5-1998

ACOUSTIC AND PERCEPTUAL COMPARISONS OF IMITATIVE PROSODY IN KINDERGARTNERS WITH AND WITHOUT SPEECH DISORDERS

Robin Harwell Rodriguez
University of South Florida

Graduate School
University of South Florida
Tampa, Florida

CERTIFICATE OF APPROVAL

Master's Thesis

This is to certify that the Master's Thesis of

ROBIN HARWELL RODRIGUEZ

with a major in Speech - Language Pathology has been approved by
the Examining Committee on April 24, 1998
as satisfactory for the thesis requirement
for the Master of Science degree

Examining Committee:

____
Co-Major Professor: Ruth Huntley Bahr, Ph.D.

____
Co-Major Professor: Carolyn Ford, Ph.D.

____
Member: Winifred Strange, Ph.D.

ACOUSTIC AND PERCEPTUAL COMPARISONS OF IMITATIVE PROSODY IN KINDERGARTNERS WITH AND WITHOUT SPEECH DISORDERS

by

ROBIN HARWELL RODRIGUEZ

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
Department of Communication Sciences and Disorders
University of South Florida

May 1998

Co-Major Professor: Ruth Huntley Bahr, Ph.D.
Co-Major Professor: Carolyn Ford, Ph.D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued).

LIST OF FIGURES (Continued).

ACOUSTIC AND PERCEPTUAL COMPARISONS OF IMITATIVE PROSODY IN

KINDERGARTNERS WITH AND WITHOUT SPEECH DISORDERS by

ROBIN HARWELL RODRIGUEZ

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
Department of Communication Sciences and Disorders
University of South Florida

May 1998

Co-Major Professor: Ruth Huntley Bahr, Ph.D.
Co-Major Professor: Carolyn Ford, Ph.D.

This study investigated the affiliation of prosody with childhood articulation disorders. The Tennessee Test of Rhythm and Intonation Patterns, T-TRIP (Koike & Asp, 1981), was used to determine if kindergartners with linguistic (i.e. phonological) speech disorders, oral-motor speech disorders, or normal speech performed differently on imitative prosody tasks. Performance was assessed perceptually with T-TRIP overall and subtest scores, and acoustically with measurements of individual prosodic variables (amplitude, duration, and fundamental frequency) on selected items from the rhythm and intonation subtests. Perceptual and acoustic data were examined for characteristic patterns of performance by individual subjects and by groups.

A Kruskal-Wallis ANOVA of the perceptual scores revealed that the three groups performed differently on the T-TRIP rhythm, and intonation subtests, and on the total score. Specifically, the oral-motor group had the lowest range of scores and was clearly separated from the other two groups. No group cut-off scores were established since the linguistic group's scores slightly overlapped the control groups' range of scores.

Acoustic results generally supported the findings of earlier studies of stress and intonation. Correct responses contained a wide selection of acoustic patterns, while incorrect responses consisted of error patterns resembling those of younger children. Subjects with speech disorders demonstrated several characteristic error patterns: linguistic subjects tended to add syllables and to lexicalize items, while oral-motor subjects tended to delete syllables and to convert iambic stress into trochaic.

Overall, whether T-TRIP responses were examined by perceptual or acoustic methods, the oral-motor group's imitative prosody ability was significantly different than the other groups' performance. The clinical implications of this finding are that the

T-TRIP has the potential to be used as a screening tool to identify subjects whose difficulties with imitative prosody are consistent with oral-motor speech disorders, specifically DVD.

Abstract Approved: ____

Co-Major Professor: Ruth Bahr, Ph.D.
Assistant Professor, Department of Communication Sciences and Disorders

Date Approved: ____________________________

Abstract Approved: ____

Co-Major Professor: Carolyn Ford, Ph.D.
Clinical Instructor, Department of Communication Sciences and Disorders

Date Approved: ____________________________

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Overview

Prosody is an important component of language and speech. Considered from a linguistic point of view, prosody functions to signal many communicative distinctions including lexical, grammatical, and emotional attributes of speech (Crystal, 1969). Viewed from an articulatory standpoint, it functions to segment the continuous stream of speech sounds into decipherable units. Abnormal prosody, or dysprosody, may negatively influence the perceived intelligibility of spoken messages (Wingfield, Lombardi, & Sokol, 1984). A common example of this phenomenon is the reduced intelligibility of non-native speakers of English who retain the prosody of their native language. Prosodic disturbances have been associated with childhood articulation disorders, although few detailed studies exist to support this association. Additionally, most existing studies use subjective judgements of perceived prosody, rather than more objective acoustic measures. Recently, it has been proposed that a prosodic disturbance may provide a diagnostic marker for the difficult-to-diagnose childhood articulation disorder of developmental verbal dyspraxia (Shriberg et al., 1997a). The affiliation of prosody with childhood articulation disorders merits further investigation, and is the focus of the present study.

Definition of Prosody

Crary and Tallman (1993) define *prosody* as the "suprasegmental features of fundamental frequency, duration and intensity that contribute to...the melody of speech production (p. 245)." Linguists agree that these three features are the main components of prosody in English (Crystal, 1969; Lehiste, 1970; Lieberman, 1966). However, the terms used may differ, depending upon whether they are being considered from the speaker's point of view (physiological/production), the listener's point of view (perception), or as an acoustic manifestation (measurement). Table 1 was adapted from Lehiste (1970) to demonstrate this difference in terminology.

Table 1. Terminology Describing Suprasegmental Features according to Reference Point-of-view and Linguistic Function. (Adapted from Table 1.1, Lehiste 1970, p. 4)

	Reference	Point of	View:	Linguistic	Function:
Suprasegmental	*Speaker's/ Physiological/ Production*	*Acoustic Manifestation/ Measurement*	*Listener's/ Perception*	*Word Level*	*Sentence Level*
Quantity Features	Timing of articulatory sequences	Time dimension of the acoustic signal	Perception of duration	Quantity	Tempo
Tonal Features	Phonation	Fundamental frequency	Perception of pitch	Tone	Intonation
Stress Features	Effort level of manipulation of physiological mechanism	Intensity and amplitude	Perception of loudness and of stress	Word stress	Sentence-level stress

Segmental vs. Suprasegmental

Since these features (duration, fundamental frequency, and intensity) are also present on a segmental level, how does the suprasegmental use of them differ? According to Lehiste (1970), the segmental use of these features is inherent, while their suprasegmental use is a secondary, overlaid function. The phonetic features of voicing, duration, and intensity are already present in every segment; manipulation of these features creates the suprasegmental features of intonation, tempo, and stress. For example, "the fundamental frequency of a voiced segment may serve simultaneously to identify the segment as voiced, and to constitute part of the manifestation of a tonal or intonational pattern" (Lehiste, 1970, p. 2). Also, suprasegmental features are identified by comparing two or more adjacent segments. Conversely, the identification of a segmental feature only involves inspection of the phonetic segment in which it occurs. For example, a vowel can be identified as rounded or unrounded without reference to adjacent sounds, but it cannot be identified as stressed or unstressed (or high or low in pitch, or long or short in duration) without comparison to other sounds (Lehiste, 1970). Once suprasegmental features have been identified by the listener, what can they contribute to the speaker's message?

Functions of Prosody

Suprasegmental features communicate linguistic and non-linguistic information from speaker to listener. Linguistic *stress* differentiates noun-verb pairs and emphasizes new or important information within an utterance. *Intonation* functions syntactically to signal questions or statements. Pragmatically, intonation may also help a speaker hold the conversational floor, by signaling "unfinished business" with a high, terminal fundamental frequency (Lehiste, 1970). Factors such as a speaker's dialect and emotional mood affect their *tempo* or rate of speech, and therefore the duration of all segments. Klatt (1976) reports that some speakers slow their rate of speech as a technique to draw attention to important phrases, and that a word preceding a phrase boundary will be lengthened. However, "it is not known whether a speaker learns to lengthen segments at the ends of phrase boundaries in order to help the listener decode the message, or if there is a natural tendency to slow down at the ends of all motor sequences or planning units" (Klatt, 1976, p. 1212). Research continues into the linguistic motivations underlying suprasegmental features in speech, but this is not the focus of the present study. Instead, this study focuses on the ability (or inability) of speakers to volitionally produce discernible suprasegmental variations in the absence of linguistic content. Two suprasegmental features form the nucleus of this study: stress and intonation.

Stress

While the linguistic functions of stress are well-known, the intricacies of the differentiation of stressed syllables from adjacent syllables are more difficult to explain. Crystal (1969) defines *stress* as "those variations in linguistically contrastive prominence primarily due to loudness." Stressed syllables are spoken and heard with more prominence than adjacent syllables. However, the perception of stress is not exclusively due to loudness, it may also be due to increases in duration or in fundamental frequency (f0) or to a combination of elements. Klatt (1976) reports that a 10% to 20% increase in duration is one acoustic correlate of emphatic or contrastive stress. Lehiste (1970, p. 125) states that "the perception of stressedness appears to be based on a number of factors, the most influential of which is fundamental frequency." If the concept of stress is divided into production and perception aspects, the issue may become less confusing.

Production of Stress

According to Lehiste (1970), the production of increases or decreases in stress are ultimately due to increases or decreases in the speaker's physical effort. "A stressed syllable is one that the speaker consciously utters with greater effort than neighboring syllables - the listener hears this syllable as louder than unstressed syllables" (Lehiste, 1970, p. 106). Greater effort results from an increase in activity of the speaker's respiratory muscles, which causes an increase in subglottal pressure, and in turn causes greater amplitude (and usually greater frequency) of vocal fold vibration. However, partly because of this relationship between vocal fold amplitude and frequency of vibration, the perception of stress may involve more than increases in amplitude.

Perception of Stress

The relative importance of intensity, f0, and duration to the perception of stress has been studied with no conclusion. Duration is reported to be the central cue for perception of stress by some researchers (i.e., Klatt, 1976), while others report than listeners relied on f0 (Crystal, 1969), or on intensity (Lehiste, 1970). In some cases, stress may be signaled with a segmental cue, such as vowel reduction. For instance, "in English there is a tendency for most vowels in weakly stressed syllables to approach schwa in quality" (Lehiste, 1970). Adding to the confusion over which of these cues signal stressed syllables, studies of adult speech have revealed inconsistency in how stress is indicated both between and within speakers (Lehiste, 1970). Given such inter- and intra-speaker variability, studies of suprasegmental stress should incorporate measurement of all three possible parameters - intensity, f0, and duration - as well as listener evaluation of perceived stress into their design.

Two recent studies of the acoustic correlates of children's stress included these elements of acoustic measurement and listener evaluation. Pollack, Brammer, and Hageman (1993) investigated the control of prosody in children aged 2-, 3- and 4-years- old. Their results indicated that children in the two oldest age groups (aged 36 to 38 months, and 48 to 50 months) combined increases in intensity, duration and f0 to indicate stressed syllables, but exhibited some variability in the degree of increase based on the location of the syllable within the word. For example, when compared across words, unstressed initial syllables generally had greater peak amplitude than unstressed final syllables. However, Pollack et al. attributed most of the variability in their results to the youngest children (age 24 to 26 months). This group did not consistently reduce the duration of unstressed syllables, and at times used contradictory patterns of f0 and intensity change to indicate stressed syllables, i.e. an increase in intensity, combined with a decrease in f0. Perceptual correlation studies revealed that these "contradictory" tokens were most often misidentified by listeners.

These results were questioned by Kehoe, Stoel-Gammon, and Buder (1995), who found that children as young as 18 months used all three acoustic parameters to differentiate stressed syllables, and that for the majority of responses, the contrast between stressed and unstressed syllables was perceptible. For the words which were perceived to have "unreliable" stress, incorrect stress, or level stress, the acoustic differences between stressed and unstressed syllables were much less than those perceived as correctly stressed (Kehoe et al., p. 346). Methodological differences between the two studies may explain the different results. Pollack elicited trochaic ('CVCV) and iambic (CV'CV) disyllabic novel words, i.e. *'boda* and *bo'da*. However, Kehoe et al. elicited trochaic words which were familiar to the children, contrasted with monosyllabic words with similar segmental content, i.e. *monkey* and *key*. Pollack et al.'s stimuli were more controlled experimentally, while Kehoe et al.'s were more natural. It has been suggested that "children's control over phonetic parameters may emerge sooner in natural speech than in imitated forms in an experimental task" (Kehoe et al., p. 38). Imitation tasks would provide a rigorous test of children's volitional production of stress, and of the other main prosodic aspect of speech - intonation.

