pattern of results did not change from those obtained with all 33 listeners.

In summary, the mean ICC of .72 for intra-rater reliability for DME ratings of listener effort for these 33 listeners suggests that young professionals in training who are inexperienced in listening to the speech of children with dysarthria can perform reliably on this kind of task.

Effect of Presentation Number on Measures of Listener Behaviour

Effect of Presentation Number on Accuracy Scores

The expected result that when listeners heard the same stimulus repeated multiple times, word identification scores would increase, was found. Scores were lowest for the first presentation and increased with the second and third presentations. Mean scores increased by almost 3% from the first to the third presentations. There was a significant increase in word identification scores between each of the presentation numbers. Although this increase in word identification scores was statistically significant, the effect size for the difference between presentations 1 and 3 was .13, trivial based on the classification of effect sizes used (Cohen, 1988). An increase of 3% would not be expected to have a significant effect in the life of a child with a severe intelligibility deficit.

In examining the effect of repeated presentation of a word on finer-grained measures of phonetic accuracy, small differences were also evident. For the initial consonants, scores increased only 2% from the first to the second and third presentations. There was no significant difference in scores from the second to the third presentation. For vowels, scores increased less than 2% from the first to third presentations. There was no difference in scores between the first and second presentations. For the final consonant, scores increased by 3.5% from the first to third presentations. There was no difference in scores between the second and third presentations. Listener's identification of syllable shapes did not change with repeated presentations.

Repeated presentations benefited listeners most in identifying the final consonants and the entire word. Past research reported improvements with repeated presentations between 4 and 7% (more than the 3% found in Listening Task 1). However, the past studies used spoken sentences (Hodge et al., 1999; Monsen, 1983; TyeMurray et al., 1990) or single words spoken from speech synthesizers (McNaughton et al., 1994). Additionally, one study found no effect of repetition with single words in the visual modality only (Gagne & Wyllie, 1989).

In summary, results from Listening Task 1 indicate that repeated presentations of the exact same stimulus word spoken by children with dysarthria can increase accuracy scores. This increase however is on the order of 3%, lower than what has been found for sentences presented auditorily. This increase is also lower than would be considered clinically significant (trivial effect size) and therefore not enough to support advising children with dysarthria and their communication partners to use this strategy solely in communication breakdowns.

Effect Sizes

Differences between the first and third presentations were trivial, an average of .13 for word identification scores, initial consonants, vowels and final consonants (ranged from .10 to .17). The increase was less than 3% from the first to the third presentations. Two past studies were found that provided sufficient information to calculate effect sizes for the effect of presentation number on accuracy scores. In the study by Peng et al. (2004), listeners heard sentences spoken by a child with a cochlear implant twice in a row. Listeners were asked to identify the words spoken. Mean scores for the first presentation were 67.86%, while for the second presentation of the sentence, mean scores were 71.54%. Effect size was calculated to be .09 using the standard deviations given and the same formula used in the present study[25] (Cohen et al., 1991).

McNaughton et al. (1994) had children and adults identify single words produced by two speech synthesizers. Listeners attended five sessions. Stimuli were either words the listeners had heard in previous sessions or novel words. Repeated[26] words were identified more accurately than novel[27] words, regardless of age of listener and device used. Effect sizes were calculated based on given means for repeated versus novel words for adults or children and device 1 or 2 (4 separate conditions), and standard deviations for each of the five sessions[28]. Effect sizes ranged from .07 for children listening to the DECtalk device to .31 for adults listening to the DECtalk device.

As noted previously, severity levels in the mild to moderate range were not represented in listeners' word identification responses on Listening Task 1 so conclusions cannot be drawn about the effect of repeated presentations on accuracy scores for these severity levels. The effect size of presentation number on accuracy scores in Listening Task 1 is within the range expected based on past research. The effectiveness of repeating the identical stimulus word(s) on increasing accuracy scores appears to be small, regardless of the talker population. Listeners may improve their accuracy scores more if the talker provides additional non-auditory information (e.g., eye contact, facial expression, gestures) or changed their speech signal in some way (e.g., use of clear speech strategies) on repeated presentations. Given the commonly reported use of repetition as a repair strategy, as described in the Introduction, it is hypothesized that

---

[25] The difference between the means divided by the pooled standard deviation.
[26] The identical stimulus played again.
[27] A word not yet produced by the device for that listener.
[28] Standard deviations from the five sessions were pooled. These pooled standard deviations were then pooled again for each of the four conditions and then divided by the difference between the means.

talkers and listeners do make use of non-auditory information and clear speech strategies to increase the success of word repetition as a repair strategy.

Effect of Presentation Number on Response Time

It was expected that with more presentations of the same stimulus, listeners would become faster at responding. Overall, response times did decrease by an average of .60 seconds from the first to third presentation (about 28% faster). Response times decreased significantly with each presentation. Listeners had a better idea of what their response was going to be, whether correct or incorrect, the more they heard the word. Mimura et al. (1997) found that when listeners are asked to identify whether a spoken stimulus is a real word or not, responses were .058 to .179 seconds[29] faster when the stimulus was repeated. Rugg (1985) found that listeners were .045 seconds faster at deciding whether or not a stimulus was a word or not. Average response times in the study by Rugg (1985) were .634 seconds for the first presentation and .589 seconds for the second presentation, about 7% faster. Listeners in the present study reduced their response times by much more than the studies discussed above, however, listeners were required to identify difficult to understand words as opposed to simply deciding whether or not a stimulus was a real word or not. Their overall response times were also much longer (2.160 seconds) than studies reported previously. However, listeners' responses were fast enough so that repeated presentations of the words were within the span of echoic memory (10 seconds), part of the memory system that is effected by repetition priming.

In summary, results indicate that listeners were faster at identifying the stimuli with repeated presentations. Listeners continued to be faster at identifying the stimulus even between the second and third presentations. These faster response times may

[29] Depending upon how many intervening stimuli were between the repeated items.

indicate that less effort was required of listeners with repeated presentations. Furthermore, response times were within the span of echoic memory, indicating that the increased accuracy scores and decreased response times with repeated presentations may have been due to the effect of repetition priming.

Effect Sizes

As noted previously, severity levels in the mild to moderate range were not represented in listeners' word identification responses on Listening Task 1 so conclusions cannot be drawn about the effect of repeated presentations on response times for these severity levels. The smallest effect size for response time and presentation number was .33, between presentations 2 and 3. Only one previous study that looked at the effect of repeated presentations on response time was identified that contained enough information to determine effect size. Holcomb et al. (2005) had listeners identify whether or not an auditory target was a real word or not. Target words were either primed with an unrelated word or the same word. Across conditions[30], the effect size for response times to words with repeated versus unrelated primes was .96. Effect size was calculated by converting the given standard errors, into standard deviations, pooling the standard deviations across the different conditions for words with related and unrelated primes separately and pooling the standard deviations of related and unrelated primes together. The mean response time for words with related primes was subtracted from the mean response time for words with unrelated primes. This difference was then divided by the final pooled standard deviation, as done in all other effect size calculations in this study (Cohen et al., 1991). The effect size for response time on presentation number in Listening Task 1 is lower than the study by Holcomb et al. (2005); however, there are many differences

[30] Conditions being different stimulus onset asynchronies.

between the two studies including the response required. In Listening Task 1 the effect size of repeated presentations for response time (.33), is higher than the effect size of repeated presentations for word identification scores (.09). This might be expected given past studies that suggest response time is a more sensitive measure of listeners' behaviour than is word identification scores. Furthermore, the effect size for the difference in response times between presentations 1 and 3 was .96, a very large effect.

These results suggest that increased presentations of the same stimuli result in much faster response times. The difference is considered to be large between the first and third presentations and small-medium between the second and third presentations, suggesting that significant reductions in response times can occur with at least up to three presentations of the same stimulus.

Relationships between Measures of Listener Behavior on Listening Tasks 1 and 2

The design of this study precluded investigating relationships between pairings of listeners' word identification scores, response times and effort ratings assigned to a given talker. Rather, relationships between pairings of global measures of each listener's word identification scores, response times and effort ratings were investigated. Of the three regression analyses conducted to determine the predictive relationships between listeners' word identification scores and their response times on the first presentation in Listening Task 1 and between each of these measures and the DME ratings of listener effort on Listening task 2, only the regression of word identification scores on response time yielded a significant result. This relationship was in the predicted direction (- . 42), i.e., listeners with higher word identification scores had faster response times. No published studies were located that examined listener behaviour in the same way. Two studies were found that reported both response times (as a measure of listener effort) and accuracy scores (Holcomb et al., 2005; Rugg, 1985). These studies used a lexical decision task but did not report the relationship between response times and accuracy scores. However Holcomb et al., (2005) stated in their results that as response time decreased, accuracy increased. Deciding whether a presented stimulus is a word or not generally leads to few errors; therefore, response time results are often reported only for words where the participant was accurate in making the word/nonword decision (Mimura et al., 1997; Stone & Van Orden, 1993). Findings from the present study suggest that listeners who have lower word identification scores overall also have relatively slower response times.

In the present study listeners' global word identification scores did not significantly predict their effort ratings. Past research that has used talkers as subjects as opposed to listeners, have reported strong negative relationships between ratings of the effort required to understand the speech of a talker and the talker's respective word identification scores (Munro & Derwing, 1995a; Preminger & Van Tasell, 1995; Whitehill & Wong, 2006). Although not significant, the relationship found between accuracy scores and response times in the current study was in the predicted direction ($r = -.21$). In the present study, the weakest relationship was found between response times and effort ratings ($r = -.06$) and the relationship found was opposite to that predicted (negative, rather than positive). This was surprising since in previous studies these have both been considered measures of the work that listeners do in understanding a speech signal. These results suggest that the effort ratings assigned by listeners can't be predicted by their overall response speed or ability to identify words.

Due to the within-subjects nature of the present study, global accuracy scores, response times and effort ratings were used in these regression analyses. These global scores were calculated by averaging scores across children for each listener. This resulted is highly compressed scores that no longer represented a range of talker severities, reducing the likelihood of finding significant relationships between the pairs of the three variables. Therefore, the significant result of the regression of accuracy scores on response times, given the restricted range of scores suggests that these are robust objective correlates of listener behaviour in word identification tasks. Furthermore, only the accuracy and response time scores for the 10 words used for the effort ratings, from each of the 33 listeners, were used in the regression analyses involving effort ratings. This small number of scores may be insufficient to lead to conclusive results about the relationship among these variables. Given the results from previous studies, it is predicted that a study design similar to that of Whitehill and Wong (2006) where listeners provide accuracy scores, response times and effort ratings for a larger number of children (e.g., 15 or more) with a wide range of severity of dysarthria and children are treated as the subjects, would yield a significant relationship between accuracy scores and listener effort ratings. It would be of interest to determine if this design also revealed the predicted relationship between response times and listener effort ratings as no previous studies were located that quantified the relationship of these two variables.

In summary, the predicted negative relationship between word identification scores and response times was found, despite the use of global measures of these two variables, averaged across children, for each of the 33 listeners. There appears to be a robust relationship between these two variables that supports previous findings, i.e., listeners with higher word identification scores respond faster. This design used the listeners as the subjects and provides information about listener's processing characteristics. As noted, a correlational study that uses children with a wide range of severity of speech disorder, rather than listeners, as subjects is recommended to provide a better test of hypotheses about relationships among accuracy scores, response times and effort ratings.

Listeners' Beliefs

The theme analysis of listeners' beliefs related to what made the words easy or difficult to understand and what listeners did to try and understand the words showed many parallels with factors that have been identified in the literature that influence speech intelligibility. Words that were easy to understand were described as being clear, with special effort put on the last sound. Listeners also identified words that were spoken slower and louder as easier to understand. Increasing the precision of consonant (e.g., increasing noise intensity of initial obstruents, releasing final stops) and vowel sounds and overall loudness and duration of the signal are strategies that are characteristic of "clear speech", as described by Picheny et al. (1985). Words that were difficult to understand were described as being mumbled or quiet, more affected by background noise, said quickly and without a precise final sound. These characteristics are also consistent with the clear speech literature, i.e., these characteristics represent "less clear" or "unclear" speech, and with findings from Larsby et al. (2005) that words with more background noise were identified as harder to understand. All these characteristics do influence the clarity of words and overall intelligibility. Of note is that listeners also listed several strategies that have also been identified in the literature as increasing the accuracy and/or decreasing response time of word identification, i.e., word frequency (Balota & Chumbley, 1985; Goldinger et al. (1989); Luce & Pisoni, 1998), word distinctness (phonological neighbourhood) (Goldinger et al., 1989) and increasing familiarity with the talkers (easier to understand the words at the end than the beginning of the listener task) (Hustad & Cahill, 2003). The word identification data collected in Listening Task 1 could be used to determine if they actually support the listeners' beliefs. The percent correct scores for each word, averaged across listeners, could be determined and the ranking of these could be compared with the ranking of word frequency and phonological neighbourhood density scores (the mean of these scores are listed in Appendices C and E) for each child to see if the relationships were significant. As well, similar to the study of Hustad and Cahill (2005), listener's scores for the first 25 words identified could be compared with the last 25 identified to determine if the last 25 had significantly higher scores.

Listeners' reported using a number of strategies to understand the words. The most frequently reported were those where they used phonetic information that they could get from the signal to try and figure out the rest of the sounds in the word (i.e., find unknown from known). Related to this, but identified as a separate theme, were strategies for listening to particular parts of the word (e.g., first and last letter [sound]). These behaviours fit with the fuzzy logic model of speech perception, i.e., listeners reported using all information available to make decisions about the meaning of the signal; used additional repetitions to focus on the parts of the signal that were ambiguous; and used strategies to attempt to resolve this ambiguity. Listeners also reported that they frequently repeated the word (internally or aloud) or used repetition in an alternative way (i.e., compared previous guesses with what was heard the next time, entered something just to be able to hear the word again), suggesting that listeners invoke repetition as a strategy when words are difficult to understand. Listeners' reports that they repeated the words suggest that they may have extended and enhanced their echoic memory for the stimulus even more than the repeated presentations did. It appears that talkers, based on previous studies of communication repairs reviewed in the Introduction, and listeners, as revealed by their behaviors reported in the current study, both use repetition as a common repair strategy, providing converging support for its use, if not effectiveness, as an intuitive communication repair strategy. Listeners also reported that they thought about semantic context, increased their own listening effort and focus for future presentations and made postural adjustments. Although not effective due to the signal and noise coming from the same speakers, the postural adjustments were consistent with trying to increase the signal and reduce the background noise (e.g., leaning closer). It is unknown to what degree each of these strategies helped listeners identify individual words, but it is clear that listeners used the additional presentations of the words to try and increase the accuracy of their responses.

Limitations

The purpose of this study was to investigate the effect of repetition on listeners' identification of words spoken by children with dysarthria and associated response times and effort ratings for four children who represented a wide range of severity as defined by a sentence intelligibility measure (38.0% to 86.5%). However, results for the dependent variable of words identified correctly in Listening Task 1 revealed that listeners obtained very similar scores for three of the children, ranging from 39.6% to 45.1%. The 25 single word stimuli (balanced for phonetic content, phonological neighbourhood and word frequency) for each talker did not provide this same range of talker severity and therefore prevented a full testing of the hypothesis of the interaction because the mild and moderate levels of severity were not represented. In view of this, the lack of a significant severity-presentation number interaction must be interpreted with caution as this finding is based on a listening task where listeners' responses indicated that the children's speech severity levels were in the moderate-severe to profound levels. If the range of severity observed in response to Listening Task 1 had included the mild region, a significant interaction may have been observed for word identification scores and response times.

A secondary purpose of this study was to investigate the relationships among three measures of listener behaviour when identifying words spoken by young children with dysarthria in an SNR within the range found to be typical of classrooms: word identification scores, response times and DME effort ratings. Due to the within-subjects nature of the present study, global word identification scores, response times and effort ratings were used in these regression analyses (Lorch and Myers, 1990). These global scores were calculated by averaging scores across children for each listener. The result is highly compressed scores that no longer represented a range of talker severities, reducing the likelihood of finding significant relationships between all possible pairs of the three variables. A correlational design that uses children with a wide range of severity of speech disorder, rather than listeners, as subjects is recommended to provide more suitable tests of hypotheses about relationships among word identification scores, response times and effort ratings.

CONCLUSIONS

1) As predicted, the main effect of severity of speech disorder for children who differed in the severity of their dysarthria from mild to severe, as classified by their scores on a sentence intelligibility measure, was significant for the dependent variables of word identification scores and response times for the 25 monosyllabic words presented in Listening Task 1 and for DME ratings of listener effort for a subset of these same monosyllabic words presented in Listening Task 2. Listeners had significantly lower word identification scores, longer response times and higher effort ratings for the child with the most severe speech disorder.
2) Unexpectedly, the difference of at least 10% observed between the three children with adjacent levels on the sentence measure used to represent the mild, mild-moderate and moderate levels of severity was not maintained in their word identification scores on Listening Task 1. Their word identification scores ranged from 39.6 to 45.1%, representing the moderate severe-severe region of the severity continuum and their scores did not fall in the same rank order as their *TOCS* sentence scores. Therefore the mild and mild-moderate regions of the severity continuum were not represented in the specific word stimuli presented under the noise conditions in the Listening Task 1 experiment. The smallest difference found to be significant between pairings of these three children was 5.5% words identified correctly. This had a medium effect size, suggesting that this is a meaningful difference in scores on this task.
3) Inspection of pre-existing information about the four children's scores for the subsets of 25 word stimuli used in Listening Task 1, obtained from the *TOCS*

word intelligibility measure in a quiet listening environment and segregated talker condition, revealed the same compression of scores for Child 1, 2 and 3 (all three children had scores between 67 and 68%) observed in Listening Task 1 and an unpredictable relationship with their *TOCS* sentence intelligibility scores. However, when the children's word identification scores obtained in Listening Task 1 are compared with their *TOCS* word intelligibility measures of the same sets of 25 words obtained in quiet and segregated talker design, it appears that the + 10 dB SNR and mixed talker condition of Listening Task 1 reduced scores to a similar degree across these three children (range of 22 – 27%). It is recommended that a follow-up study be undertaken that treats SNR as an independent variable, using stimuli that represent the full range of the speech disorder severity, to systematically determine the effect of more and less favourable listening conditions on word identification measures, response times and effort ratings, and how these may interact with severity and/or the effect of repeated stimulus presentations.

4) Mean listener response times for the four children's word stimuli in Listening Task 1 ranged from 1.65 – 2.90 seconds, which are longer than those found in previous studies for lexical decision-making tasks. This suggests that when conversing with children with moderate-severe to profound dysarthria (range effectively represented by conditions of Listening Task 1), listeners need to be prepared to take adequate time to listen if they want to try and understand these children's speech and children need to give their listeners adequate time to respond. A question for future research is to determine how much more time

listeners need to identify words spoken by children with various severities of dysarthria, compared to children of comparable age without speech disorders.

5) The pattern of results obtained for the effect of talker severity on response times differed from that obtained from the word identification scores for Child 1, 2 and 3 and for the ordering of talker severity using the *TOCS* sentence intelligibility measure. It is not evident why Child 3's mean response time was similar to that of Child 1 and significantly higher than that of Child 2. Results indicate that there is not a one-to-one relationship between these two measures of listener behaviour: word identification scores and response times. Additional analyses of Child 3's speech patterns appear warranted given several unexpected findings concerning her pattern of results on the measures of severity reported. Additional investigation of the stimuli used for Child 1 and 2 in Listening Task 1 is also recommended to attempt to determine why their word identification scores did not differ significantly, but their response times did.
6) As predicted, the main effect of stimulus presentation number was found to be significant for word identification scores and response times, i.e., repeated presentations of the identical word stimulus spoken by children with dysarthria can increase accuracy scores and decrease response times. However, the increase in accuracy scores found is small (in the order of 3%) and lower than what has been found for sentences presented auditorily. This increase is also smaller than what would be considered clinically significant (trivial effect size) and therefore not strong enough to warrant advising children with dysarthria and their communication partners to use this strategy solely in communication breakdowns.

However, these trivial effect sizes are comparable to those reported in published investigations that have found significant effects of repeated presentations on lexical decision-making and word identification tasks. Therefore it appears that repetition must be combined with something else to be an effective repair strategy, at least for children's words that are in the moderate-severe to profound range of severity represented in the conditions of Listening Task 1. Based on the literature reviewed and on the listeners' beliefs reported in this study about what they did to try and better understand the words presented, a combination of repetition with clear speech strategies is recommended as a next area for investigation. Another suggestion for future research is to investigate the effect of repeated presentations in a more realistic context, that is, only for words that the listener does not understand the first time. It is the case in the current study that for some words, some listeners were correct in their first response and then changed their response later. It also possible then, that scores on the second and third presentations are in fact lower than they should be. In reality, if a listener was correct the first time, a repair strategy of repeating would not be necessary. Additional analyses of the data from Listening Task 1 is underway to identify adjusted scores across repetitions for those words that were not identified correctly on the first and second presentations and compare these with the original scores to determine if a different result is found.