Intonation

As was shown in Table 1, the terms *phonation, fundamental frequency, pitch, tone* and *intonation* describe related concepts. Fundamental frequency, or f0, is a quantification of the rate of the speaker's vocal fold vibration during phonation. Although there is not a strict one-to-one relationship, the physiological phenomenon of f0 correlates to the psychological attribute of sound known as pitch. Linguists, such as Lehiste, refer to word-level contrasts in f0 as "tone", and reserve the term "intonation" to describe "the linguistically significant function of fundamental frequency at the sentence level" (Lehiste, 1970). Some languages use word-level contrasts in f0 to make semantic or phonological distinctions; however, English does not. Therefore, the term intonation will be used here to describe f0 changes at the sentence-, word-, or syllable-level. But how are these changes in f0 produced?

Production of F0

Fundamental frequency is determined by physiological factors. A speaker's age, gender, and physique affect the size of their larynx and thus the rate of vibration of their vocal folds. As with the other articulators, vocal folds with more mass have more inertia, resulting in a slower vibration and a lower f0. Therefore, differences in inherent vocal-fold mass explain the characteristic f0 ranges for infants, children, adult and female speakers, as well as the between-speaker differences in f0.

Within-speaker variations in f0 (i.e. intonation) are generally the result of increasing or decreasing vocal fold tension. A speaker may manipulate vocal fold tension by adjusting (contracting or relaxing) the intrinsic laryngeal muscles, or by adjusting (increasing or decreasing) the subglottal air pressure during expiration. Lieberman (1966) asserts that the f0 of phonation is primarily a function of subglottal air pressure. According to Lieberman, the subglottal respiratory muscles relax at the end of a breath group, causing a drop in both the subglottal air pressure and the f0 during the last 150 to 200 ms of phonation. This explains the phenomenon of “declination”, or utterance-final decrease in f0 (and intensity) that is typical of English and most other languages. However, at times an utterance-final increase in f0 occurs, signaling a change in the linguistic function of the sentence. In this case, the laryngeal muscles contract, increasing the tension on the vocal-folds and countering the effects of the drop in subglottal air pressure (Lieberman, 1966). The redundancy of these mechanisms for controlling f0 enable a speaker to manipulate the intonation of an utterance for linguistic purposes.

Perception of Pitch

Because of the design of the human auditory system, the perception of pitch is not uniform with regard to frequency change, nor is it independent of the duration and intensity of the signal. Pitch “increases less and less rapidly as the stimulus frequency is increased linearly, and more and more rapidly as the stimulus frequency is increased logarithmically” (Lehiste, 1970, p. 65). In other words, a larger frequency difference is needed between two high frequency tones than between two low frequency tones, in order for a listener to perceive a pitch difference between the tones. For example, although there is the same linear difference of 100 Hz between 100 Hz and 200 Hz, and between 11,100 Hz and 11,200 Hz, the perceived increase between each pair of pitches is not the same. Listeners may perceive 200 Hz as "twice as high" as 100 Hz, yet fail to perceive any difference between 11,100 Hz and 11,200 Hz.

The perception of pitch is also related to the perceived intensity (loudness) and duration of acoustic events. According to Lehiste, at least 12 ms, or more than three cycles of a soundwave within the speech range (250 Hz), is needed "to give the listeners a firm feeling of pitch" (1970, p. 67). This is important for studies using tones as stimuli, but should not be a factor with speech stimuli, since connected speech necessarily contains the required three cycles. A listener's sensitivity for pitch changes also increases or decreases according to the loudness level of the stimulus. For higher tones the effect is profound, however, for tones in the lower range of audible frequencies, such as speech, the effect is slight (Lehiste, 1970). Therefore, if speech stimuli are presented at a comfortable listening level, listeners should be able to judge intonation contours independently of duration and intensity. But perceptual judgements are not the only method for measuring suprasegmentals.

Measurement of Suprasegmental Features

Analysis of duration, intensity and f0 may be made via physiological measurements, perceptual judgements, or acoustic signal measurements. Physiological measurement methods such as X-ray and electromyography are not currently practical due to their expense, as well as the accompanying discomfort and danger of the procedures.

Perceptual judgements are subjective and therefore prone to greater levels of interjudge variability, than judgements made with a more-objective method such as acoustic analysis. Kent and Rosenbek (1983) list two primary advantages of the acoustic method over perceptual description. They state that acoustic analysis yields "a greater degree of quantification as well as a greater degree of sensitivity to some aspects of deviant speech production." However, the sensitivity afforded by acoustic measurement may exceed that of the human listener. Since listeners ultimately decide the success of a speaker's attempts to communicate, differences in acoustic measurement must be evaluated according to the capabilities of the human auditory system.

Perceptible Differences of Duration, f0, and Intensity

Before acoustically-measured differences in duration, f0, or amplitude can be evaluated as perceptually-salient to a human listener, the smallest perceptible difference must be known. For example, the durations of two vowels may be measured to an accuracy of fractions of milliseconds. However, for speech sounds - which usually range in duration from 30 to 300 ms - studies have shown that differences must be 10 to 40 ms in length before they are judged by listeners to have a "just-noticeable difference" in duration (Lehiste, 1970, p. 13). For overall amplitude, the just-noticeable difference, or JND, is approximately 1 dB (Lehiste, 1970). However, since the perception of pitch is not uniform with regard to frequency change and is also dependent upon the duration and intensity of the signal, determining the threshold of perceptible change of f0 is more complex. The frequency response of the human auditory system is essentially linear below 500 Hz, but logarithmic above 500 Hz (Lehiste, 1970). Therefore, f0 values in Hertz (a linear scale) must be translated into a scale which more closely matches this logarithmic design, such as the semitone scale which is used in studies of music perception. In speech intonation research, both semitones and Hz have been used to express pitch, with continuing debate over which scale is better. One semitone represents a perceptible difference in musical frequency. In this musical scale, equal distances represent equal frequency proportions which amounts to using a logarithmic frequency scale" (Hermes and van Gestel, 1991). Because of this logarithmic organization, the semitone scale was selected as more perceptually appropriate for the current study.

Prosody in Meaningful Speech vs. Nonsense Syllables

Because segmental and suprasegmental features co-exist, speech samples for prosody measurements must be selected to maximize control of all factors which might be operating. One method for minimizing the effects of linguistic influences is to use nonsense syllables, although of course this will not eliminate effects due to syllable position, i.e. utterance-final lengthening or declination. Nor will it negate the inherent variations between phonetically distinct segments, such as duration and f0 differences.

Segmentally, Lehiste (1970) states that vowels have an "intrinsic pitch", or a connection between vowel quality and the relative height of the average f0 associated with it, based on the configuration of the vocal tract. "Other factors being kept constant, higher vowels have a higher fundamental frequency" (Lehiste, 1970, p. 68). This phenomenon supports the use of reduplicated syllables for measurement of suprasegmentals such as intonation, since the intrinsic pitch will thus be kept constant, and changes in f0 can then be attributed to laryngeal management by the speaker. The use of reduplicated syllables will also facilitate the comparison of syllable durations by eliminating the inherent duration and f0 differences found in different phonetic sequences.

But are nonsense syllable studies valid reflections of spontaneous speaking habits? Several studies have addressed this question by incorporating both meaningful and nonsense stimuli.

When intonation contours were imitated in both meaningful and non-meaningful (reduplicated) sentences, the results suggested that “the meaningful/non-meaningful distinction produced no significant differences for any variable” (Crary & Tallman, 1993, p. 256). According to Klatt, “comparisons between measurements of ... nonsense syllable sequences, ... read discourse, and ... spontaneous speech suggest that the similarities are greater than the differences” (1976, p. 1209). Further, when rules generated from studies of spontaneous speech were applied to nonsense syllables in order to predict vowel durations, the rules accounted for 97% of the total variance for each vowel (Klatt, 1976, p. 1216). These findings imply that the use of reduplicated nonsense syllables is an appropriate method to control for linguistic and segmental confounds, since it allows investigation of the production of prosody on a phonetic level.

Prosody in Children’s Speech

Before applying adult-based prosody research to the speech of children, several questions must be addressed. First, when does the use of prosodic features develop in children’s speech? Second, do children use the acoustic features of intensity, f0, and duration in the same manner as adults when imitating rhythmic and intonation patterns? The answers to both questions are related to the age of the children studied. Lieberman (1966) states that “intonation, stress, and prosody are primarily responses to the periodicity, amplitude, spectral character, and duration of the output of the larynx.” Naturally, as children mature and gain increasing control of their laryngeal output, they will gain increasing control of both segmental and suprasegmental aspects of their speech.

Suprasegmental components of speech, including prosody, appear early developmentally. Crystal (1979) identifies five stages of prosodic development. Stage I begins at birth when infant vocalizations appear as the "prelinguistic antecedents of prosodic features." Prosodic features become progressively more linguistic in use, concluding with the emergence of Stage V at approximately eighteen months of age. This final stage is characterized by pairs of identical two-word utterances whose linguistic functions are distinguished by stress and intonation (Crystal, 1979, p. 45). During the intervening three stages, linguistic intent appears in segmental as well as suprasegmental dimensions of speech. According to Crystal (1979), the prosodic dimension stabilizes earlier and is more readily elicited than the segmental dimension. The early development of prosody is confirmed by other researchers. Lenneberg (1967) states that,

> the first feature of natural language to be observed in a child's babbling is the contour of intonation. The sound sequences that are produced are not meaningful and they don't have definable phoneme structure, but they have recognizable intonation such as occurs in questions, exclamations, or affirmations, (p. 279).

Is this early-developing prosody meaningful or simply imitative? Menyuk and Bernholtz (1969) used a single-subject design to explore whether a 20-month-old child used prosody meaningfully or only imitatively. They concluded that their subject demonstrated the ability to change intonation and stress patterns in one-word utterances generatively to create sentence types, rather than simply imitating prosodic features. Substantial prosodic development, therefore, occurs during the first two years of life, with imitation preceding volitional production. Thereafter, production of suprasegmental features (and segmental features) continues to stabilize. For example, unstressed syllable duration is one suprasegmental aspect which appears to stabilize over time. Allen and Hawkins (1980) observed that children have difficulty reducing the length or vowel quality of unstressed syllables, particularly in word-initial position. They proposed that English children have a “bias in learning trochaic disyllables” and therefore often deleted initial unstressed syllables, resulting in trochaic rather than iambic stress patterns (Allen & Hawkins, 1980, p. 240). Learning to shorten unstressed syllables, according to these researchers, is part of learning to control stress in English. By kindergarten, i.e. at approximately four to six years of age, children should be able to use prosody in spontaneous speech as well as imitatively during experimental tasks. But do children use prosody in the same way as adults do? A review of recent studies of prosodic aspects of children’s speech may help to answer that question.

Studies of Prosody in Children with Normal Speech

Table 2 contains summaries of ten studies of children’s prosody. Comparisons between the studies are made difficult by the fact that methodology between studies differed, especially in the age and number of subjects studied and in the manner of eliciting and analyzing the subjects’ speech.

Two studies used single subject designs (Menyuk & Bernholtz, 1969; Scukanec & Watson, 1995). These studies were also the only studies to examine the subjects' production of suprasegmentals over time. However, while Scukanec and Watson measured their subject at two-month intervals for a period of 26 months (thus yielding a longitudinal study designed to examine developmental changes), Menyuk and Bernholtz simply recorded the subjects over a two month period, perhaps to ensure that the child's productions were stable. The remaining eight studies used multiple subjects (n=4 to n=30) in order to examine intra-subject variability, and seven of these studies grouped the children by age in order to examine between-group differences for trends in development.

Most studies only focused on a single aspect of prosody. *Stress* was the aspect studied most often (included in nine of ten studies). *Intonation* was included in only four of the studies, with Loeb and Allen (1993) focusing solely on intonation. Two studies ostensibly examined both stress and intonation variables (Schwartz, Petinou, Goffman, Lazowski, & Cartusciello, 1996; Menyuk & Bernholtz, 1969). Instead, these two studies effectively examined the relationship of these variables to perceived differences in communicative intent and concluded only that such differences existed acoustically and could be objectively measured, in addition to being subjectively apparent from the context of the utterance. Only one study (Koike & Asp, 1981) examined multiple aspects of prosody: stress, intonation, and tempo.

Speech samples were gathered by various methods. Only one study recorded spontaneous speech (Menyuk & Bernholtz , 1969). Elicitation of target words through conversational repairs or play was the method of choice; seven of ten studies used this method. The remaining two studies (Koike & Asp, 1981, Loeb and Allen, 1993) used direct imitation of an adult model. For these two studies, the subjects' imitative responses were judged as correct or incorrect in comparison to the model.