7) Unexpectedly, a significant interaction of speech disorder severity level and stimulus presentation number was not found for either word identification scores or response times. In this investigation, a range of speech disorder severity from

mild to severe was desired to test hypotheses about the interaction between talker severity and presentation number. However, the 25 single word stimuli for each talker, presented in noise with a mixed talker design in Listening Task 1, resulted in a severity range from moderate-severe to profound. This prevented a full testing of the hypothesis of the interaction because the mild and moderate levels of severity were not represented. An implication of these findings for future research is that caution must be taken when classifying talker severity, if a full range and even distribution among levels is desired. The levels of talker severity must be identified based on stimuli very similar to the stimuli being used. To test hypotheses about the interaction of speech disorder severity and presentation number in a mixed talker design, using less than ideal SNRs, across the full range of speech disorder severity in young children, future research may need to use children up to seven years of age and include children without speech disorders.

8) The ordering of DME effort ratings reflected that of the *TOCS* sentence intelligibility measure and the vowels and syllable shapes correct scores obtained from responses to Listening Task 1. Child 1 had the lowest ratings of effort and the highest vowel and syllable shape correct scores, Child 4 had the highest ratings of effort and the lowest vowel and syllable shape correct and Child 2 and 3 had scores in the middle. However, unlike the *TOCS* sentence intelligibility measure, but similar to the vowels and syllable shapes correct measures in Listening Task 1, the effort ratings for Child 2 and 3 did not differ from each other. Of interest is that Child 1 had significantly lower effort ratings than Child 2, despite Child 2 having slightly higher word identification scores in Listening

Task 1. This suggests that listeners are using something about the single words other than, or in addition to, the effort involved in the task of identifying a whole word (e.g., vowel accuracy, syllable shape information, voice quality, coarticulation cues, word frequency, phonological neighbourhood), and more than response time alone, in generating their effort ratings. Again, these findings are limited to talker severity levels at the moderate-severe to profound region of the severity continuum, as represented by the word stimuli used in Listening Task 2.

9) Effort ratings by adult listeners with little or no previous exposure to young children with dysarthria using DME were found to have acceptable reliability. This suggests that even when listening to children with dysarthria at the severe end of the talker severity continuum, with as few as eight repeated items in a mixed talker design with added noise, inexperienced listeners can make reliable judgments of effort using DME. A next step would be to compare listener effort ratings using DME and a linear rating scale and plot the responses against each other to identify whether rating listening effort for children with dysarthria is prothetic or metathetic in nature and consequently, the more appropriate method to measure listeners' subjective effort.

10) The predicted negative relationship between word identification scores and response times was found, despite the use of global measures of these two variables, averaged across the four children, for each of the listeners. There appears to be a robust relationship between these two variables that supports previous findings, i.e., listeners with higher word identification scores have shorter response times. A correlational study that uses children with a wide range

of severity of speech disorder, rather than listeners, as subjects is recommended to provide a better test of hypotheses about relationships among word identification scores, response times and effort ratings.

11) Three themes were identified from listeners' responses about what makes words easy or difficult to understand. These themes included articulation/voice characteristics (the most frequently reported theme for what makes words both easy and difficult to understand), word characteristics and other. Many of the articulation and voice characteristics identified are similar to those of clear speech. These results support the observations from previous research that clear speech can increase word identification scores. Six themes were identified for what strategies listeners used to try and understand the words. These themes included finding out unknown from known using phonetic information, using additional repetitions, body and postural adjustments, specific word characteristics, listening to particular parts, increasing mental effort and other. The most frequently reported themes were trying to find out unknown from known using phonetic information and using additional repetitions. These results suggest that listeners in this study had good insights into what makes words easy or difficult to understand and what strategies to use to increase the likelihood of their success in identifying a word, particularly when they knew that they would hear it a second and third time. It was apparent that listeners also invoked the strategy of repetition (i.e., said the word silently or aloud) to help them identify the words, suggesting that this strategy is intuitive for both listeners and talkers when speech is not understood.

FUTURE RESEARCH

This study identified several areas for future investigation. Suggestions include repeating the study and ensuring that the mild-moderate end of the continuum is represented; and treating noise as an independent variable so that its influence on word identification scores, response times and effort ratings and the potential interaction of noise with talker severity and presentation number could be determined. This would likely require including children without speech disorders or adults with or without dysarthria. Ensuring that a range of severity from mild to profound was represented would reveal an interaction among or between pairings of talker severity, noise and repetition number, if present.

For the range of severity investigated in this study, repeated presentations had a significant, but trivial, effect, even between the first and third presentations. From the perspective of coaching children to use effective repair strategies, it would be beneficial to investigate what additional behaviours, that require minimal cognitive effort, would enhance repetition so that the combined effect would yield an effect size that might make a functional difference. For example, the effect of repetition alone versus repetition and use of clear speech strategies could be compared.

Another possibility would be to examine the subset of words that listeners did not identify correctly on the first presentation in Listening Task 1 to determine how successful listeners were in identifying these correctly on subsequent presentations. The current study looked at the total number correct for each presentation number. It is possible that some listeners were correct in their first response and then refined their response later. It is also possible then, that scores on the second and third presentations are in fact lower than they should be. In reality, if a listener was correct the first time, a repair strategy of repeating would not be necessary. Repair strategies are only necessary when the original attempt is not understood.

Use of the DME task to obtain measures of listener effort had acceptable reliability for young professionals in training who are inexperienced in listening to the speech of children with dysarthria. As observed in the Introduction, it is unknown if rating listener effort for understanding the speech of young children with dysarthria is prothetic or metathetic in nature. A follow-up study could be conducted using speech produced by children who represent a wide range of severity of speech disorder where children are the subjects and listeners rate effort under two conditions: DME and equal interval scaling. The ratings under each condition could then be compared to determine which kind of scale appears more appropriate for rating listener effort.

Lastly, the word identification data collected in Listening Task 1 could be used to determine if the word identification scores actually support the listeners' beliefs. The percent correct scores for each word, averaged across listeners, could be determined and the ranking of these could be compared with the ranking of word frequency and phonological neighbourhood density scores for each child to see if the relationships were significant. As well, the effect of familiarity with the talkers could be investigated similarly to the study of Hustad and Cahill (2003). Specifically, listener's scores for the first 25 words identified could be compared with the last 25 identified to determine if the last 25 had significantly higher scores.

REFERENCES

Adobe Systems Incorporated. (2004). *Adobe Audition 1.5.* San Jose, CA: Adobe Systems Incorporated.

Alexander, D., Wetherby, A., & Prizant, B. (1994). The emergence of repair strategies in infants and toddlers. *Seminars in Speech and Language, 18*, 197-212.

American Speech and Hearing Association. (1985). Guidelines for identification of audiometry. *American Speech and Hearing Association Journal, 27,* 409-453.

Ansel, B. M., McNeil, M. R., Hunker, C. J., & Bless, D. M. (1983). The frequency of verbal and acoustic adjustments used by cerebral palsied dysarthric adults when faced with communicative failure. In W. Berry (Ed.), *Clinical dysarthria* (pp. 85-106). San Diego, CA: College Hill Press Inc.

Ashcraft, M. H. (2002). *Cognition* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Austin, D. & Shriberg, L. D. (1997). Lifespan reference data for ten measures of articulation competence using the speech disorders classification system (SDCS). Phonology Project Technical Report No. 3. Retrieved on March 10, 2006 from http://www.waisman.wisc.edu/phonology/search/search.html

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Baddeley, A. D. (1976). *The psychology of memory.* New York: Basic Books, Inc., Publishers.

Balota, D. A. & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language, 24,* 89-106.

Barefoot, S. M., Bochner, J. H., Johnson, B. A., & Eigen, B. A. V. (1993). Rating deaf speakers' comprehensibility: An exploratory investigation. *American Journal of Speech-Language Pathology, 2*, 31-35.

Beliveau, C., Hodge, M. M., & Hagler, P. H. (1995). Effects of supplemental linguistic cues on the intelligibility of severely dysarthric speakers. *Augmentative and Alternative Communication, 11*, 176-186.

Benki, J. R. (2003). Analysis of English nonsense syllable recognition in noise. *Phonetica, 60,* 129-157.

Benki, J. R., & Felty, R. (2005). Recognition of English phonemes in noise. *149th Meeting: Acoustical Society of America, 114*, 2568.

Bentin, S., Kutas, M., & Hillyard, S. A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology, 30,* 161-169.

Bergerbest, D., Ghahremani, D. G., & Gabrieli, J. D. E. (2004). Neural correlates of auditory repetition priming: Reduced fMRI activation in the auditory cortex. *Journal of Cognitive Neuroscience, 16,* 966-977.

Bernthal, J. E. & Bankson, N. W. (1998). *Articulation and phonological disorders* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Berry, W. R., & Sanders, S. B. (1983). Environmental education: The universal management approach for adults with dysarthria. In W. Berry (Ed.), *Clinical dysarthria* (pp. 203-216). San Diego, CA: College Hill Press, Inc.

Beukelman, D. R. & Yorkston, K. M. (1980). Influence of passage familiarity on intelligibility estimates of dysarthric speech. *Journal of Communication Disorders, 13,* 33-41.

Bloom, M. & Fisher, J. (1982). *Evaluating practice: Guidelines for the accountable professional.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Borowsky, R. & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 63-85.

Bostrom, R. N. & Waldhart, E. S. (1988). Memory models and the measurement of listening. *Communication Education, 37,* 1-13.

Bradlow, A. R. & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener- and item-related factors. *Journal of the Acoustical Society of America, 106,* 2074-2085.

Brinton, B., Fujiki, M., Loeb, D. F., & Winkler, E. (1986). Development of conversational repair strategies in response to requests for clarification. *Journal of Speech and Hearing Research, 29,* 75-81.

Brown, C. R., & Rubenstein, H. (1961). Test of response bias explanation of word-frequency effect. *Science, 133,* 280-281.

Caissie, R., & Wilson, E. (1995). Communication breakdown management during cooperative learning activities by mainstreamed students with hearing losses. *Volta Review, 97,* 105-121.

Carlson, N. R., Buskist, W., Enzle, M. E., & Heth, C. D. (2000). *Psychology: The science of behaviour* (Canadian Ed.). Scarborough, ON: Prentice-Hall Canada, Inc.

Ciocci, S. R., & Baran, J. A. (1998). The use of conversational repair strategies by children who are deaf. *American Annals of the Deaf, 143,* 235-245.

Cohen, J. (1988). *Statistical power for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J., Welkowitz, J., & Ewen, R.B. (1991). *Introductory Statistics for the Behavioral Sciences* (4th ed.). New York: Harcourt Brace Jovanovich.

Collins, L. J. (1998). Select MMPI-2 scales and consistency of physical effort during trunk extension exercise as predictors of treatment outcome for functional restoration rehabilitation in chronic back pain patients. Doctoral dissertation, University of Detroit Mercy, 93. (UMI Order #AAI9910820.)

Cooper, J. C. & Cutts, B. P. (1971). Speech discrimination in noise. *Journal of Speech and Hearing Research, 14,* 332-337.

Crandell, C. C. & Smaldino, J. (2002). Room acoustics and auditory rehabilitation technology. In J. Katz, R. F. Burkard, & L. Medwetsky (Eds.), *Handbook of clinical audiology* (pp. 607-630). Philadelphia, PA: Lippincott, Williams, & Willkins.

Crandell, C. C., Smaldino, J., & Flexer, C. (1995). *Sound field FM amplification: Theory and practical applications.* San Diego: Singular Publishing Group.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology, 3,* 255-267.

Deacon, D., Dynowska, A., Ritter, W., & Grose-Fifer (2004). Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology, 41,* 60-74.

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

Dirks, D. D., Morgan, D. E., & Dubno, J. R. (1982). A procedure for quantifying the effects of noise on speech recognition. *Journal of Speech and Hearing Disorders, 47,* 114-122.

Downs, D. W. (1982). Effect of hearing aid use on speech discrimination and listening effort. *Journal of Speech and Hearing Disorders, 47,* 189-193.

Downs, D. W., & Crum, M. A. (1978). Processing demands during auditory learning under degraded listening conditions. *Journal of Speech & Hearing Research, 21,* 702-714.

Doyle, M. C., Rugg, M. D., & Wells, T. (1996). A comparison of the electrophysiological effects of formal and repetition priming. *Psychophysiology, 33, 132-147.*

Eadie, T. L. & Doyle, P. C. (2004). Scaling of voice pleasantness and acceptability in tracheoesophageal speakers. *Journal of Voice, 19,* 371-383.

Erber, N. P. (Ed.). (1996). *Communication therapy for adults with sensory loss* (2nd ed.). Clifton Hill, Victoria, Australia: Clavis Publishing.

Feltz, C., McClure, K., & O'Hare, J. (2002). Speech Intelligibility Probe for Children with Cleft Palate (SIP-CCLP): A Preliminary Assessment of Validity and Reliability. Unpublished manuscript, University of Alberta, Edmonton, Canada.

Field, A. (n.d.). A bluffer's guide to ... sphericity. *BPS-MSC Newsletter, 6, 1.* Retrieved on January 30, 2007 from http://www.ling.ohio-state.edu/~kjohnson/ling795q/papers/sphericity.pdf

Gagne, J., Stelmacovich, P., & Yovetich, W. (1991). Reactions to requests for clarification used by hearing-impaired individuals. *Volta Review, 93,* 129-143.

Gagne, J. -P. & Wyllie, K. M. (1989). Relative effectiveness of three repair strategies on the visual-identification of misperceived words. *Ear and Hearing, 10,* 368-374.

Gallagher, T. M. (1977). Revision behaviours in the speech of normal children developing language. *Journal of Speech and Hearing Research, 20,* 293-302.

Ghahremani, D. (2005). Experience-driven neural plasticity: Evidence from neuroimaging of repetition priming. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 65* (11-B), 5566. [full text]

Givens, G. & Greenfield, D. (1982). Revision behaviors of normal and hearing-impaired children. *Ear and Hearing, 3*, 274-279.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28,* 501-518.

GoldWave Incorporated. (2005). *GoldWave v5.12.* Retrieved on March 22, 2006 from http://www.goldwave.com

Gordon-Brannan, M. & Hodson, B. W. (2000). Intelligibility/Severity measurements of prekindergarten children's speech. *American Journal of Speech-Language Pathology, 9,* 141-150.

Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *The Quarterly Journal of Experimental Psychology, 58A,* 981-998.

Griffiths, R. (1990). Speech rate and nonnative speaker comprehension. *Language Learning, 40,* 311-336.

Hanson, E. K., Yorkston, K. M., & Beukelman, D. R. (2004). Speech supplementation techniques for dysarthria: A systematic review. *Journal of Medical Speech-Language Pathology, 12*(2), ix-xxix.

Hicks, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech, Language, and Hearing Research, 45*, 573-584.

Hodge, M. M. (1996, February). Measuring speech intelligibility in young children with dysarthria. Paper presented at the Biennial Conference on Motor Speech, Amelia Island, FL.

Hodge, M., Spooner, L., & Wellman, L. (1999). Increasing the clinician "friendliness" of measuring children's speech intelligibility. Annual Conference, Speech-Language-Hearing Association of Alberta, Edmonton, Alberta.

Hodge, M. M. & Wellman, L. (1999). Management of children with dysarthria. In A. Caruso & E. Strand (Eds.). *Clinical management of motor speech disorders in children* (pp. 209-280). New York: Thieme Medical Publishers, Inc.

Holcomb, P. J., Anderson, J., & Grainger, J. (2005). An electrophysiological study of cross-modal repetition priming. *Psychophysiology, 42*, 493-507.

Humes, L. E., Christensen, L., Thomas, T., Bess, F. H., Hedley-Williams, A., & Bentler, R, (1999). A comparison of the aided performance and benefit provided by a linear and a two-channel wide dynamic range compression hearing aid. *Journal of Speech, Language, and Hearing Research, 42,* 65-69.

Hustad, K. C., & Cahill, M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology, 12,* 198-208.

Hustad, K. C., & Beukelman, D. R. (2001). Effects of linguistic cues and stimulus cohesion on intelligibility of severely dysarthric speech. *Journal of Speech, Language, and Hearing Research, 44*, 407-510.

Jamieson, D. G., Kranjc, G., Yu, K., & Hodgetts, W. E. (2004). Speech intelligibility of young school-aged children in the presence of real-life classroom noise. *Journal of the American Academy of Audiology, 15,* 508-517.

Kent, R. D. (1988). Speech intelligibility. In D. Yoder & R. Kent (Eds.). *Decision making in speech-language pathology* (pp. 120-121). Toronto: B. C. Decker.

Kent, R. D. (1992). Speech intelligibility and communicative competence in children. In A. P Kaiser & D. B. Gray (Eds.). *Enhancing children's communication: Research foundations for intervention* (Vol. 2, pp. 223-239). Baltimore: Brookes Publishing Co.

Kent, R. D., Miolo, G., & Bloedel, S. (1994). The intelligibility of children's speech: A review of evaluation procedures. *American Speech-Language-Hearing Association, 3,* 81-95.

Kent, R. D., Weismer, G., Kent, J. F., Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54,* 482-499.

Kirk, R. (1968). *Experimental design: Procedures for the behavioral sciences*, Belmont. CA: Brooks/Cole Publishing Company.

Konefal, J. A., & Fokes, J. (1984). Linguistic analysis of children's conversational repairs. *Journal of Psycholinguistic Research, 13,* 1-11.

Kučera, H. & Francis, W. N. (1967). *A computational analysis of present-day American English*. Providence, R. I.: Brown University Press.

Kutas, M. & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161-163.

Laing, A. C., Frazer, M. B., Cole, D. C., Kerr, M. S., Wells, R. P., & Norman, R. W. (2005). Study of the effectiveness of a participatory ergonomics intervention in reducing worker pain severity through physical exposure pathways. *Ergonomics, 48*, 150-170.

Larsby, B., Hällgren, M., Lyxell, B., & Arlinger, S. (2005). Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *International Journal of Audiology, 44,* 131-143.

Lorch, R. F. & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 149-157.

Luce, P. A. & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19,* 1-36.

Mackersie, C., Neuman, A. C., & Levitt, H. (1999). A comparison of response time and word recognition measures using a word-monitoring and closed-set identification task. *Ear and Hearing, 20,* 140-148.

Macromedia. (2001). *Authorware 6.0.* San Francisco, CA: Macromedia, Inc.

Markham, D., & Hazan, V. (2004). The effect of talker- and listener-related factors on intelligibility for a real-word, open-set perception test. *Journal of Speech, Language, and Hearing Research, 47,* 725-737.

Massaro, D. W. (1989). A fuzzy logical model of speech perception. In D. Vickers & P. L. Smith (Eds.), *Human information processing: Measures, mechanisms and models* (pp. 367-379). Amsterdam: North-Holland.

Max, L. & Onghena, P. (1999). Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research. *Journal of Speech, Language, and Hearing Research, 42,* 261-270.

McGurk, J. & MacDonald, T. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

McHenry, M., Whatman, J., & Pou, A. (2002). The effect of botulinum toxin A on the vocal symptoms of spastic dysarthria: A case study. *Journal of Voice, 16,* 124-131.

McNaughton, D., Fallon, K., Tod, J., Weiner, F., & Neisworth, J. (1994). Effect of repeated listening experiences on the intelligibility of synthesized speech. *Augmentative and Alternative Communication, 10,* 161-168.

Milenkovic, P. & Read, C. (1992). *CSpeech Version 4.0, Laboratory Version.* Madison, WI: University of Wisconsin-Madison.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41,* 329-335.

Miller, G. A., & Nicely, P. A. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27,* 338-352.

Mimura, M., Verfaellie, M., & Milberg, W. (1997). Repetition priming in an auditory lexical decision task: Effects of lexical status. *Memory and Cognition, 25,* 819-825.

Monsen, R. B. (1983). The oral speech intelligibility of hearing-impaired talkers. *Journal of Speech and Hearing Disorders, 48, 286-296.*

Most, T. (2002). The use of repair strategies by children with and without hearing impairment. *Language, Speech, and Hearing Services in Schools, 33*, 112-123.

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

Munro, M. J. & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45,* 73-97.

Munro, M. J. & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38,* 289-306.

Norman, G. R. & Streiner, D. L. (1997). *PDQ Statistics (2nd Ed.).* Hamilton, ON: B.C. Decker Inc.

Ogata, Y., Yokote, E., Hashimoto, M., Fukuda, T., Nakano, Y., et al. (2000). A study on resource-based relative values for visiting nursing services. *Japanese Journal of Public Health, 47,* 973-989. [abstract]

Ogata, Y., Kobayashi, Y., Fukuda, T., Mori, K., Hashimoto, M., & Otosaka, K. (2004). Measuring relative work values for home care nursing services in Japan. *Nursing Research, 53*, 145-153.

Peng, S. C., Spencer, L. J., & Tomblin, J. B. (2004). Speech intelligibility of pediatric cochlear implant recipients with seven years of device experience. *Journal of Speech Language & Hearing Research, 47,* 1227-1236.