The type of dependent variables in each study differed. Two used only perceptual judgements, while four used only acoustic measurements, and four used both perceptual and acoustic data. Perceptual judgements included the transcription of perceived syllable- and word-stress, the interpretation of the subject's communicative intent, and quality rating of imitative responses as correct, incorrect, or partially correct. Acoustic data included f0, duration and amplitude values for words, syllables, vowels or utterances. Despite the many methodological differences, the results of these studies add specific knowledge about how children use the prosodic features of stress and intonation.

As mentioned above, Kehoe, Stoel-Gammon, and Buder (1995) found that children aged 18-30 months were manipulating f0, intensity and duration to produce stress contrasts, although they were not completely adult-like in their productions. The adults in the Kehoe et al. study used intensity and duration increases more often than f0 increases to differentiate stressed syllables. Conversely, the children used f0 increases more often, perhaps because of poor control of duration contrasts. Nevertheless, Kehoe et al. found that all subjects used greater intensity in stressed syllables, and that there was a tendency for the intensity contrast to increase with age. Likewise, Pollock, Brammer, and Hageman (1993) concluded that 2-, 3-, and 4-year-old children all used f0, intensity and duration to indicate word stress, although the youngest children did not consistently reduce the duration of unstressed syllables. Hence, the Pollack et al. and Kehoe et al. findings suggest that intensity and f0 may provide the main cues which differentiate stressed syllables in the speech of 2-, 3- and 4-year-old children, due to their poor control of duration.

Other acoustic studies of children's speech confirmed these conclusions regarding the inconsistent use of acoustic features by 2-year-old children, but disagreed as to which cues were inconsistent. Schwartz et al.(1996) examined acoustic stress patterns in the speech of fourteen 22- to 28-month-old children. They found that stressed syllables had significantly higher peak amplitude and peak f0 values, and longer durations (syllable duration and vowel duration) than unstressed syllables. Contrary to Kehoe et al.'s (1995) findings, Schwartz et al. found the adult's vowel and syllable durations were actually *longer* than the children's. These findings were explained by the fact that the adult productions were "child-directed", i.e. the adult provided an over-exaggerated model for the novel words. Overall, Schwartz et al. concluded that their subjects used all three acoustic parameters to indicate stress, but that the magnitude of their contrasts was less than that of the adult model. Again, the use of "child-directed" speech for the adult measurements probably resulted in over-exaggerated contrasts not necessarily to be expected in adult productions of "adult-directed" speech.

Konefal and Fokes (1985) elicited conversational repairs for the acoustic measurement of stress contrasts in three groups of children, aged 2- to 6-years-old. They found that the oldest children were more likely to combine acoustic cues to indicate stress, increasing at least two parameters 90% of the time. None of the oldest children *decreased* acoustic parameters during the repaired word, while nine out of twenty (45%) of the younger subjects (i.e. 2- and 3- year-olds) decreased one or more acoustic parameters over 60% of the time during repaired words. As Konefal and Fokes point out in their discussion, this may be an alternate strategy to draw attention to the repaired word, or it may simply indicate uncertainty regarding the linguistic-correctness of the repair. Decreases in acoustic parameters were also noted in a study of communicative intent (Furrow, Podrouzek, & Moore, 1990) which concluded that *intensity* was a reliable cue for differentiating assertive versus directive intent, while f0 was at times increased, at times decreased, and at times kept constant by four 23-month-old children.

Other studies merely concluded that the subjects were able to convey linguistic stress or intonation to their listeners, without concluding which acoustic features were used to do so (Menyuk & Bernholtz, 1969; Hornby & Hass, 1970; Scukanec & Watson, 1995). Two studies focused on the imitation of sub-groups of prosodic features. Koike and Asp (1981) compared the abilities of 3- and 5-year-old children to imitate stress, tempo and intonation patterns of nonsense syllables, while Loeb and Allen (1993) compared their abilities to imitate intonation contours of meaningful sentences. Both found that five-year-old children correctly imitated the adult model more often than three-year-old children did. Also, both studies revealed that 3-year-old subjects had more difficulty with rising than with falling intonation contours. These findings may indicate that rising terminal-f0 is mastered late developmentally, as was suggested regarding the reduction of unstressed syllable duration (Allen & Hawkins, 1980).

Overall, despite the intra- and inter-study variations regarding children's use of prosodic aspects of speech, these results indicate that prosodic features are found in the speech of very young children, and that their use and control of them increases with age. Since normal young children's production of prosodic features may not be adult-like, disordered children's their prosodic abilities should not be compared with adult models, but instead should be compared to the abilities of children within the same chronological age group. Assessment of children's prosodic abilities may be useful for several reasons, particularly in children with speech disorders.

Studies of Prosody in Children with Disordered Speech

Prosodic disturbances have been associated with numerous childhood speech disorders. Disordered prosody has been reported in the speech of children with disorders such as developmental verbal dyspraxia, or DVD, (Rosenbek & Wertz, 1972; Edwards, 1973) and childhood dysarthria (Love, 1992). Dysprosody has also been identified in the speech of children with hearing impairments (Most & Yael, 1994; Murphy, McGarr, & Bell-Berti, 1990), dysfluent speech (Prins, Hubbard, & Krause, 1991), and autism (Baltaxe & Simmons, 1985).

Increasingly, prosodic disturbances are being targeted as part of speech therapy in order to improve intelligibility (Hargrove & McGarr, 1994). Abnormal prosody, or *dysprosody*, may negatively influence the perceived intelligibility of spoken messages (Wingfield, et al., 1984). Lack of prosodic variation in speech, or *aprosody*, can cause the listener to lose interest in the speaker's words. Intact prosody, on the other hand, has been correlated with increased perceived intelligibility, increased accuracy of articulation for stressed syllables, and retained meaning of sentences in the presence of degraded segments (Wingfield et al., 1984). These associations suggest that targeting prosodic disturbances in therapy may improve intelligibility. To this end, Hargrove, Roetzel, and Hoodin (1989) had favorable results from a single-subject experiment designed to modify the prosody of a six-year-old language-impaired child, indicating that prosodic skills are modifiable.

Before efforts to remediate prosodic disturbances in the speech of unintelligible children can be effective, specific information is needed about how their prosody is atypical. To date there have been few studies aimed at obtaining this information, and fewer diagnostic tools developed to incorporate this information into clinical assessment. Table 3 summarizes ten recent studies of prosodic aspects of speech in children with disordered speech. As in prosodic studies with normally developing children, methodology varied greatly among the studies.

First the number of subjects per study varied, with between seven and 87 speech-disordered subjects, and between zero and 60 normally-developing controls included in each study. Second the studies of disordered speech reviewed here included older subjects with a larger age range (between three and 19 years old), than the studies of normally-developing children reviewed earlier. Third, two groups of studies placed subjects into two age-groups in order to make developmental comparisons between the groups. Studies contrasting hearing-impaired and normally-hearing subjects (Most & Yael, 1994; Murphy et al., 1990), and those contrasting speech disorders with speech delays (Henry, 1990; Shriberg et al., 1997b; Shriberg et al., 1997c), all featured developmental comparisons between groups.

The aspect (or aspects) of prosody studied also varied. Five studies focused exclusively on *stress*, and two focused exclusively on *intonation*. Only three studies included multiple aspects of prosody; one included both *stress* and *tempo* (Henry, 1990), one included *stress* and *intonation* (Baltaxe & Simmons, 1985), while the other

included *stress, tempo,* and *intonation* (Shadden, Asp, Tonkovich, & Mason, 1980). The latter two studies both used all or part of the Tennessee Test of Rhythm and Intonation Patterns, or T-TRIP (Koike & Asp, 1981), to obtain speech samples for analysis.

The T-TRIP samples imitative speech in the context of reduplicated nonsense syllables presented by a tape-recorded model. Imitation of tape recorded models was also used by Most and Yael (1994) and Crary and Tallman (1993) to obtain speech samples, although those studies included meaningful as well as nonsense stimuli. The three studies with hearing-impaired subjects (Most & Yael; Murphy et al., 1990) or dysfluent subjects (Prins et al., 1991) permitted oral-reading of meaningful stimuli. Only Baltaxe and Simmons (1985) elicited conversational repairs and only the studies by Shriberg et al., (1997b; 1997c) elicited conversational speech for analysis, although Most and Yael elicited a picture description, which they termed "conversational speech." The design and measurement of dependent variables differed. Six studies employed only perceptual judgements, while two employed only acoustic measurement. Acoustic data included the expected measurements of f0, duration, and amplitude of vowels, syllables, words, or sentences, as well as values for f0 range and intersyllabic pause duration. Perceptual judgements included the transcription of syllable- and word-stress, and the location of stuttering events, along with judgements regarding the correctness or incorrectness of imitations of T-TRIP stimuli. Additionally, Shriberg et al. (1997b; 1997c) used trained research transcriptionists to assess suprasegmental variables with the Prosody Voice Screening Profile, PVSP, (Shriberg, Kwiatowski, & Rasmussen, 1990). The T-TRIP and PVSP will be reviewed in greater detail later. However, neither the T-TRIP or PVSP findings in the studies examined here, were supported with any acoustic data. Both acoustic and perceptual data were reported by Baltaxe and Simmons (1985) and Murphy et al. (1990).

The studies covered a broad range of speech and language disorders - aphasia, autism, stuttering, hearing impairments, unspecified articulation inadequacies, phonological speech delay, and developmental verbal dyspraxia — which made comparing their results difficult. Also, several studies used different terminology to describe apparently identical disorders, while some identically-labeled groups turned out to have differing inclusion criteria. After excluding groups with speech disorders attributed to hearing impairments, stuttering, and autism, the remaining six groups comprise overtly-similar articulation-disordered children. The terminology and criteria used to classify these six groups are listed in Table 4. Shriberg et al.'s two groups are distinguished from each other, but not necessarily from the other four groups, by the Speech Disorder Classification System, or SDCS, developed by Shriberg (1993). The other four subject groups in Table 4 have different labels but may or may not represent the same population. The exclusion criteria for all groups were similar, generally including only subjects with normal hearing, receptive language and non-verbal intelligence, and with no known neurological or oral-structural abnormalities. Inclusion criteria were also similar for the first four studies, but with important factors omitted from some. For example, neither Baltaxe and Simmons (1985), nor Shadden et al. (1980) included assessments of their subjects' motor-speech performance. Also, although all groups were stated to have articulation problems, the relative severity of their impairments was difficult to judge due to the use of different assessment techniques. Accordingly, the studies' subjects may not be comparable (based on the nature and severity of their articulation disorders), and so comparisons among them are limited. Despite these limitations, the studies' add valuable knowledge about the prosody of children with speech disorders.

Table 4. Terminology, Inclusion, and Exclusion Criteria for Disordered Speech Groups

Study	**Term(s)**	**Study Exclusion Criteria**	**Group Inclusion Criteria**
Baltaxe & Simmons (1985)	"aphasic"	Utterances outside of target range: 2 to 4 MLU	"Severe oral-language handicap, but non-verbal IQ in the normal range"
Crary & Tallman (1993)	"severe speech disorder"	Hearing WNL, normal language comprehension abilities, no known or obvious neurologic deficit, no known or obvious voice disorder.	"Poor speech articulation abilities characterized by multiple-articulation errors, and poor motor-speech performance as judged by diadochokinetic (DDK) rates"
Shadden et al.(1980)	"inadequate articulation"	Hearing WNL, no neurologic or voice disorders, receptive NSST† scores at or above 25th %-ile	"Inadequate articulation scores as judged by Templin-Darley ††"
Henry (1990)	"speech disorders"	Hearing, language comprehension, and intelligence WNL, no structural abnormality of speech mechanism.	"severe articulation problems, multiple articulation errors on previous testing, plus unintelligible connected speech to anyone outside their family circle"
Shriberg et al. (1997b; 1997c)	"suspected DAS" *	Hearing WNL at time of assessment, IQ normal (by report or standard scores), dysarthria ruled-out.	Speech disorder categorized (i.e., as DAS) using the SDCS ‡
Shriberg et al. (1997b; 1997c)	"speech delay (SD) or residual errors (RE)" **	Hearing WNL at time of assessment, IQ normal (by report or standard scores), dysarthria ruled-out.	Speech disorder categorized (i.e., as SD or RE) using the SDCS ‡

* DAS = developmental apraxia of speech. ** speech delay if C.A. < 7, residual errors if C.A. > 7. †= Northwestern Syntax Screening Test, NSST (Lee, 1971) †† = Templin-Darley Test of Articulation (Templin & Darley, 1960; Templin, 1953)
‡ = Speech Disorder Classification System, SDCS (Shriberg, (1993).

Similar results were attained in all three studies by Shriberg et al. (1997b; 1997c) which used the PVSP perceptual assessment. Four groups participated in the Shriberg et al. (1997b, 1997c) studies: a younger group (less than 7-years-old) and an older group (7-years-old or older) with suspected developmental apraxia of speech (DAS), and younger and older groups diagnosed with either speech delay (if younger than seven) or residual error (if seven or older) according to the SDCS (Shriberg, 1993). The terms "speech delay" (SD) and "residual error" (RE) describe the same type of articulation disorder that the term phonological disorder does. Of the seven suprasegmental aspects of speech assessed, *stress (lexical/phrasal)* was judged to be appropriate more often in groups diagnosed with SD/RE articulation disorders, and inappropriate or questionable more often in the younger group of subjects with suspected DAS. Inappropriate stress was further described as "excess, equal, or misplaced stress" according to the PVSP (Shriberg et al., 1990). Findings regarding other suprasegmental aspects of prosody did not match the significance of findings regarding *stress*. All three studies by Shriberg et al. (1997b; 1997c), therefore, support the hypothesis that inappropriate stress may serve as a marker to distinguish at least one subtype of developmental apraxia of speech. However, since acoustic measurements were not used in the series, Shriberg et al. do not provide any specific acoustic information regarding *how* the stress was inappropriate.

Other studies employing acoustic measurement may help build hypotheses about *how* the variables of f0, duration and amplitude will differ in disordered populations. For hearing-impaired subjects, Most and Yael (1994) found that rising intonation contours were often produced incorrectly by subjects in both age groups (five-six years old and nine-12 years old). This judgement depended on acoustic, rather than perceptual, criteria, i.e., a change in f0 of at least 10 Hz in the appropriate direction. In a study of intra-word stress contrasts produced by 14 hearing impaired subjects, Murphy et al. (1990) utilized acoustic measurement with only four subjects. The responses judged to have incorrect stress placement were generally also identified as having less contrast between syllables, i.e., an inadequate degree of contrast. Normally-hearing aphasic subjects exhibited restricted f0 range with less terminal f0 fall (declination) in a study by Baltaxe and Simmons (1985). Crary and Tallman's (1993) speech disordered subjects demonstrated similar restricted f0 ranges and terminal fall. In addition, they produced longer overall response durations than the control subjects. On the basis of these results, it may be hypothesized that the prosody of disordered populations may contain a narrower range, less acoustic contrast between stressed and unstressed syllables, and less proficiency with rising intonation contours than the prosody of children with normally developing speech.

Studies without acoustic measurements may also suggest possible differences in performance when different groups are assessed with the same test. Both Shadden et al. (1980) and Henry (1990) used the T -TRIP to compare the performance of children with and without articulation disorders. However, Shadden et al.'s results were not significant, while Henry's were. Henry attributes this discrepancy to the greater severity of her subjects' disordered speech, and supports this assertion by pointing out that none of Shadden et al.'s subjects exhibited vowel deviations, in contrast to her subjects. Other explanations are possible for the conflicting results. Shadden et al. suggest that the effect of therapy for segmental errors might have a facilitating effect on suprasegmental features of the disordered children's speech. Additionally, several methodological errors identified in the Shadden et al. study render their results inconclusive at best. First, the study compared the T-TRIP *percentage correct* scores with the Templin-Darley *raw* scores (total correct). Second, the parametric statistics used to test the significance of the Shadden et al. data were inappropriate for their non-parametric sample, largely due to the fact that the sample size was small (n=10 pairs of scores), and not normally distributed. The possibility that the two studies' apparently similar speech-disordered groups actually were *not* homogenous must also be addressed. Shadden et al. did not report structural or functional assessment of their subjects' oral-mechanisms, although Henry did for her subjects: "None of the group was diagnosed as having any structural abnormality but some showed some difficulty in neuromotor control of the speech apparatus and additional fine motor co-ordination difficulties" (p. 123). A replication of the previous studies with two groups of disordered subjects, matched for severity of articulation disorder, but differing in their non-speech oral-motor capabilities, might explain the contradiction between Shadden et al.'s and Henry's results.

Overall, a review of existing studies of prosodic aspects of speech in articulation-disordered children reveals numerous differences in methodology. In spite of these incongruencies, disordered acoustic patterns emerge, i.e., more restricted f0 ranges, longer sentence durations, and less contrast in amplitude between adjacent stressed and unstressed syllables than used by control subjects. These tentative patterns must be confirmed by further research, which would also benefit from designs comparing the performance of a control group with two groups of carefully defined speech-disordered subjects.

Is there a single underlying cause for all of these prosodic disturbances? That question may be answered with a look at a model of speech production which incorporates prosody generation.

Levelt's Model of Speech Production

Figure 1 diagrams an adaptation of Levelt's "blueprint for the speaker," containing three interactive levels of processing leading to speech production (Levelt,

1989). Level one is called the “conceptualizer”; at this level the intention and ideas (content) for speech originate. Level two, the “formulator,” transforms these abstract ideas into words and sentences, and prepares a phonetic plan for use by the third level. At this level, the “articulator” executes the phonetic plan by activating the appropriate musculature (respiratory, laryngeal, and supralaryngeal) and overt speech is produced (Levelt, 1994; Levelt, 1989). Level two generates both segmental and suprasegmental patterns for speech and so is the level of primary interest here; it has three sub-stages: “grammatical encoding,” “phonological encoding,” and “phonetic encoding.”

Grammatical Encoding

Grammatical encoding follows the speaker’s conceptualization of the message to be spoken. Once this preverbal message is generated, the grammatical encoder retrieves appropriate words from the lexical storehouse (mental lexicon). Two kinds of information are retrieved from the lexicon: “lemma” and “lexeme”. The *lemma* is the word’s semantic and syntactic identity. For example, the word “sparrow” is retrieved along with its meaning (a type of small bird) and identity (a noun). The *lexeme* contains the word’s morphological and phonological information. For the word “dangerous”, the lexeme reveals that “danger” is the root and “-ous” is the suffix, that the first of its three syllables is accented, and which phonemes comprise its citation form (Levelt, 1989). Through retrieving *lemma* and *lexeme* information, grammatical encoding produces a preliminary structure (called *surface structure* by Levelt) for the speaker’s utterance. Next, this surface structure is transformed by phonological encoding into a string of syllables that level three, the *articulator*, can pronounce.

Phonological Encoding

During phonological encoding, the citation-form surface structure undergoes several procedures to transform it into the form to be used in connected speech. This transformation is necessary because some aspects of the word's phonetic shape are not predictable enough to be stored in the lexicon, and must be generated each time the word is spoken. "The same word, in different contexts, can be spoken with very different segmental and syllable structure, intonation, duration and amplitude" (Wheeldon & Levelt, 1995). Morphological/metrical spellout and segmental spellout procedures, along with the prosody generator, modify the surface structure to fit the requirements of the context.

Morphological/Metrical Spellout

During *morphological spellout*, the surface structure may be changed to various allomorphs that fit the context. For example, the lemma for the verb *to eat* contains multiple possibilities for inflection, including *eat, eats, eating, ate*. Morphological spellout procedures also consider the context of the entire utterance when choosing allomorphs for each citation word. For example, "I have bought it" may be changed to "I've bought it." *Metrical spellout* then specifies the number of syllables and the number of syllabic peaks (if any) for each morpheme. For example, the word "segmented" has three syllables with a syllabic peak on the second one (Levelt, 1989, p. 322). While morphological/metrical spellout procedures are underway, another procedure known as segmental spellout operates simultaneously (Levelt, 1992).

Segmental Spellout

This procedure generates the segmental composition of each word based on the morphemic and metrical information it receives. Its input is in the form of empty frames for stems and affixes (morphemic information), with the number of syllables and any syllabic peaks (metrical information) for each frame specified. These frames are filled with the phonemic composition of the word, probably organized into onset, nucleus and coda for each syllable. At this point the segments are in “citation form”, not programmed for coarticulation as they will be during connected speech. To transform the phrase’s successive words into connected speech, the segmental spellout for each work in the phrase is fed into the *segment-to-frame association area*.

Prosody Generator

In addition to metrical spellout, the prosody generator’s input includes surface structure, which contains the phrasal syntactic information. It also receives as input executive, attitudinal and emotional information about the “intonational meaning” of the utterance. *Intonational meaning* affects the speaking rate, intentional pausing, and general loudness level used by the speaker, as well as the location and direction of the utterance’s pitch contour. The speaker must also select a *key* and *register* for each intonational phrase. *Key* is the range of pitch movement in an intonational phrase, expressed as *high, mid,* or *low* intonation peaks; *register* is the pitch level of the baseline and conveys tension and emotion (Levelt, 1989, p. 315). The prosody generator uses this information to prepare its input for phonetic encoding.

Pitch contours are assigned to groups of syllables, forming intonational phrases. The phonological structure of each intonational phrase is then re-organized to make pronunciation easier. For example, “I want to go” may become “I wanna go” through cliticization. Individual segments may be lost or added, especially at word boundaries. Examples include “just fine” changing to “jus fine ” and “I’ve got you” becoming “I’ve got (ch)you" (Levelt, 1989, p. 364). The result is new syllables which cross word boundaries, creating *phonological words.* The three words "gave him it" unite as one phonological word: /geivɪmɪt/. The prosody generator then sets certain free parameters, including the duration, stress, and pitch of successive syllables, and the insertion of pauses between them (Levelt, 1989). These *parameters* correspond to the already-discussed suprasegmental features of speech, and are realized as metrical frames for phonological words. Once these parameters are set, the output of the prosody generator is channeled to the segment-to-frame association area.

Segment-to-Frame Association Area and Syllabification

In this area, the output of the segmental spellout procedure is united with the phonological words metrical frame. According to Levelt (1992), the metrical spellout procedure operates faster than the segmental spellout procedure. This enables the metrical frame to arrive at the segment-to-frame association area before the string of spelled-out segments arrives. As successive segments are spellout out, they are absorbed into the metrical frame (Levelt, 1992, p.21). This is followed by *syllabification,* "the chunking of successive speech sounds into pronounceable syllables" (Levelt, 1994, p. 100). After the phonological word has undergone syllabification, it enters the final stage of Level two, phonetic encoding.

Phonetic Encoding

Phonetic encoding fills these metrical frames with the motor plans for specific phonetic segments.   The phonetic segments have already been spelled-out by the segmental spellout procedure, then united with the metrical frames from the prosody generator. During phonetic encoding, syllabic gestural scores are retrieved from the speaker's mental *syllabary,* a storehouse for sets of articulatory gestures that will create the intended syllable (Levelt & Wheeldon, 1994). The parameters from the prosody generator are overlaid on these phonetic syllable plans, and the resulting program is ready for execution by the articulator.

Possible Sites for Breakdown

When disordered articulation or prosody results, it presumably points to an error at one of the stages of speech production. Errors toward the beginning are more "linguistic", while the latter stages are more "motoric" in nature. A useful method of discussing hypothetical sites where speech production breaks down, involves different populations known to have articulation and/or prosody disturbances. The speech of children who have dysarthria, developmental verbal dyspraxia, hearing disorders or phonological disorders may all be severely unintelligible, but for different reasons. The articulation errors of children with dyspraxia or dysarthria are primarily due to motoric disorders, while children with hearing impairments or phonological disorders make articulation errors primarily for linguistic reasons (Hoffman, Schuckers & Daniloff, 1989).

Linguistically-based Articulation Disorders

Hoffman, Schuckers, and Daniloff (1989) state that "hearing impaired children should logically have difficulties learning the relationship between acoustic aspects of the speech signal and the expression of meaning." In particular, their prosody may be disordered because their lexemes do not contain information about syllabic peaks (lexical stress), and they may not have access to other speakers' intonational meaning in order to produce their own. Similarly, their segmental misarticulations may be due to an improperly perceived phonology system, which forms the basis for improperly produced phonemes in their own speech.

Phonological errors are also characteristic of children without hearing impairments. According to Crary (1993), the term phonological disorder "implies a degree of disorganization within the rule system used to organize phonemes." Children with phonologically disordered speech may keep using natural phonological processes to simplify speech production beyond the age at which such processes have usually disappeared. Possible sites for breakdown include the mental lexicon, since it may contain simplified word forms (i.e., /ki / for /kip/) or the segmental spellout procedure since it may contain simplified forms of phones and phone-combinations (i.e., /tu/ for /skul/). Because the breakdown is at a linguistic level, before the message enters the prosody generator, prosodic disturbances would not be expected to co-exist with phonological disorders unless the syllable structure itself were altered by coarticulation. Instead, these children exhibit articulation errors involving multiple phonemes. Often one or more natural classes of phonemes (i.e. fricatives) are affected, although a phone may be correct in one sound environment, but incorrect when coarticulated with other sounds (Hoffman et al., 1989; Crary, 1993). This inconsistency distinguishes children with linguistically-based phonological-impairments from those with dysarthria, a motorically-based impairment which features more consistent errors.