Picard, M., & Bradley, J. S. (2001). Revisiting speech interference in classrooms. *Audiology, 40,* 221-244.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing: I. intelligibility differences between clear and conversational speech. *Journal of Speech & Hearing Research, 28,* 96-103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing: II. acoustic characteristics of clear and conversational speech. *Journal of Speech & Hearing Research, 29,* 434-445.

Pickett, J. (1999). *The acoustics of speech communication: Fundamentals, speech perception theory, and technology.* Needham Heights, MA: Allyn & Bacon.

Pratt, R. L. (1981). On the use of reaction time as a measure of intelligibility. *British Journal of Audiology, 15,* 253-255.

Preminger, J. W. & Van Tasell, D. J. (1995). Quantifying the relation between speech quality and speech intelligibility. *Journal of Speech and Hearing Research, 38,* 714-725.

Rakerd, B., Seitz, P. F., & Whearty, M. (1996). Assessing the cognitive demands of speech listening for people with hearing losses. *Ear and Hearing, 17,* 97-106.

Reiger, J. (2002). *The effect of automatic speech recognition on speaking behaviors in individuals with and without spinal cord injury.* Doctoral Thesis. University of Alberta, Edmonton, Alberta, Canada.

Rogers, C. L., Dalby, J., & Nishi, K. (2004). Effects of noise and proficiency on intelligibility of Chinese-accented English. *Language and Speech, 47,* 139-154.

Ross, M. (Ed.). (1992). *FM auditory training systems: Characteristics, selection, and use.* Timonium, Maryland: York Press, Inc.

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology, 22,* 642-647.

Samar, V. J. & Metz, D. E. (1988). Criterion validity of speech intelligibility rating-scale procedures for the hearing-impaired population. *Journal of Speech and Hearing Research, 31,* 307-316.

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. *Journal of Speech and Hearing Research, 24,* 441-445.

Schmidt, M. (1984). Intelligibility and the child with multiple articulation deviations. In H. Winitz (Ed.), *Treating articulation disorders: For clinicians by clinicians* (pp. 67-73). Baltimore, MD: University Park Press.

Shriberg, L. D. (1986). *PEPPER: Programs to examine phonetic and phonologic evaluation records*. Hillsdale, NJ: Lawrence Erlbaum.

Shriberg, L. D. & Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders, 47,* 242-256.

Solomon, N. & Robin, D. (2005). Perceptions of effort during handgrip and tongue elevation in Parkinson's disease. *Parkinsonism & Related Disorders, 11,* 353-361.

Southwood, H. (1996). Direct magnitude estimation and interval scaling of naturalness and bizarreness of the dysarthria associated with amyotrophic lateral sclerosis. *Journal of Medical Speech-Language Pathology, 4,* 13-25.

Southwood, M. H. & Weismer, G. (1993). Listener judgements of the bizarreness, acceptability, naturalness, and normalcy of the dysarthria associated with amyotrophic lateral sclerosis. *Journal of Medical Speech-Language Pathology, 1,* 151-161.

SPSS for Windows (Ver. 13) [Computer Software]. (2004). www.spss.com

Stevens, S. S. (1975). *Psychophysics.* New York: Wiley.

Stone, G. O. & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception & Performance, 19,* 744-774.

Thorndike, E. L. & Lorge, I. (1944). *The teacher's word book of 30, 000 words.* Columbia University: Teachers College.

Tomasello, M., Farrar, M. J., & Dines, J. (1984). Children's speech revisions for a familiar and an unfamiliar adult. *Journal of Speech & Hearing Research, 27,* 359-363.

TyeMurray, N., Purdy, S. C., Woodworth, G. C., & Tyler, R. S. (1990). Effects of repair strategies on visual identification of sentences. *Journal of Speech & Hearing Disorders, 55,* 621-627.

Volden, J. (2004). Conversational repair in speakers with autism spectrum disorder. *International Journal of Language and Communication Disorders, 39,* 171-189.

Weismer, G., & Laures, J. (2002). Direct magnitude estimation of speech intelligibility: Effects of a chosen standard. *Journal of Speech, Language, Hearing Research*, 45, 421-433.

Whitehill, T. L. & Wong, C. C-Y. (2006). Contributing fact ors to listener effort for dysarthric speech. *Journal of Medical Speech-Language Pathology, 14,* 335-341.

Whitehill, T. L., Lee, A. S. Y., & Chun, J. C. (2002). Direct magnitude estimation and interval scaling of hypernasality. *Journal of Speech, Language and Hearing Research, 45,* 80-88.

Wilcox, K. A. & Morris, S. R. (1999). *The Children's Speech Intelligibility Measure.* San Antonio, TX: The Psychological Corporation.

Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders, 11,* 499-512.

Yorkston, K. M. & Beukelman, D. R. (1980). A clinician-judged technique for quantifying dysarthric speech based on single-word intelligibility. *Journal of Communication Disorders, 13,* 15-31.

Yorkston, K. M., & Beukelman, D. R. (1981). *Assessment of Intelligibility of Dysarthric Speech.* Austin, TX: Pro-Ed, Inc.

Yorkston, K. M., Beukelman, D. R., Strand, E. A., & Bell, K. R. (1999). *Management of motor speech disorders in children and adults.* Austin, Texas: Pro-Ed, Inc.

APPENDICES

Appendix A – Power Calculation

This power calculation followed the procedures described by Norman & Streiner, (1997) and is based on the effect of repeated exposures on speech identification scores. (This effect is expected to be slightly smaller than the effect of severity of speech disorder on word identification scores. No relevant information to estimate the expected effect size of repeated exposures on listeners' response time or of severity of speech disorder on DME measures of listener effort was located.) Using an alpha level of .01 and a beta level of .2, this calculation takes into account the smallest difference that is thought to be important for an intelligibility gain (estimated at 10%) and a standard deviation (estimated at 10%) for the same listeners judging a given child with dysarthria on a word identification task, where a child may range in severity from mild to severe (Hodge, 1996). An intelligibility gain of 10% reflects an average of 2.5 more words, from twenty-five words, being understood correctly after repeated presentations. This value was chosen so that the difference between the first presentation and a repeated presentation was greater than reported test-retest reliability differences. On a comparable single word intelligibility measure for adults (Yorkston & Beukelman, 1981), test-retest reliability is reported to be within 8% with 95% confidence. Therefore values greater than this might be considered a meaningful change in intelligibility scores.

Equation to calculate effect size (Cohen,Welkowitz & Ewen, 1981):

$$\text{Effect size} = (\mu_1 - \mu_2)/SD$$

$$= 10\%/10\% \quad = 1.00$$

Power Calculation

For an effect size of 1.00, an alpha level of .01, a power level of .80 and the factor group with four levels, the power tables in Kirk (1968) were used to determine that the minimum number of subjects required per cell is 33. As this study is a within-subjects design, 33 subjects in total were required.

Appendix B – Shriberg's Conventions for Calculating PCC, PVC and PPC

Percent of Consonants Correct (PCC) and Percentage of Vowels/Diphthongs Correct (PVC) are calculated by identifying each consonant and vowel/diphthong respectively in a continuous speech sample as omitted, substituted, distorted or correct, The number correct divided by the total number yields a percentage score (Austin & Shriberg, 1997). To create Percent Phonemes Correct (PPC), the transcriber combines the PCC and PVC raw scores for a total number of phonemes correct and then calculates a percentage. Shriberg's conventions for deciding how many words from a continuous speech sample should be included for a representative sample for calculation of PCC were followed. The minimum number of words required using any of these three procedures was used:

- 90 non-questionable first occurrence words
- 70 utterances each containing at least one non-questionable word
- 225 non-questionable words

These words were obtained from a spontaneous speech sample recorded from each of the four children.

Appendix C – Word List for Listening Task 1

Word	Child 1[a]	Child 2[b]	Child 3[c]	Child 4[d]
Practice Items				
1 hoot				
2 pup				
Test Items				
1	bead	back	beach	bait
2	bet	bag	boo	bee
3	chew	chip	bud	chop
4	come	comb	bug	cub
5	don	dye	chin	down
6	fees	feet	duck	fit
7	full	fill	fat	foot
8	heat	hen	hoe	hatch
9	jaw	jay	jam	hid
10	log	leap	key	hot
11	mow	lip	lose	jab
12	nip	low	mud	lawn
13	pad	match	nib	line
14	pan	nap	pain	mug
15	pin	peas	paw	nose
16	shoot	pop	pipe	pen
17	sue	pot	pull	pooh
18	tap	pout	shot	pool
19	thick	shut	shout	sheet
20	town	sick	suit	shoe
21	tub	thin	thing	sing
22	type	toad	tin	thumb
23	wait	tongue	toes	tip
24	watch	tube	top	toe
25	wing	whoa	wet	walk

Continued on next page…

	Child 1[a]	Child 2[b]	Child 3[c]	Child 4[d]
Phonological Neighbourhood[e]				
mean	21.00	22.00	22.20	21.00
SD	5.71	5.83	5.60	4.94
Phonetic Frequency[f]				
mean	81.88	59.04	85.60	62.36
SD	213.52	199.44	330.31	199.70
Lorge Written Frequency[g]				
mean	683.72	664.33	567.60	591.48
SD	1251.51	1328.31	1256.87	1141.40
Number in most frequent 2000[h]				
	12	12	13	13
Number in most frequent 5000[i]				
	21	23	21	21

[a]Child 1's speech severity rating is Mild. [b]Child 2's speech severity rating is Mild-Moderate. [c]Child 3's speech severity rating is Moderately-Severe. [d]Child 4's speech severity rating is Severe. [e]Defined as the number of words that differ in one phoneme from the target word (C. Westbury, personal communication, April, 2006). [f]Defined as the spoken frequency of the words out of 1,000,000 words (Westbury, personal communication, April, 2006 from Baayen et al., 1995). [g]Defined as the number of occurrences of each word in approximately 4 ½ million words taken from "recent and popular magazines" (Thorndike & Lorge, 1944). [h]Defined as the written frequency of the word, or, the number of times each word was found written out of a total number of 1,000,000 words from a summary of four counts of written frequency. Total number of words (out of 25) that were of the 2000 most frequent words (Thorndike & Lorge, 1944). [i]Total number of words (out of 25) using the same definition found in (h) that were of the 5000 most frequent words (Thorndike & Lorge, 1944).