Motorically-based Articulation Disorders

Children with dysarthric speech may not be able to fully execute the phonetic plans for speech because of neurologic damage or muscular weakness, breakdowns at the level of Levelt's *Articulator.* Love (1992) defines childhood *dysarthria* as "a neurogenic speech impairment caused by dysfunction of the motor control centers of the immature central and/or peripheral nervous systems and marked by disturbances of strength, speed, steadiness, coordination, precision, tone, and range of movement in the speech musculature" (p 8). This disorder affects both segmental and suprasegmental aspects of speech, because both are achieved with the same speech musculature. The *Articulator* is pinpointed as the likely site for breakdowns in speech production for dysarthric speakers.

Apraxia, another motorically-based articulation disorder, may be due to a breakdown at an earlier stage: during the retrieval of articulatory programs from the syllabary. Levelt states that: "the skilled language user has an inventory of syllable plans, a stock of frequently used motor programs" (1989, p. 327). However, some children may be "unskilled language users" who experience difficulty with organizing syllable plans for storage or retrieval. This corresponds to many descriptions of apraxic or, in the case of children, dyspraxic speech.

Love (1992) defines childhood *developmental verbal dyspraxia*, or DVD, as "an impaired ability of the child, in the absence of obvious muscular disturbance of the speech mechanism, to execute voluntarily the expected motor gestures and programming of gestures needed for the articulation of speech" (p 9). In addition to segmental articulation disorders, prosodic disturbances have also been reported in the speech of apraxic children (Edwards, 1973; Rosenbek & Wertz, 1972; Shriberg et al., 1997b). This indicates that, at least in these children, the breakdown may occur within the prosody generator itself, possibly when the metrical frames containing both the suprasegmental parameters and the re-syllabified segmental spellout are generated.

Differential Diagnosis of Articulation Disorders

The complexity of Levelt's model, and the corresponding large number of sites where breakdowns may occur, underscores the fact that no two children will have exactly the same type of speech disorder. Developmental speech disorders, whether linguistically-based or motorically-based, by definition involve children who are changing. Hodge (1994) points out that since children's development speech processes develop over time, and since some children use compensatory techniques more effectively than others, variability and differing severity levels should be expected in childhood disorders. Therefore, although prosodic disturbances appear likely to occur in motorically-based speech disorders, some children with motoric disturbances may not exhibit prosodic disturbances. Likewise, some children with linguistically-based speech disorders may have co-existing disorders which create prosodic disturbances which would not otherwise be expected. This reinforces the importance of carefully defining the subjects in any study of prosody in children with speech disorders.

Shriberg et al. (1997c) suggested that future research into the viability of prosodic disturbances as a diagnostic marker for DVD should incorporate comparison to populations with other childhood speech disorders in which DVD has been reported or suspected. Children with severe phonological disorders are often unintelligible and may be misdiagnosed with DVD. Likewise, children with DVD may be misdiagnosed as having severe phonological disorders. Both consequently receive treatment which is ineffective. Earlier it was suggested that a replication of Shadden et al.'s (1980) and Henry's (1990) studies with two groups of disordered subjects, matched for severity of articulation disorder, but differing in their non-speech oral-motor capabilities, might explain the contradiction between their results. DVD and phonological disorders are logical choices for comparison, given the fact that each is often misdiagnosed as the other.

However, this fact also makes it difficult to establish reliable groups for comparison.

Severity alone does not distinguish DVD from phonological disorders (Velleman, Huntley, & Lasker, 1991). Researchers even disagree about whether DVD exists only as a homogenous group, “pure DVD”, or may also be found in combination with other disorders, such as dysarthria. Therefore, instead of attempting to locate a group of subjects with “pure DVD,” the present study will classify disordered-speech subjects according to the presence or absence of oral-motor difficulties. Subjects will be matched for severity of articulation disorder using the Assessment of Phonological Processes - Revised, APP-R, (Hodson, 1986). The Oral and Limb Praxis Test, OLPT (adapted from Dewey, Roy, Square-Storer, & Hayden, 1988), will be used to subdivide the disordered group into a subgroup with motorically-based oral-motor speech disorders, and a subgroup with linguistically-based speech disorders.

Assessment of Prosody

Shriberg et al. (1997a) speculated that inappropriate stress marks at least one subtype of DVD. Three studies supporting this hypothesis (Shriberg et al., 1997b; Shriberg et al., 1997c) were described earlier. The Shriberg et al. studies used the Prosody Voice Screening Profile, PVSP (Shriberg et al., 1990), to assess the use of prosodic features by articulation-disordered children. However, other studies of children’s prosody used different assessment methods, i.e., the Tennessee Test of Rhythm and Intonation Patterns, T-TRIP (Koike & Asp, 1981). This leads to the question of how best to assess prosody in clinical populations. The few diagnostic instruments currently being used for such assessment are detailed in Table 5.

Table 5: Diagnostic Instruments Available for the Assessment of Prosody

Instrument	**Areas of Prosody Addressed**	**Context of Sample**	**Scoring Method**
Profile of Prosody, PROP (Crystal, 1992)	Intonation, tempo, stress	Spontaneous speech samples	Guidelines provided to help clinician determine presence or absence of prosodic problem
Voice Assessment Protocol, VAP (Pindzola, 1987)	Pitch, loudness, duration, tempo	Various tasks, i.e., sustained vowel phonation, conversational & automatic speech	Guidelines provided to help clinician decide presence or absence of prosodic problems. Tape recorded samples aid in pitch matching.
The TAKI Task (Allen, 1983)	Lexical stress	Novel words with contrasting stress patterns, i.e., *'taki* vs. *ta'ki*	Developed for experimental purposes; scoring not specified.
Screening Test for Developmental Apraxia of Speech, STDAS (Blakely, 1980)	Rate, stress, phonemic spacing, and inflection	Conversational speech samples	Primarily a test for DVD, prosody assessed only in one subtest. Prosody is subjectively rated using a 3- point scale. No examples are provided to guide ratings.
Prosody Voice Screening Profile, PVSP (Shriberg et al., 1990)	Phrasing, rate, pitch, stress, loudness, laryngeal quality, & laryngeal resonance	Conversational speech samples	Clinicians follow explicit directions for coding of utterances. Training tapes and exercises instruct clinician re: coding
Tennessee Test of Rhythm and Intonation Patterns, TRIP (Koike & Asp, 1980)	Stress, tempo, intonation, rhythm	Imitation of reduplicated nonsense syllable "mɑ"	Examiner judges correctness of imitative response. Best 1 of 2 attempts scored. Total % correct score obtained

Evaluation of prosody may be perceptual or instrumental (Hargrove & MacGarr, 1994). Clinicians make either perceptual judgements of its subcomponents (i.e. pitch, loudness, and relative timing of syllables) or quantitative measurements of their acoustic correlates (fundamental frequency contours, amplitude or intensity, and duration of syllables). According to Hargrove and McGarr, quantitative instrumental measurements should merely *augment* perceptual assessments, since a speaker's prosody is judged by human listeners. However, although perceptual judgements may be superior at ascertaining the *presence* of a prosodic problem, acoustic measurement is superior at *quantifying* the problem. Therefore, a combination of perceptual and instrumental assessment is preferred. All of the assessment tools reviewed use only perceptual evaluation. Two of these assessment tools look only at narrow regions of prosody, thus limiting their effectiveness. The Voice Assessment Profile, VAP, (Pindzola, 1987), is primarily aimed at the identification of voice disorders. Prosody is just one component of this analysis. The TAKI Task (Allen, 1983) was developed for experimental purposes and focuses only on lexical stress. The other tools take a broader view of prosody, but some are more suited for the task of evaluating the prosody of speech disordered children than others.

Blakely's Screening Test for Developmental Apraxia of Speech,, STDAS (1980), Crystal's Profile of Prosody, PROP, (1992), and Shriberg et al.'s PVSP (1990), all utilize conversational speech samples. However, the STDAS includes assessment of prosody only in a subtest, since the primary purpose of the STDAS is to identify DVD, not to assess prosody. Rosenbek and Wertz's (1972) statement that prosodic disturbance may "*occasionally* be a distinguishing difference between children with functional articulation disorders and children with DVD" provides the rationale for inclusion of a prosody subtest. Conversational speech samples are observed for "deviance in rate, phonemic spacing, inflection, or stress" (Blakely, 1980). The STDAS uses the following three-point scale to rate prosody: 1 = normal prosody; 2 = slight but apparent deviance in prosody; 3 = readily apparent deviance which adds a distinct characteristic to speech (Blakely, 1980). The use of such a rating scale addresses prosody as a whole rather than providing separate assessment of individual aspects of prosody. Moreover, the test as a whole is based on characteristics of DVD which are not universally agreed upon and therefore the STDAS is not widely used.

The PROP (Crystal, 1992) uses spontaneous speech samples to determine the presence or absence of prosodic problems. The PROP examines "intonation (direction of nuclear pitch accent), tempo (phrasing), and stress (phrasal), along with strategies for producing stress" (Hargrove & McGarr, 1994). Tape recorded examples of impaired prosody are provided with the PROP. According to Hargrove and McGarr (1994), the tape is not appropriate for training listeners to judge prosodic errors reliably, and some aspects of disturbed prosody (i.e., intonation and phrasal stress) are addressed in more detail than others (i.e., rate and rhythmic problems).

The PVSP (Shriberg et al., 1990) screens selected parameters of prosody and of voice on an individual basis. It includes judgements of seven suprasegmental aspects of conversational speech; phrasing, rate, and stress are considered prosodic aspects, while loudness, pitch, laryngeal quality, and laryngeal resonance are assessed as aspects of voice. Training tapes are provided to instruct clinicians in the rigid criteria for segmentation, inclusion, exclusion, and coding of utterances, as well as to ensure intra- and inter-judge reliability (Hargrove & McGarr, 1994). Since PVSP speech samples must be coded by trained research transcriptionists who are provided only with the child's age and gender, these requirements limit the clinical applicability of the PVSP. In addition, a more structured assessment task than conversational speech may yield more specific information about children's error patterns.

Speech samples for evaluation may be elicited by various methods: unstructured conversation (i.e., spontaneous speech), structured conversation (i.e., eliciting contrastive stress in target words during play), or structured imitation tasks (using nonsense or meaningful stimuli). Conversational prosody is used in most diagnostic tools. Since conversational intelligibility is the desired functional outcome, this has ecological validity. However, in some cases it is important to have more control over the speech sample.

The imitative method may be more effective than conversational elicitation in assessing the prosody of children with DVD. Rosenbek and Wertz (1972) suggested that imitated prosody rather than conversational prosody will be disturbed in dyspraxic children. Their hypothesis builds on the common observation that voluntary production of phonemes by dyspraxic speakers is often in error although the same phonemes may be correctly produced in spontaneous speech (Edwards, 1973; Love, 1992; Shriberg et al., 1997a). Imitation tasks also allow for control of task length and complexity, other factors that affect speech production in children with DVD. From a practical test administration standpoint, the perceptual judgement of same or different, required by imitative tasks, is more accessible to clinicians scoring items "on-line," than lengthy transcription and coding, required by conversational tasks such as Shriberg et al.'s PVSP.

In their discussion of the rationale behind the development of the Tennessee Test of Rhythm and Intonation Patterns, T-TRIP, Shadden, et al. (1980) state that "ideally, for clinical and research purposes, systematic investigation of suprasegmental patterns requires a test that is both sensitive and repeatable." The T-TRIP was developed using 3- and 5-year-olds with normal hearing and language skills and was designed to assess imitation of syllable stress, timing, and intonation contours. The T-TRIP has been shown to be sensitive to developmental differences between 3- and 5-year-old children (Koike & Asp). Haak and Darling (1993) have even shown through foreign language studies with the T-TRIP that the test can discriminate prosodic deficits across languages.

Several aspects of the T-TRIP's design support its use with speech disordered subjects. In order to increase task length and complexity, the T-TRIP items vary in stress, timing and intonation patterns, as well as in number of syllables per item. The 25 test items are composed of reduplications of the nonsense syllable /mɑ/. The T-TRIP's authors chose /mɑ/ because of its articulatory simplicity. This simplicity makes the test appropriate for use with children with severe articulatory difficulties since their prosodic ability can be assessed independent of their articulatory ability. Should prosody disturbances prove to be the diagnostic marker that Shriberg et al. propose, the T-TRIP would be a viable tool for assessment of imitative prosody in children suspected of having DVD.

Purpose

The purpose of the present study was to use the T-TRIP to assess and compare the imitative prosody of children with normal speech development, motorically-based speech disorders, and linguistically-based speech disorders. Currently, differential diagnosis of oral-motor speech disorders, like DVD, is complicated by similarities to severe phonological disorders. Early diagnosis of oral-motor speech disorders would enable more effective treatment to take place earlier, which should in turn produce better treatment outcomes.

Studies such as the current one, in which the prosody of children with suspected DVD is compared to the prosody of children with similar severe articulation disorders, will provide more information about the viability of prosody as a diagnostic marker. The T-TRIP's perceptually judged scores will also add to the knowledge base regarding the imitative prosodic abilities of children with disordered speech, whether or not a difference in prosody among children with different types of disordered speech is found.

Design

Acoustic analysis of individual responses to the T-TRIP was incorporated into the design of the present study with the expectation that this method might reveal more information about the manner in which the prosody of children with DVD differed from that of their peers with normal speech. The present study fills a gap in the literature by incorporating a comparison among the prosody of children diagnosed with different types of speech disorder, as well as between the prosody of children with disordered speech and their age-matched peers with normally developing speech. In addition, the design of the present study is unique in employing acoustic analysis to describe children's imitative prosody, as well as comparing acoustic and perceptual assessment results.

Since T-TRIP scores only provide judgements about the subjects' response as a "correct" or "incorrect" imitation of the model, an additional type of perceptual assessment was added to the study. Three experienced listeners transcribed the stress and intonation patterns of each of the acoustically analyzed items, without benefit of knowing the target pattern. This second, "more-rigorous" perceptual assessment added to the study by confirming that the items which showed the correct acoustic pattern were in fact perceived as having that pattern. In addition, the transcriptions specified what the perceived stress or intonation pattern of the *incorrect* items were, in order to complete a more thorough error-analysis.

Research Questions

The present study attempted to answer four questions regarding the following three groups of children:

a) linguistically-based speech disorders,

b) motorically-based speech disorders, and

c) normal speech development.

1) Do the overall percentage correct scores on the T-TRIP differ by group?

2) Do the subtest percentage correct scores on the T-TRIP differ by group?

3) Do the three groups differ from each other in their use of amplitude, frequency and duration to imitate prosodic patterns on selected items from the T-TRIP's rhythm subtest (Part I)?

4) Do the three groups differ from each other in the magnitude or direction of frequency change used to imitate prosodic patterns on selected items from the T-TRIP's intonation subtest (Part III)?

CHAPTER 2

METHOD

Subjects

Nine kindergarten children from west central Florida participated in the study. Eight of the children were selected from a classroom in Pinellas County composed of children with articulation disorders and their peers with normal speech skills. The classroom was a unique environment in that all of the children participated in oral-motor exercises, regardless of their speech skills. Three children with normal speech served as age-matched controls for six children diagnosed with disordered speech. The disordered speech participants were divided into two groups: those diagnosed with motorically-based (i.e., oral-motor) speech disorders and those diagnosed with linguistically-based (i.e., phonological) speech disorders. In addition, one subject from the University of South Florida's Communication Disorders Clinic with an oral-motor speech disorder was included to balance the groups. Two boys and one girl were in each group, and the age range and mean age for each group was comparable as shown in Table 6.

Table 6. Age Range and Mean Age for Subject Groups at Time of Testing.

Group Label	Speech Diagnosis	Age Range	Mean Age
A	Phonological Speech Disorder	68 mos to 71 mos (5-8 to 5-11)	69.0 mos
B	Oral-Motor Speech Disorder	66 mos to 72 mos (5-6 to 6-0)	68.7 mos
C	Normally Developing Speech	66 mos to 67 mos (5-6 to 5-7)	66.5 mos

Groups were labeled as follows: Linguistic Speech Disordered (Group A); Oral-

Motor Speech Disordered (Group B); and Normal Speech (Group C). The children in the disordered speech groups were receiving treatment for moderate to severe articulation disorders at the time of their participation in this study. Random selection methods were not employed due to the limited number of participants. Rather, participants were selected and assigned to groups according to the following criteria.

Inclusion & Exclusion Criteria

Inclusion criteria for all participants in the study included: chronological age from four years to six years five months, normal hearing ability, age-appropriate receptive language skills, intelligence within normal range, no evidence suggesting attentional disorder or emotional disturbance, no evidence suggesting a learning difficulty, the absence of any known neurological insults, and normal oral-peripheral structure and functioning. Normal hearing sensitivity was determined by the county's annual hearing screening. Receptive language ability was determined by scores on the Clinical Evaluation of Language Fundamentals-Preschool, CELF-P (Semel, Wiig, and Kwiatowski, 1987), the Preschool Language Scales-3rd Edition, PLS-3 (Zimmerman, Steiner & Pond, 1992), or by teacher report. Intelligence was determined to be normal as reported by the classroom teachers. To ensure that subjects were free of known neurological insults, all relevant and available school, medical records, and clinical records were examined. Available records were also examined for labels of Attention Deficit Disorder (ADD), Attention Deficit-Hyperactivity Disorder (ADHD), emotional disturbance, or learning disability. All children included in the study were labeled as "speech impaired" only.

All of the children with disordered speech had phonological deviancy scores of 40 or higher and were rated as moderate to severe on the Assessment of Phonological Processes-Revised, APP-R (Hodson, 1986). Based on their performance on the Oral and Limb Praxis Test, OLPT (adapted from Dewey et al., 1988), a non-standardized series of limb and oral praxis tasks, the children with speech disorders were categorized into either the oral-motor or the phonological speech disorders group. The praxis tasks are listed in Appendix A.

Subjects' performance on each of the praxis tasks was rated on a four point scale. A rating of three indicated an immediate and correct response to verbal command, two indicated a correct response following a model, one indicated a correct response following both a model and a tactile cue, and zero indicated no correct response was elicited using any cues. The total points obtained were then divided by the total points possible to obtain a percent correct score. Children with a percent correct score of less than 90% who also displayed evidence of struggle or groping during oral or articulatory behaviors were placed in the oral-motor speech disorders group. Children with a mean score of 90% or higher who displayed no evidence of struggle or groping during oral or articulatory behaviors were placed in the phonological speech disorders group.

The same battery of tests was administered to the control group of children. The control group demonstrated age-appropriate receptive and expressive language skills, were rated as having age-appropriate speech articulation, evidenced no oral or limb apraxia, and exhibited no struggle or groping during oral or articulatory behaviors. Control subjects scored in the mild range (0 - 19) on the APP-R and had OLPT scores of

90% or higher. APP-R scores, OLPT scores, and information on hearing screening results, receptive language ability, and cognitive level for all subjects are included in Table 7.

Table 7. Scores for Inclusion and Exclusion Criteria.

Group - Subject #	**APP-R**	**OLPT**	**Hearing Screening**	**Receptive Language**	**Cognition**
A-1	56	90 %	Pass	Normal by Teacher Report	Normal by Teacher Report
A-2	42	96 %	Pass	Normal by Teacher Report	Normal by Teacher Report
A-3	54	93 %	Fail*	Normal by Teacher Report	Normal by Teacher Report
B-1	50	81 %	Pass	PLS-3 AC = 93 (Zimmerman et al., 1992)	Normal by Teacher Report
B-2	55	68 %	Pass	Normal by Teacher Report	Normal by Teacher Report
B-3	50	45 %	Pass	CELF-R EL = 97 (Semel et al., 1987)	Normal by Teacher Report
C-1	0	100 %	Pass	Normal by Teacher Report	Normal by Teacher Report
C-2	0	100 %	Pass	Normal by Teacher Report	Normal by Teacher Report
C-3	12**	100 %	Pass	Normal by Teacher Report	Normal by Teacher Report

* Hearing screened prior to surgery to place P.E. tubes. Testing occurred after surgery.
** Reflects liquid /r/ and /l/ gliding only

Stimuli

The items from the Tennessee Test of Rhythm and Intonation Patterns, or T-TRIP, (Koike & Asp, 1981) were used to elicit a sample of each child's imitative prosody. The T-TRIP contains 25 items composed of reduplications of the nonsense syllable /mɑ/. Items vary in stress, timing and intonation patterns and are presented in three subtests: rhythm, tempo, and intonation. A transcription of the T-TRIP is included in Appendix B (Koike & Asp, 1981). Musical notations accompany each item to indicate the appropriate tempo. However, since the symbols used by Koike and Asp are uncomplicated, knowledge of musical notation is not necessary to decipher the transcription. Each syllable is represented by a circle. Stress is indicated by the circle's size, small for unstressed and large for stressed. Pitch level (low, mid, or high) is indicated by the location of each circle on the top, middle or lowest line. When more than one syllable comprises one beat, the syllables are joined by a curved line. Pauses are indicated with a vertical line between circles. Therefore in part I, the rhythm subtest, item 2 is "ma MA´ ", while item 14 is "MA´ ma (pause) MA´ ". For the tempo subtest, part II, the syllables all have equal pitch and stress, with three beats per test item. Syllabic duration varies as the tempo increases from one to three syllables per beat. Accordingly, item 15 is "ma ma ma", while item 16 is "mama mama mama." In part III, the intonation subtest, rising and falling intonation contours are indicated with arrows pointing upward or downward. Gradually changing intonation is symbolized by a longer, straight, dotted line attached to the arrow head, while fast changing intonation is symbolized by a shorter, curving, solid line. For example, items 18 and 19 have the same degree and rate of change of their intonation contours, although the direction of the contour changes. Item 18 is spoken as a statement with falling intonation: "ma." Item 19 is spoken with rising intonation, as a question: "ma?".

Koike and Asp (1981) recorded the T-TRIP stimulus items spoken by an adult male speaker. Wide- and narrow-band spectrographic analyses were employed to verify the accuracy of the speaker's production of the test items (Koike & Asp, 1981). In addition, perceptual evaluation of the recorded stimuli was performed by three judges trained in narrow phonetic transcription. Interjudge agreement was 100% regarding stress, intonation, number and length of syllables in each test item. Complete details of the recording and validation of the T-TRIP stimuli can be found in Koike and Asp (1981).

Testing Procedure

The T-TRIP was given by either a graduate clinician or a Ph.D. level speech-language pathologist trained in its administration. Each child was seated in a small, quiet room facing the examiner. Koike and Asp's (1981) pre-recorded stimuli were presented in the sound field from a Tascam portable DAT player (model DA-P1 by Teac) with (Radio Shack AMX 9) model portable speakers. The stimuli were presented at a comfortable loudness level of approximately 72 dBA SPL. Responses were recorded onto high quality tape (3M's AVX60 Professional cassettes) with a portable Marantz Superscope C0200LP tape recorder. A lapel microphone (omnidirectional impedance model 33-3003 by Radio Shack) was clipped to the child's shirt approximately six inches from his or her mouth.

The child was instructed that he or she was going to play a game similar to "Follow the Leader," and to "listen carefully and repeat exactly what the leader on the tape says." Five pre-recorded training items (corresponding to items 1, 2, 3, 4 and 13) were each presented twice, with time allowed after each presentation for the child's response. If the child did not successfully perform the task, these practice items could be repeated by the examiner until the child demonstrated understanding of the testing procedure. This was not necessary except in the case of the last practice item, corresponding to item 13. This item contains a one-beat rest (pause) and the subjects

tended to respond to the item within the pause rather than wait for the model to complete the item. No more than one repetition of this item was needed per subject.

The 25 pre-recorded test items were then presented twice via the recorded model and each response was scored immediately by the clinician. In the case of a child not responding within the allotted time, the tape player was paused and the item was repeated a third time by the examiner (live voice). If a child's response was interrupted by the taped presentation of the next stimulus item, the item was re-administered by the examiner in order to get a clear recording of the subjects' response for later analysis.

Scoring Procedure

According to Koike and Asp's (1981) criteria, a correct response included the appropriate number of syllables and the same stress, tempo, and intonation pattern as the stimulus item. Intelligibility of the syllable /mɑ/ was not required for a correct response, as suprasegmental information was paramount. Examiners used these criteria during on-line scoring and subsequent re-scoring for reliability. Both attempts at each of the 25 items were scored, although only one accurate imitation was required for an item to be scored as correct. If the child responded more than twice to each item, only the first two responses were scored. In the case of an immediate self-correction, however, the corrected response was scored in lieu of the original responses. Because the stimuli were played in the sound field, during on-line scoring the examiners were able to judge each response as "same or different" compared to the model. Responses to live voice presentation of the test items were not scored, except when the subject's response overlapped the taped presentation of the model, rendering the response uninterpretable.