Appendix D – Phonetic Balancing Listening Task 1

	Child 1[a]	Child 2[b]	Child 3[c]	Child 4[d]
Syllable Shape				
CVC	21	21	21	21
CV	4	4	4	4
TOTAL	**25**	**25**	**25**	**25**
Initial Consonant				
p	3	4	4	3
b	2	2	4	2
m	1	1	1	1
t	4	3	3	2
d	1	1	1	1
n	1	1	1	1
ch	1	1	1	1
dj	1	1	1	1
f	2	2	1	2
s	1	1	1	1
sh	1	1	2	2
l	1	3	1	2
k	1	1	1	1
w	3	1	1	1
h	1	1	1	3
th	1	1	1	1
TOTAL	**25**	**25**	**25**	**25**
Final Consonant				
p	4	5	2	2
b	1	1	1	2
m	1	1	1	1
t	3	4	5	5
d	2	1	2	1
n	4	2	3	4
ch	1	1	1	1
z	1	1	2	1
l	1	1	1	1
k	1	2	1	1
g	1	1	1	1
ng	1	1	1	1
TOTAL	**21**	**21**	**21**	**21**

Continued on next page…

	Child 1[a]	Child 2[b]	Child 3[c]	Child 4[d]
Vowel				
a (hot)	4	2	3	4
o (boat)	1	4	2	2
u (boot)	3	1	3	3
uh (cup)	3	2	5	3
u (foot)	0	0	0	1
au (house)	1	1	1	1
i (pie)	3	3	2	2
I (pit)	4	5	4	4
E (pet)	1	1	1	1
ae (pat)	3	4	2	2
ai (my)	1	1	1	1
e (bait)	1	1	1	1
tense	12	11	11	12
lax	11	12	12	11
high	10	9	9	10
mid	6	8	9	7
low	7	6	5	6
front	12	14	10	10
central	3	2	5	3
back	8	7	8	10
dipthongs	2	2	2	2
TOTAL	**25**	**25**	**25**	**25**

[a]Child 1's speech severity rating is Mild. [b]Child 2's speech severity rating is Mild-Moderate. [c]Child 3's speech severity rating is Moderately-Severe. [d]Child 4's speech severity rating is Severe.

Appendix E – Word List for Listening Task 2

Word	Child 1[a]	Child 2[b]	Child 3[c]	Child 4[d]
Modulus Talker				
1 pat				
Practice Items				
1	boot	pig	sheep	hug
Test Items				
1	bead	back	boo	bee
2	come	dye	duck	down
3	heat	feet	fat	hid
4	mow	peas	mud	mug
5	pad	pot	paw	pen
6	shoot	shut	shout	sheet
7	sue	thin	thing	sing
8	thick	toad	tin	thumb
9	town	tongue	toes	toe
10	wing	whoa	wet	walk
Phonological Neighbourhood[e]				
mean	22.20	22.70	22.10	20.60
SD	6.42	5.46	5.80	3.95
Phonetic Frequency[f]				
mean	143.60	109.50	170.10	118.40
SD	330.32	325.19	533.11	311.40
Lorge Written Frequency[g]				
mean	948.70	1097.89	853.90	912.90
SD	1829.65	2110.41	1967.24	1734.35
Number in most frequent 2000[h]				
	6	5	4	6
Number in most frequent 5000[i]				
	8	9	9	9

[a]Child 1's speech severity rating is Mild. [b]Child 2's speech severity rating is Mild-Moderate. [c]Child 3's speech severity rating is Moderately-Severe. [d]Child 4's speech severity rating is Severe. [e]Defined as the number of words that differ in one phoneme from the target word (C. Westbury, personal communication, April, 2006). [f]Defined as the spoken frequency of the words out of 1,000,000 words (Westbury, personal communication, April, 2006 from Baayen et al., 1995). [g]Defined as the number of occurrences of each word in approximately 4 ½ million words taken from "recent and popular magazines" (Thorndike & Lorge, 1944). [h]Defined as the written frequency of the word, or, the number of times each word was found written out of a total number of 1,000,000 words from a summary of four counts of written frequency. Total number of words (out of 25) that were of the 2000 most frequent words (Thorndike & Lorge, 1944). [i]Total number of words (out of 25) using the same definition found in (h) that were of the 5000 most frequent words (Thorndike & Lorge, 1944).

Appendix F – Phonetic Balancing Listening Task 2

	Child 1[a]	Child 2[b]	Child 3[c]	Child 4[d]
Syllable Shape				
CVC	8	8	8	8
CV	2	2	2	2
TOTAL	**10**	**10**	**10**	**10**
Initial Consonant				
p	1	2	1	1
b	1	1	1	1
m	1	0	1	1
t	1	2	2	1
d	0	1	1	1
f	0	1	1	0
s	1	0	0	1
sh	1	1	1	1
k	1	0	0	0
w	1	1	1	1
h	1	0	0	1
th	1	1	1	1
TOTAL	**10**	**10**	**10**	**10**
Final Consonant				
m	1	0	0	1
f	0	0	0	0
t	2	3	3	1
d	2	1	1	1
n	1	1	1	1
z	0	1	1	0
k	1	1	1	1
g	0	0	0	1
ng	1	1	1	2
TOTAL	**8**	**8**	**8**	**8**

Continued on next page…

	Child 1[a]	Child 2[b]	Child 3[c]	Child 4[d]
Vowel				
a (hot)	0	1	1	1
o (boat)	1	2	1	1
u (boot)	2	0	1	0
uh (cup)	1	2	2	2
au (house)	1	0	1	1
i (pie)	2	2	0	2
I (pit)	2	1	2	2
E (pet)	0	0	1	1
ae (pat)	1	1	1	0
ai (my)	0	1	0	0
	10	10	10	10
tense	5	5	3	4
lax	4	4	6	5
high	6	3	3	4
mid	2	4	4	4
low	1	2	2	1
front	5	4	4	5
central	1	2	2	2
back	3	3	3	2
dipthongs	1	1	1	1
TOTAL	**10**	**10**	**10**	**10**

[a]Child 1's speech severity rating is Mild. [b]Child 2's speech severity rating is Mild-Moderate. [c]Child 3's speech severity rating is Moderately-Severe. [d]Child 4's speech severity rating is Severe.

Appendix G – Recruitment Poster (Education)
(PRINTED ON LETTERHEAD)

Education Students Needed
for a Study of Children's Speech

✓Participation involves listening to children with speech disorders.
✓Results of the study will identify strategies to understand these children better.
✓Participation will take place in Corbett Hall and last 1 hour.
✓Participants will be reimbursed for parking/travel. costs.

Participants Must:

- Be between 18 and 35 years of age
- Have normal hearing (we will screen your hearing)
- Be monolingual speakers of Canadian English
- Be a student in the Faculty of **Education,** University of Alberta

Contact Kim to participate: kjc@ualberta.ca

Appendix H – Information Letter
(PRINTED ON LETTERHEAD)

INFORMATION LETTER

Title of Research Study: Effects of Signal Clarity and Dose on Listener Judgments of the Speech Intelligibility of Young Children with Dysarthria

Principal Investigator: Dr. Megan Hodge, Professor, Department of Speech Pathology and Audiology, University of Alberta Telephone: 492-5898

Co-Investigator: Kimberley Cote-Reschny, Graduate Student, Department of Speech Pathology and Audiology, University of Alberta. Telephone: 492-0833; Email: kjc@ualberta.ca

Purpose and Background

Dysarthria is a speech disorder that results from conditions that impair muscle function for speech. An example of such a condition is cerebral palsy. When their speech is not understood, persons with dysarthria often repeat what they just said. Background noise can also interfere with how much listeners understand. The purpose of this project is to examine if hearing a word more than one time helps listeners to understand children with dysarthria. Words spoken by children with dysarthria will be presented several times in background noise. This project will also provide information about the mental effort that it takes to understand these words. We are recruiting 33 adult listeners to participate in this study. This is a graduate student project supervised by Dr. Megan M. Hodge.

Procedure

If you agree to take part in this study, you will come to a sound booth at Corbett Hall at the University of Alberta. First, we will ask you several questions to confirm your eligibility to participate. After reading this information letter you will be asked to sign the consent form. Then we will screen your hearing. If you do not pass the hearing screening you will be provided with information about obtaining a hearing evaluation and your participation will end. If you pass the hearing screening, you will be given a short typing task to complete using a computer keyboard. Then you will begin the listening tasks. You will hear words spoken by several children who have a speech disorder. Words will be played through speakers from a computer. You will hear each word three times. You will type into the computer what you think the child is saying. Then you will rate how much effort it takes to understand a subset of the words spoken by these same children. You will compare how much effort was required to understand each word to how much effort it took to understand a "model" word. Then you will be asked three general questions about the first listening task. Your participation will take about one hour in total.

Risks and Benefits
There are no known risks for participating in this study. There are no direct personal benefits for participating in this study. However, results will be used to guide recommendations given to children and their families about what to do when their speech is not understood.

Questions
You can ask any questions about the research project and its procedures at any time.

Confidentiality
All information will be held confidential, except when professional codes of ethics or legislation require reporting. The information collected will be kept for at least five years after the study is done. The information will be kept in a secure area. This is in a locked filing cabinet in Dr. Hodge's laboratory at Corbett Hall. Your name or any other identifying information will not be attached to the information collected. Only the principal investigator and her research assistants/students will have access to the data you provide. Your name will also never be used in any presentations or publications of the study results.

The information gathered for this study may be looked at again in the future. It may be used to help us answer other study questions. If so, the study will be reviewed by the ethics board to ensure the information is used ethically.

Freedom to Withdraw
During the session, you have the right to end your participation at any time. There is no penalty for withdrawing early. If you show up for the study you will be eligible for the parking/travel reimbursement of ten dollars.

Additional Contacts:
If you have any concerns about any aspect of the study, please contact the Office of the HREB at 492-0302. If you have any questions for the supervisor of this project, you can contact her at 492-5898 (Dr. Megan Hodge). If you have any questions for the student, you can contact her at 492-0833 (Kimberley Cote-Reschny).

Appendix I – Consent Form
(PRINTED ON LETTERHEAD)

CONSENT FORM

Title of Research Study: Effects of Signal Clarity and Dose on Listener Judgments of the Speech Intelligibility of Young Children with Dysarthria

Principal Investigator: Dr. Megan Hodge, Professor, Department of Speech Pathology and Audiology, University of Alberta Telephone: 492-5898

Co-Investigator: Kimberley Cote-Reschny, Graduate Student, Department of Speech Pathology and Audiology, University of Alberta. Telephone: 492-0833; Email: kjc@ualberta.ca

	Yes	No
Do you understand that you have been asked to be in a research study?	☐	☐
Have you read and received a copy of the attached Information Letter?	☐	☐
Do you understand the benefits and risks involved in taking part in this research study	☐	☐
Have you had an opportunity to ask questions and discuss this study?	☐	☐
Do you understand that you are free to withdraw from the study at any time, without having to give a reason?	☐	☐
Has the issue of confidentiality been explained to you?	☐	☐
Do you understand who will have access to the data you provide?	☐	☐

Who explained this study to you? ______________________________

I agree to take part in the study. Yes: ☐ No: ☐

Signature of Research Participant: ____________________________

Printed Name: ______________________________

Date: ______________________

I believe that the person signing this form understands what is involved in the study and voluntarily agrees to participate.

Signature of Investigator: _____________________

Appendix J – Listening Task 1

Listener Protocol

Introduction Screen: Participant's Identification Code is entered. Two separate randomizations are created, one for the 100 test items for Listening Task 1; one for the 40 items for Listening Task 2.

Listening Task 1

Verbal Instructions: "You are going to hear words spoken by children with a speech disorder. These children have a range of severity of speech disorder, so some are very difficult to understand. Just try your best. Type in the word you think the child is saying. If you can't decide between two words, make your best guess. If you can't recognize it as a real word, type the letters that match the sounds you hear. So if your best guess is a nonsense word, that is ok. There will also be background noise playing. Just try to listen to what the child is saying and not the rest. You will hear each word three times in a row, so you will have three chances to type in what you hear the child saying. Don't worry about capitals or punctuation. It is possible that you will hear a word repeated later on in the task. A break screen will come up twice during this task to give you a bit of a break. Please take at least a 30 second break to stretch, drink some water, etc. You can take up to 2 minutes and then press continue to move on. When you are done this part, a screen will come up that asks you to continue to the rating task. I'll come back in and see how it is going and then you can continue."

Priming Words Screen

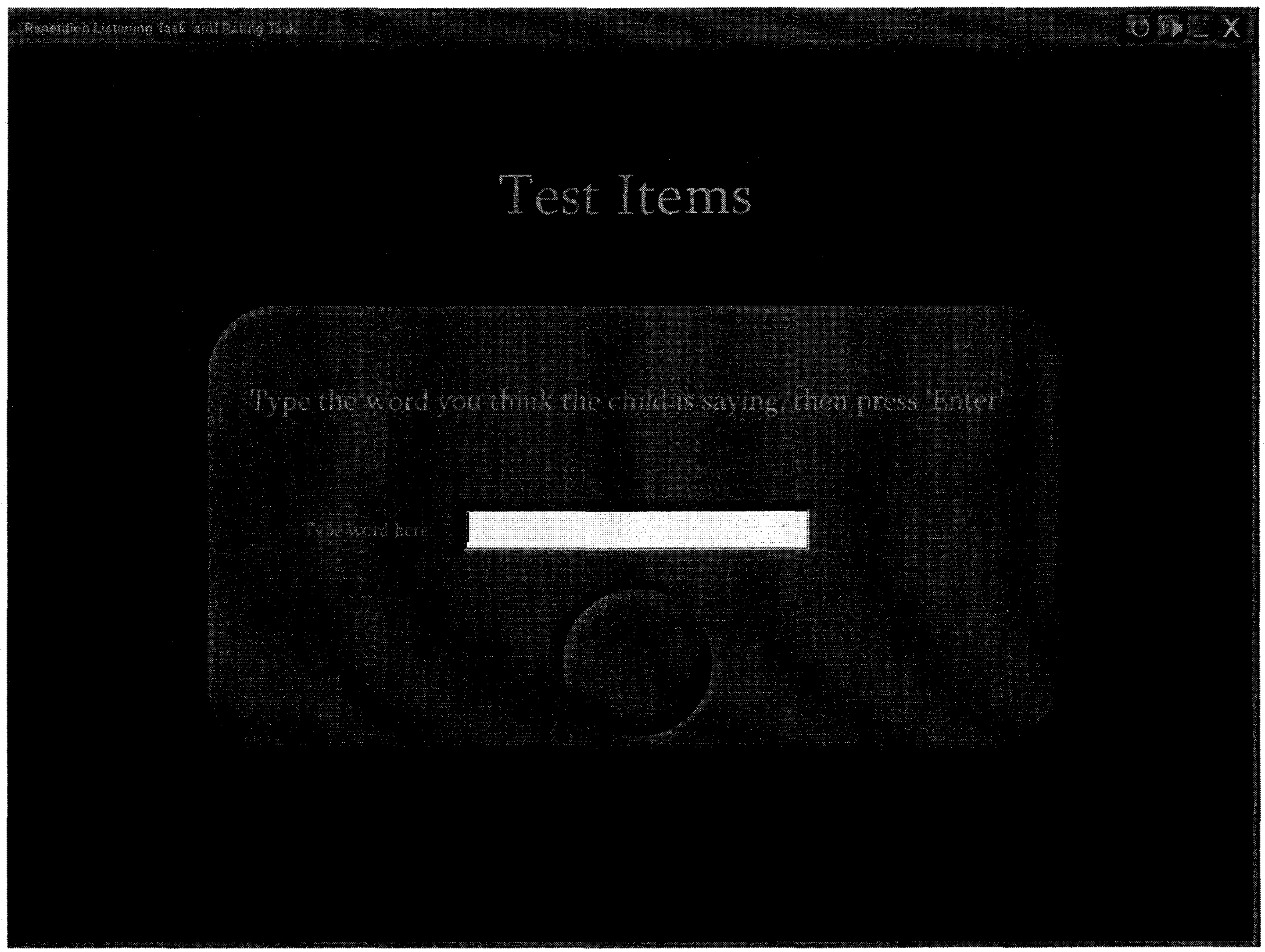
Written Instructions: "You will hear 100 individual words spoken by different children with a motor speech disorder. Please type the word you think the child is saying. Press 'Enter' when you have completed your entry. You will hear each word three times in a row. You may type in the same word each time if you think that it is the word the child said. You may also change your mind and type in a different word. If you have any questions, please feel free to ask."

Practice Items: Listener heard 2 words, each repeated three times consecutively. They typed in the word they heard (total of 6 word entries)

Test Items: The listener heard the unique randomization of the 100 test Items, each repeated three times consecutively. They typed in what they heard after each one (Total of 300 word entries)

Appendix K – Priming Word Screen

Appendix L – Listening Task 1 Screen



Appendix M – Listening Task 2

Listening Task 2

Verbal Instructions: "You are going to hear some of the same words spoken by those same children. This time, we want you to pay attention to the amount of effort it takes to understand them. Read the instructions on the screen and ask if you have any questions."

Written Instructions: "You are going to be presented with a number of words produced by children with a motor speech disorder. Your task is to judge how much EFFORT is required to listen to each speech sample. The definition of EFFORT is 'the amount of cognitive resources or mental energy required to understand what is being said'. The task is to assign numbers to each sample."

"The first item is the standard stimulus, or modulus. Give this item a rating of 100. Some of the items that you will hear will require more effort or less effort. Please rate each speech sample with a number that reflects the degree of effort relative to the modulus. That is, assign a number to each speech sample that reflects how much more or less effort is required to listen to the sample compared to the effort required to listen to the modulus. The more effort required compared to the modulus, the bigger your number response; the less effort compared to the modulus, the smaller your number response. Assign numbers in such a way that they reflect your subjective impression. If a speech sample requires twice as much effort, assign the number 200. If the speech sample requires half as much effort, assign the number 50. You may use as many numbers as you wish as long as they correspond to your perception of how much effort is required to listen to the speech sample compared to the modulus

"Remember, you are rating the amount of EFFORT required to listen to the speech of children with a motor speech disorder. This is relative to the first speech sample. So the more effort, the higher the numerical rating should be relative to the first sample. You will hear each item only once. The stimuli will be grouped into blocks of 10. After every 10 words, you will be reacquainted with the modulus."

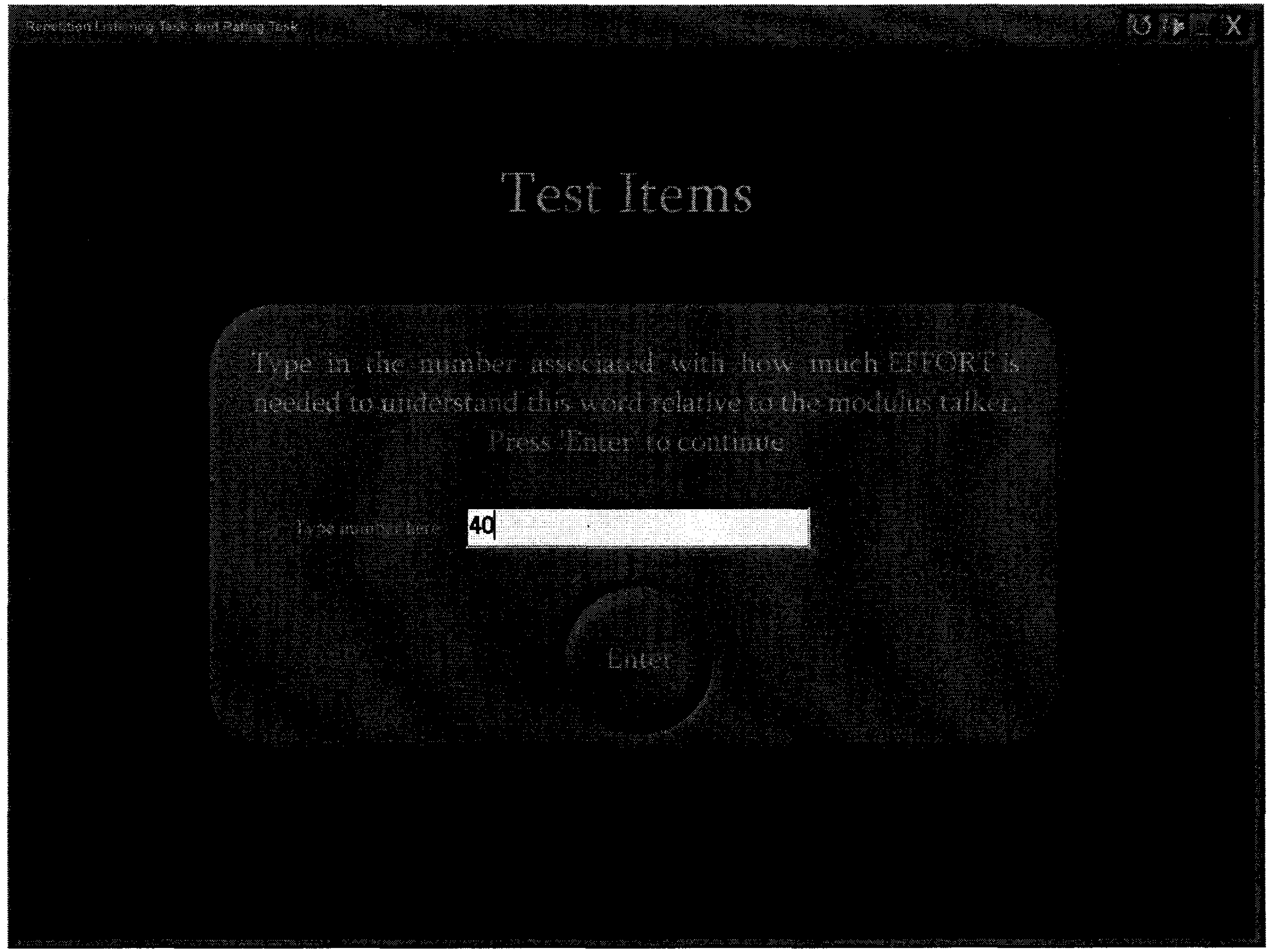
Modulus Talker: The listener heard one word spoken by the modulus talker. They were told that the amount of effort required to understand this talker is 100.