Re-scoring Procedure

A graduate-level clinician then re-scored each subject's complete test from the recording, which contained both the model and the response, in order to establish inter-rater reliability. However, the on-line scores were uniformly less than the scores generated by re-scoring. On-line scores were theorized to be less reliable because of the demands of overseeing both the subject and the equipment, as well as the fact that on-line scores were made by three different test administrators. Also, the on-line scores of the test administrators could have been biased by knowledge of which speech diagnosis group the subjects belonged while the re-scoring was done without such knowledge. Additionally, during re-scoring, the subjects' responses could be played multiple-times if needed to determine the accuracy of the response. Accordingly, it was decided that two more graduate-level clinicians should re-score the test from the recordings, in order to get a consensus. Reliability of these three scores was 90%. Scores from the three clinicians were averaged to obtain each subject's T-TRIP subtest scores and total percentage correct.

Statistical Analysis of Perceptual Scores

Since the number of subjects in each diagnostic group was small (n=3), non-parametric statistics were chosen to evaluate the significance of the scores on the 25 T-TRIP items. A Kruskal-Wallis 2-way ANOVA test for independent samples was selected to analyze results both by test score and by group.

Acoustic Analysis

The T-TRIP results of other normal and speech-disordered subjects involved in ongoing research at the University of South Florida's Communication Disorders Center (none of which participated as subjects in the present study) were used to identify which of the 25 test items were best suited for acoustic analysis. Items 1 and 2 were excluded because almost all subjects responded correctly to these two-syllable items. The subjects who missed them seemed to be just “warming up” and went on to correctly imitate longer items. Items 13 and 14 were excluded because most subjects responded separately to each half of the model, without waiting for the complete item to be presented. The tempo subtest (Part II - items 15, 16, and 17) was excluded because it contained only equal-stress, equal-pitch items which primarily tested timing and therefore did not require measurement of multiple parameters by acoustic analysis. Examples from the intonation subtest (items 18 - 25) were not included by Koike and Asp (1981) in the practice items, therefore most subjects used items 18 and 19 to become attuned to the task. Items 18 and 19 were excluded for this reason. Each subjects’ response to the remaining items, 3 - 12 and 20 - 25 was digitized for further acoustic analysis.

For each subject, the second response from each of the sixteen selected test items was used, except when the first response contained the correct number of syllables while the second response was incorrect due to syllable addition or omission . In these cases (10 of 144 responses = 7 %), the first response was used for analysis. The second response was preferred since it allowed the subjects to have a “ practice” attempt at imitating the pattern during the first response.

Responses were digitized at a sampling rate of 20 kHz using Kay Elemetrics Computerized Speech Laboratory model 4300-B. During analysis, a spectrogram with an LPC formant history superimposed, and a pitch contour were displayed in addition to the waveform. An example of the CSL’s display screen with each analysis window is included in Appendix C. Settings for each of the analysis windows are listed in Appendix D. All measurement values were saved to a file and later exported to a spreadsheet for statistical analysis. Further measurements were chosen for each group of subtest items in order to reflect the differing task requirements.

The CSL's *pitch synchronous* option was used for f0 measurements. This option involved a subroutine which placed *impulse markers* corresponding to each glottal pulse on the waveform. The wavelength between each pair of impulse markers was used to calculate f0 with the CSL's 'pitch extraction' subroutine. The *pitch synchronous* option also enabled manual-editing of impulse marks in cases where the subroutine omitted markers, increasing the wavelength and consequently lowering the mean f0. This occurred primarily on low amplitude final syllables, on which the gain (amplitude) could be temporarily increased in order to facilitate editing of impulse marks on the waveform by visual inspection. As a secondary method of f0 measurement, *pitch asynchronous* f0 values for each 20 ms frame of speech were made using the 'pitch extraction' and 'numerical results' subroutines without impulse marks. This yielded numerical f0 values for the entire item, which were saved to a file for later calculation of the mean f0 for each syllable in case the primary method was unsatisfactory. However, this secondary f0 calculation was not needed.

Rhythm Subtest

The information of interest for the rhythm subtest was the *relative stress* given to each syllable. Accordingly, f0, amplitude and duration were measured for each syllabic nucleus. Using both the waveform and wideband spectrogram displays, the vowel nucleus of each syllable was identified and marked with cursors. The CSL's 'result statistics' subroutines for pitch extraction and energy calculation were used record the duration of the marked segment, and to generate mean, median, and range values of the f0 and amplitude for each syllabic nucleus.

Intonation Subtest

Stimuli on the intonation subtest required imitation of rising or falling intonation at different rates and over increasing numbers of syllables. The information of interest was the child's ability to imitate the *pitch contour's* degree of change, rate of change, and direction of change. The degree of change was measured by calculating the f0 range for each syllable with the CSL's 'result statistics' subroutine for pitch extraction. The rate of change was then calculated by dividing the range of f0 by the duration of the syllabic nucleus. The direction of the intonation contour was obtained by visual inspection of the pitch contour display and a narrow-band spectrogram. Intonation was judged to be increasing if the contour inclined, decreasing if the contour declined, or unchanging if the display revealed no slope. Rise-fall or fall-rise contours were also noted. Direction of change was confirmed by the CSL's 'numerical results' subroutine for pitch extraction for each syllable.

Reliability

A graduate assistant digitized the responses and assigned pseudonyms to each subject so that their identity and speech diagnosis were protected and acoustic measurement could be done without bias by the author. Ten percent (10 %) of the selected test items were re-analyzed by another graduate-level clinician to establish inter-rater reliability of measurement. Five percent (5 %) of the items were re-analyzed by the author to establish intra-rater reliability of measurement. Reliability was determined separately for each parameter measured in the two subtests. The difference between the original value and the re-measurement value was divided by the original to obtain the percentage discrepancy between the two measurements. These values were subtracted from 100% to obtain the percentage of agreement between the two measurements. Inter- and intra-judge reliability values are presented in Table 8.

Table 8. Intra- and Inter-judge Reliability for Acoustic Measurement

Subtest:	**Rhythm**	**Rhythm**	**Rhythm**	**Intonation**	**Intonation**
Parameter:	**Duration**	**RMS Amplitude**	**Mean f0**	**Duration**	**Change in f0**
Intra-rater	96%	100%	100%	85%	87%
Inter-rater	90%	99%	99%	73%	89%

Examination of Table 8 reveals that measurements of the duration of the syllabic nucleus were least reliable. This was attributed to the difficulty of establishing acoustic landmarks between the nasal phoneme /m/ and the following vowel. For the rhythm subtest, the steady-state portion of the vowel was used to isolate the vocalic nucleus. Items in the intonation subtest did not contain steady-state vowels because of the rising or falling pitch contours, therefore syllable duration was even more difficult to measure for these items. For both subtests, however, the other acoustic parameters demonstrate good intra- and inter-rater reliability. This is particularly important because the RMS amplitude, mean f0, and change in f0 values for each item were calculated based on the portion of the vowel identified during the duration measurement. Despite the difficulty of measuring duration, the other acoustic parameters had intra-rater measurement reliability of 87% to 100%, and inter-rater measurements had reliability of 89% to 99%.

Transcription of Selected Items

In addition to on-line scores and re-scoring of the subject's responses, the stress or intonation pattern of each response was transcribed for each of the acoustically analyzed items. Transcriptions were used to judge the acoustically analyzed items as correct or incorrect. The transcriptions were made by three experienced listeners - Ph.D. level faculty of the University of South Florida's Communication Disorders Clinic. Judges were provided with a list of target items to "match" with responses, but were cautioned that patterns other than the targets were to be expected. Transcription was made without knowledge of what the actual model for each item was and the judges were not informed as to which group the subjects had been assigned. The responses of all nine subjects to the ten items from part I (rhythm subtest) were presented in random order via the CSL. The first ten responses were presented for practice. These responses appeared again at the end of the random order; the second transcription of the practice responses was used. The procedure was repeated with the six items from part III (intonation subtest), so that responses were blocked by subtest but not by speaker. Six intonation responses were presented as practice and repeated again at the end of the random order. The responses were each played twice, and judges could request additional re-plays if needed.

For the ten items in the rhythm subtest, judges transcribed primary and secondary stress. The target stress pattern was then compared to the transcription of each response in order to judge whether the subject's response was perceived as a "correct" or "incorrect" imitation of the model. According to Koike and Asp (1981), T-TRIP items had only two levels of stress - "stressed" or "unstressed." A third stress level - "secondary stress" was added to the transcription procedures in deference to the four items with only one stressed syllable. These items (4, 5, 8, & 9) were considered correct even if an unstressed syllable was transcribed with secondary stress, providing the target stressed syllable had received primary stress and the response contained the appropriate number of syllables. Agreement among the three transcription judges was 73% for items in the rhythm subtest. The two most experienced judges (both phoneticians) had an agreement of 90%; the third judge had extensive experience with phonetic transcription but limited experience transcribing stress patterns. Items where two or three judges transcribed the same stress pattern as the target were considered to have been perceived as "correct"; items where only one or zero judges transcribed the target stress pattern were considered "incorrect".

For the items in the intonation subtest, the intonation contour for each syllable was transcribed as *rising, falling,* or *level*. In addition, one judge transcribed level tones as *high, medium,* or *low* in pitch. All judges commented that transcribing intonation was a difficult task for which they had had limited practice. In addition, instructions for the intonation transcription task were inadequate. Due to a literal reading of the T-TRIP's musical notation, the investigator neglected to tell the judges to mark level pitch.

According to the musical notation, the intonation subtest contains no syllables with level pitch. (See Appendix B for T-TRIP musical notation.) Items 20 and 21 are marked with "fast fall" and "fast rise" intonation; items 22, 23, 24 and 25 are marked with "slow rise" or "slow fall" intonation. However, narrowband spectrographic analysis of the pre-recorded model for these items revealed the use of level pitch for one or more syllables in each of the items with "slow rise" or "slow fall" intonation. Since

transcription judges were not instructed to expect level pitch, this was confusing. The spectrograms of the model used a bandwidth of 45 Hz, and are included in Appendix E.

Agreement between the two most experienced transcription judges was less than

44%. Because of the difficulty of the task and the poor inter-judge reliability, the transcriptions were not used to judge items on the intonation subtest as “correct” or “incorrect.” Instead, narrowband spectrograms of each response and graphs of the acoustic measurements were examined to determine “correctness.” Narrowband spectrograms of the pre-recorded model were also used for comparison with the responses.

Analysis of Acoustic Data

Each of the items selected for acoustic analysis was graphed to enable visual inspection of the relative values for each acoustic parameter. Appendix F includes graphs of items 3 - 12. Graphs of items 20 - 25 from the intonation subtest were included in Appendix G.

Rhythm Subtest Items

Duration, RMS amplitude, and mean f0 values for each syllable’s vocalic nucleus were graphed separately for items in the rhythm subtest. This separation facilitated examination of the graphs for any preferences for the use of particular parameters to indicate stress patterns. Since f0 was measured in Hz, these values were converted to semitones using the following formula: $x = 12(\log_{10} f_2 - \log_{10} f_1)/\log_{10} 2$, where f_1 = 16.35 Hz (0 semitones) and f_2 = the value (in Hz) to be converted (Baken, 1987). Each response was marked as “correct” or “incorrect” according to the transcription procedure described above. The transcriptions provided confirmation that the items which showed the correct pattern of acoustical parameters were actually *heard* as having the correct pattern.

Perceptually “correct” items were examined to determine which acoustic parameters had been used to indicate the stress pattern, i.e. duration, amplitude, f0 or some combination of the three. A parameter was considered to have been used if the response demonstrated that target stressed syllables exceeded the target unstressed syllables by the minimum perceptual value for that parameter. These minimum values were established according to the *just noticeable differences* (JNDs) discussed in Chapter 1. Target stressed syllables had to exceed target unstressed syllables by 1 dB in amplitude, or by 1 semitone in pitch, or by 10 ms in duration. For example, if the target stress pattern was *SuS (stressed-unstressed-stressed),* and the first and third syllables exceeded the second syllable by 1 dB in RMS amplitude, the parameter of *amplitude* was employed. An exception was made if all syllables except the final syllable reflected the target stress pattern in their *duration* values. Since utterance-final syllables are often prolonged (Klatt, 1976) by 60 - 200 ms due to pre-pausal lengthening, items which demonstrated this phenomenon were accepted as having used *duration* to replicate the target stress pattern. Therefore, if the target contained a final *unstressed* syllable which was *lengthened*, this was accepted as correct. However, if the target contained a final *stressed* syllable which was *shortened*, this was incorrect.

Each subject's *pattern* of acoustic parameter use was tallied for examination of preferred parameters or combinations of parameters. Responses which were perceptually "incorrect" were examined in a similar way. The perceived "error" stress pattern and the acoustic parameters which were used to produce the "error" stress pattern were noted.

These acoustic patterns were tallied and examined for preferences. The possible combinations are listed below.