Practice Items: The listener heard four words, one spoken by each of the four children who provided test words. They typed in the number associated with how much EFFORT was needed to understand each word, relative to the modulus talker.

Test Items: The listener heard ten words spoken by the same children involved in Listening Task 1. They heard 8 of these words (two from each child) a second time for the purpose of intra-rater reliability. They typed in the number associated with how much EFFORT was needed to understand each word, relative to the modulus talker.

Re-orientation to the Modulus Talker: The listener heard the original word spoken by the modulus talker after rating 10, 20, 30 and 40 test items. They were reminded that the amount of effort required to understand this talker was 100.

Appendix N – Listening Task 2 Screen



Appendix O – Effect Size Calculations

Effect of Talker Severity on Word Identification Scores (smallest significant difference)

$SD_{pooled} = \sqrt{(SD_{C2})^2 + (SD_{C3})^2}$
$SD_{pooled} = \sqrt{(1.79)^2 + (2.23)^2}$
$SD_{pooled} = 2.86$
Effect size = $(\mu_{C2} - \mu_{C3})/SD_{pooled}$
Effect size = (11.27 – 9.89)/2.86
Effect size = .48

Effect of Presentation Number on Word Identification Scores (smallest significant difference)

$SD_{pooled} = \sqrt{(SD_{P3})^2 + (SD_{P2})^2}$
$SD_{pooled} = \sqrt{(4.02)^2 + (4.09)^2}$
$SD_{pooled} = 5.73$
Effect size = $(\mu_{P3} - \mu_{P2})/SD_{pooled}$
Effect size = (8.88 – 8.61)/5.73
Effect size = .05

Effect of Presentation Number on Word Identification Scores (Presentation 1 vs. Presentation 3)

$SD_{pooled} = \sqrt{(SD_{P3})^2 + (SD_{P1})^2}$
$SD_{pooled} = \sqrt{(4.02)^2 + (3.89)^2}$
$SD_{pooled} = 5.59$
Effect size = $(\mu_{P3} - \mu_{P1})/SD_{pooled}$
Effect size = (8.88 – 8.18)/5.59
Effect size = .13

Effect of Talker Severity on Initial Consonants Correct Scores (smallest significant difference)

$SD_{pooled} = \sqrt{(SD_{C3})^2 + (SD_{C4})^2}$
$SD_{pooled} = \sqrt{(1.86)^2 + (2.01)^2}$
$SD_{pooled} = 2.74$
Effect size = $(\mu_{C3} - \mu_{C4})/SD_{pooled}$
Effect size = (16.05 –14.27)/2.74
Effect size = .65

Effect of Presentation Number on Initial Consonants Correct Scores (smallest significant difference)

$SD_{pooled} = \sqrt{(SD_{P2})^2 + (SD_{P1})^2}$
$SD_{pooled} = \sqrt{(2.36)^2 + (2.22)^2}$
$SD_{pooled} = 3.24$
Effect size = $(\mu_{P2} - \mu_{P1})/SD_{pooled}$
Effect size = (16.55 –16.04)/3.24
Effect size = .16

Effect of Presentation Number on Initial Consonants Correct Scores (Presentation 1 vs. Presentation 3)

$SD_{pooled} = \sqrt{(SD_{P3})^2 + (SD_{P1})^2}$
$SD_{pooled} = \sqrt{(2.24)^2 + (2.22)^2}$
$SD_{pooled} = 3.15$
Effect size = $(\mu_{P3} - \mu_{P1})/SD_{pooled}$
Effect size = (16.58 –16.04)/3.15
Effect size = .17

Effect of Talker Severity on Vowels Correct Scores (smallest significant difference)

$SD_{pooled} = \sqrt{(SD_{C1})^2 + (SD_{C3})^2}$
$SD_{pooled} = \sqrt{(1.44)^2 + (2.31)^2}$
$SD_{pooled} = 2.72$
Effect size = $(\mu_{C1} - \mu_{C3})/SD_{pooled}$
Effect size = (19.59 – 17.57)/2.72
Effect size = .74

Effect of Presentation Number on Vowels Correct Scores (smallest significant difference)

$SD_{pooled} = \sqrt{(SD_{P3})^2 + (SD_{P2})^2}$
$SD_{pooled} = \sqrt{(3.26)^2 + (3.24)^2}$
$SD_{pooled} = 4.60$
Effect size = $(\mu_{P3} - \mu_{P2})/SD_{pooled}$
Effect size = (16.66 – 16.37)/4.60
Effect size = .06

Effect of Presentation Number on Vowels Correct Scores (Presentation 1 vs. Presentation 3)

$SD_{pooled} = \sqrt{(SD_{P3})^2 + (SD_{P1})^2}$
$SD_{pooled} = \sqrt{(3.26)^2 + (3.47)^2}$
$SD_{pooled} = 4.76$
Effect size = $(\mu_{P3} - \mu_{P1})/SD_{pooled}$
Effect size = (16.66 – 16.19)/4.76
Effect size = .10

Effect of Talker Severity on Final Consonants Correct Scores (smallest significant difference)

$SD_{pooled} = \sqrt{(SD_{C2})^2 + (SD_{C3})^2}$
$SD_{pooled} = \sqrt{(1.47)^2 + (1.71)^2}$
$SD_{pooled} = 2.25$
Effect size = $(\mu_{C2} - \mu_{C3})/SD_{pooled}$
Effect size = (15.64 – 14.69)/2.25
Effect size = .42

Effect of Presentation Number on Final Consonants Correct Scores (smallest significant difference)

$SD_{pooled} = \sqrt{(SD_{P2})^2 + (SD_{P1})^2}$
$SD_{pooled} = \sqrt{(4.31)^2 + (4.53)^2}$
$SD_{pooled} = 6.25$
Effect size = $(\mu_{P2} - \mu_{P1})/SD_{pooled}$
Effect size = (12.99 – 12.36)/6.25
Effect size = .10

Effect of Presentation Number on Final Consonants Correct Scores (Presentation 1 vs. Presentation 3)

$SD_{pooled} = \sqrt{(SD_{P3})^2 + (SD_{P1})^2}$
$SD_{pooled} = \sqrt{(4.27)^2 + (4.53)^2}$
$SD_{pooled} = 6.23$
Effect size = $(\mu_{P3} - \mu_{P1})/SD_{pooled}$
Effect size = (13.11 – 12.36)/6.23
Effect size = .12

Effect of Talker Severity on Syllable Shapes Correct Scores (smallest significant difference)

$SD_{pooled} = \sqrt{(SD_{C1})^2 + (SD_{C3})^2}$
$SD_{pooled} = \sqrt{(1.41)^2 + (2.67)^2}$
$SD_{pooled} = 3.02$
Effect size = $(\mu_{C1} - \mu_{C3})/SD_{pooled}$
Effect size = (21.33 – 19.13)/3.02
Effect size = .73

Effect of Talker Severity on Response Times (smallest significant difference)

$SD_{pooled} = \sqrt{(SD_{C2})^2 + (SD_{C3})^2}$
$SD_{pooled} = \sqrt{(.51)^2 + (.53)^2}$
$SD_{pooled} = .74$
Effect size = $(\mu_{C2} - \mu_{C3})/SD_{pooled}$
Effect size = (2.20 – 2.03)/.74
Effect size = .23

Effect of Presentation Number on Response Times (smallest significant difference)

$SD_{pooled} = \sqrt{(SD_{P2})^2 + (SD_{P3})^2}$
$SD_{pooled} = \sqrt{(.48)^2 + (.54)^2}$
$SD_{pooled} = .72$
Effect size = $(\mu_{P2} - \mu_{P3})/SD_{pooled}$
Effect size = $(2.09 - 1.85)/.72$
Effect size = .33

Effect of Presentation Number on Response Times (Presentation 1 vs. Presentation 3)

$SD_{pooled} = \sqrt{(SD_{P1})^2 + (SD_{P3})^2}$
$SD_{pooled} = \sqrt{(.49)^2 + (.54)^2}$
$SD_{pooled} = .73$
Effect size = $(\mu_{P1} - \mu_{P3})/SD_{pooled}$
Effect size = $(2.55 - 1.85)/.73$
Effect size = .96

Effect of Talker Severity on DME Effort Ratings (smallest significant difference)

$SD_{pooled} = \sqrt{(SD_{C2})^2 + (SD_{C1})^2}$
$SD_{pooled} = \sqrt{(57.72)^2 + (30.40)^2}$
$SD_{pooled} = 65.24$
Effect size = $(\mu_{C2} - \mu_{C1})/SD_{pooled}$
Effect size = $(112.16-81.65)/65.24$
Effect size = .47

Effect of Talker Severity on Response Times (largest difference: Child 1 vs. Child 4)

$SD_{pooled} = \sqrt{(SD_{C4})^2 + (SD_{C1})^2}$
$SD_{pooled} = \sqrt{(.65)^2 + (.49)^2}$
$SD_{pooled} = .81$
Effect size = $(\mu_{C4} - \mu_{C1})/SD_{pooled}$
Effect size = $(2.45 - 1.97)/.81$
Effect size = .59

Effect of Talker Severity on Word Identification Scores (largest difference: Child 2 vs. Child 4)

$SD_{pooled} = \sqrt{(SD_{C2})^2 + (SD_{C4})^2}$
$SD_{pooled} = \sqrt{(1.79)^2 + (1.18)^2}$
$SD_{pooled} = 2.14$
Effect size = $(\mu_{C2} - \mu_{C4})/SD_{pooled}$
Effect size = $(11.27 - 2.42)/2.14$
Effect size = 4.13

Note. Raw scores are used in effect size calculations.