1. Amplitude, f0, and duration *(All)*
2. Amplitude and f0
3. F0 and duration
4. Duration and Amplitude
5. Duration only.
6. Amplitude only.
7. F0 only.
8. No parameters matched target pattern.

The combinations were then collapsed to reveal how often each of the main acoustic parameters (amplitude, f0, and duration) was utilized, alone or in combination with other parameters. These values were obtained by summing the possible combinations listed above. For example, to obtain the percentage of times a subject employed amplitude, the number of times they used combinations 1, 2, 4, or 6 were totaled. The total was divided by the subject's total number of correct responses to get a percentage of use of amplitude. The process was repeated for f0 (combinations 1, 2, 3 and 7 were totaled), and for duration (combinations 1, 3, 4, and 5 were totaled). Since some combinations were included in two or more parameters (i.e., combination 1 is included in all three), when the percentage of use for f0, amplitude, and duration is added, the result may exceed 100%.

Statistical Analysis of Acoustic Preferences

The number of times each subject *successfully* utilized an acoustic parameter (or combination of parameters) to imitate the target stress pattern was tallied and converted to a percentage of their total correct responses. For "correct" responses, there were no instances of subjects' responses using *no parameters* which matched the target. The percentage of times each subject used the remaining seven possible combinations was arc-sine transformed and then analyzed for significant group differences using a Kruskal-Wallis ANOVA with the Median Test. The *percentage of use* of the three main parameters was also examined using a Kruskal-Wallis ANOVA with the Median Test.

Examination of Error Responses for Expected Error Patterns

"Incorrect" responses were examined to see if they showed any of the errors which previous studies had shown might be expected of young children or children with speech disorders. These expected errors include: deletion of initial unstressed syllables or syllable-addition to form trochaic stress patterns, use of reduced contrast between stressed and unstressed syllables, and reduced range of intonation change.

Intonation Subtest Items

Intonation subtest items were graphed to highlight the information of interest - pitch change. Duration and f0 values from the beginning and end of each syllabic nucleus were plotted to illustrate the amount and direction of pitch change over time. Items were judged to be "correct" or "incorrect" imitations of the target intonation pattern based on visual comparisons between the narrowband spectrograms for each subjects' response and the narrowband spectrogram of the model. Visual inspection of the direction of pitch change on the graphs was used to confirm the judgement of "correct" or "incorrect" for each response. Items which exhibited different patterns of production than the model were further examined to establish which differences were "errors" and which were acceptable patterns of production. The productions of the normal control subjects were used as the standard of acceptable variation. If a subject from a disordered group produced a response with the same variation as a control subject, it was considered acceptable; if the disordered subject's production was unlike both the model and the control subjects', it was considered to be an error.

CHAPTER 3

RESULTS

In the present study, the prosodic abilities of nine kindergarten children were examined to investigate the relationship between segmental and suprasegmental speech disorders. Two speech-disordered groups were matched for severity and each contained three subjects, either with linguistic speech disorders or oral-motor speech disorders. An age-matched group of three of their peers with normally developing speech served as controls. All groups were given the Tennessee Test of Rhythm and Intonation Patterns, T-TRIP (Koike & Asp, 1981), a test of imitative prosody using reduplicated nonsense-syllables as stimuli.

Recordings of subjects' responses were made and used for perceptual scoring of all 25 items and for acoustic analysis of 16 selected items. Three judges scored the T-TRIP according to Koike and Asp's (1981) criteria; the scores of all three judges were averaged for each subtest and for the overall percent correct score of each subject. Acoustic measurements of amplitude, duration, and fundamental frequency (f0) were made with the Computerized Speech Lab (CSL) for ten *rhythm* subtest items and six *intonation* subtest items. Accurate imitation of the target pattern was verified by phonetic transcription for rhythm items and by examination of narrowband spectrograms for intonation items. Acoustic data was examined for patterns of performance by individual subjects and by groups. The acoustic values were graphed to illustrate subjects' use of each parameter to produce rhythm and intonation patterns for selected items.

Results from the Perceptual Scoring of the T-TRIP

Since on-line scoring for each subject was done by one of three different test administrators, three graduate level clinicians re-scored the T-TRIP for all subjects to provide intra-judge reliability. Scores from the three clinicians were averaged to obtain each subject's mean T-TRIP subtest scores and total percentage correct. Figure 2 presents a graph of the total percent correct scores for each subject. Figure 3 contains a graph of the average subtest scores for each subtest. Note that each subtest had a different number of items. The *rhythm* subtest had 14 total items, the *tempo* subtest had three items, and the *intonation* subtest had eight items.

Examination of Figure 2 reveals that the three groups performed differently overall. All subjects in the linguistic group and the control group attained higher total percent correct scores than subjects in the oral-motor group. The range of scores overlapped only slightly for the linguistic group (75% to 93%) and the control group (92% to 99%), while the oral-motor groups' scores (31% to 61%) did not overlap the ranges of either group. As shown in Figure 3, control subjects also scored highest on each subtest, followed by linguistic subjects with the second highest scores, and oral-motor subjects with the lowest scores. Overall there was a continuum of ability levels, with the controls at the high end, the linguistic subjects in the center, and oral-motor subjects at the low-end. This suggests that the T-TRIP is able to differentiate the two speech-disordered groups based on their imitative prosody abilities.

In order to examine group differences on individual subtests, T-TRIP subtest and overall scores for the three subjects in each group were averaged to get mean group scores. These scores were then graphed in Figure 4. Visual inspection of Figure 4 reveals that the groups performed similarly on each subtest. All groups performed best on the *intonation* subtest and the two disordered groups performed most poorly on the *tempo* subtest. Each group's performance on the *rhythm* subtest was similar to its' *total score;* this was not surprising since the rhythm subtest contained more than half of the total items (14 of 25 total items, 56%). Individual performance patterns seen in Figures 2 and 3 are confirmed by examining group mean scores in Figure 4. The control group attained the highest mean score for each subtest and for the overall T-TRIP score. The linguistic group attained the second highest score, and the oral-motor group the lowest score. None of the mean group scores overlapped. The level of significance for between-group differences was examined using a non-parametric test of significance.

Statistical Significance of Perceptual Scores

Due to the small number of subjects in each group, the non-parametric Kruskal-Wallis ANOVA was selected to analyze the significance of the groups' test scores. Results are presented in Table 9. Since each subtest contained a different number of items, these scores were converted to percentage correct scores and arcsine transformed

Table 9. Kruskal-Wallis Results (Group as Independent Variable)

Dependent Variable	**Significance Level**	**Significant at $p < .05$?**
Rhythm Subtest Score	p = .0429	Yes
Tempo Subtest Score	p = .0764	No
Intonation Subtest Score	p = .0429	Yes
Total T-TRIP Score	p = .0429	Yes

to normalize them. The scores were analyzed with *group* as the independent variable and *test score* as the repeated measure. Each subtest score, as well as the overall score, was included in the analysis. The results verified that a significant difference existed among groups on the *rhythm, intonation,* and *total* T-TRIP scores. Since the *tempo* subtest contained only 3 of the 25 items, the lack of a significant difference on this subtest was not surprising.

Acoustic Measurement Results

Each of the 16 T-TRIP items selected for acoustic analysis was graphed to aid in the visualization of patterns. Appendix F contains graphs of each response to items 3 - 12 in the rhythm subtest. Separate graphs were made for each acoustic parameter (amplitude, duration, and f0) measured on this subtest. Graphs of items 20 - 25 in the intonation subtest appear as Appendix G. The degree of change of f0 for each syllable was plotted in semitones for each subjects' response to items on this subtest. For both subtests, the graphs illustrate the acoustic patterns which were *produced* as each subject attempted to imitate the target rhythm or intonation pattern.

"Correct" Rhythm Subtest Items

To determine if subjects' responses to rhythm subtest items were *perceived* as "correct" or "incorrect", the stress patterns were transcribed by three listeners experienced in phonetic transcription. The results of the three transcriptionists were combined to form a consensus. If two out of three judges transcribed the response pattern the same as the target pattern, the response was judged perceptually correct. If one or no judges transcribed the response pattern the same as the target pattern, the item was judged to be perceptually incorrect. Table 10 summarizes the transcription results for each subject.

Sixty-three (63) of the ninety (90) responses were judged as "correct"; the other 27 were judged "incorrect."

Acoustic Parameter "Combinations" Used to Indicate Stress

Rhythm subtest items selected for acoustic analysis, i.e. items 3 - 12, were examined to determine which acoustic parameters (or "combinations" of parameters) were used to imitate the target stress pattern. If a response displayed a difference between stressed and unstressed syllables of more than the *JNDs* for a given acoustic parameter, that parameter was considered to have been used to imitate the stress pattern. For example, as the graph of item 3 in Appendix F illustrates, the target stress pattern for this item was *SuS* (or "stressed-unstressed-stressed"). Subject B-3 produced the correct stress pattern for this item only with the parameter of f0; the amplitude and durations of the unstressed syllable were higher than the initial (stressed) syllable and so were not representative of the target stress pattern. Therefore subject B-3 was considered to have used only the parameter of f0 to indicate the stress pattern for item 3. In many cases "combinations" of two or three parameters were used to indicate the stress pattern. These "combinations" were totaled for all items which had been judged correct through transcription, and are included as Table 11, which shows the number of times each combination was used by individual subjects. The overall frequency with which each combination was used by all subjects for items judged as correct is presented as a pie chart in Figure 5.

Table 11. Acoustic Parameters used to Indicate Stress in Items 3 - 12

Subjects	**A-1**	**A-2**	**A-3**		**B-1**	**B-2**	**B-3**		**C-1**	**C-2**	**C-3**		***Total***
Amplitude, f0, and duration	4	2	2		2	1	5		3	4	6		*29*
Amplitude and f0	0	0	0		0	0	0		2	1	0		*3*
Duration and f0	2	0	3		0	0	0		2	2	3		*12*
Amplitude and duration	1	0	1		0	1	0		0	0	0		*3*
Only duration	1	5	1		1	0	1		2	1	1		*13*
Only amplitude	0	0	0		0	0	0		0	1	0		*1*
Only f0	1	0	0		0	0	1		0	0	0		*2*
Total Correct Items (of 10)	*9*	*7*	*7*		*3*	*2*	*7*		*9*	*9*	*10*		*63*

All three parameters were combined to indicate stress nearly half (45% - or 29 of 63 responses) of the time. Duration was used as the sole indicator of the stress pattern in 21% (13 of 63 items). By contrast, the other parameters were rarely used alone. Amplitude alone indicated the stress pattern 2% of the time (1 of 63), and f0 alone 3% (2 of 62) of the time. The two parameters most used in combination were duration and f0. They were paired to indicate the stress pattern in 19% (12 of 63) of the "correct" responses. The other combinations - duration and amplitude, and amplitude and f0 - were each only used 5% of the time (3 of 63).

Figure 6, the frequency of parameter use by each group, presents a slightly different picture. The control group and the linguistic group were similar in the proportion of time they used each parameter or combination of parameters. Both groups indicated stress most frequently by combining all three parameters (46% for the control

group, 35% for the linguistic group). The control group and the linguistic group also demonstrated flexibility by utilizing other parameters and combinations of parameters numerous times. Each group combined f0 and duration approximately 25% of the time (25% for the control group, 22% for the linguistic group). The linguistic group used duration alone 30% of the time, while the control group relied solely on this parameter less often (14%). Other combinations were utilized only minimally. For example, the control group combined amplitude and f0 11% of the time, corresponding to three of the 28 correct responses by this group. The linguistic group had 23 correct responses overall to rhythm subtest items, and used amplitude and duration paired in only two of those responses (4%).

The oral-motor group used fewer acoustic parameter combinations to indicate stress. However, since this group responded correctly only 12 times of 30 opportunities (i.e. 10 items x 3 oral-motor subjects = 30 opportunities), they admittedly had less opportunity to do so. The oral-motor group used the combination of all three parameters in the majority of their correct responses (67% - 8 of 12). Duration alone was used 14% of the time (2 of 12), while f0 and the combination of amplitude and amplitude were each used 4% of the time (1 of 12).

Statistical Analysis of Acoustic Parameter Combinations

Each group's use of acoustic parameter combinations for correct responses were converted to percentages and arc-sine transformed for analysis of significant differences. The Median Test was chosen as a non-parametric test permitting group comparisons. Results are presented in Table 12. The only parameter combination to vary significantly among groups was the combination of all three parameters (amplitude, f0, and duration).

Table 12. Median Test Results for Group Differences in Acoustic Parameter Use.

Acoustic Parameters Used to Indicate Stress	Significance Level	Significant at .05% level?
Amplitude, f0, and duration	***p* = .0429**	**Yes**
Amplitude and f0	*p* = .0764	No
Duration and f0	*p* = .5258	No
Amplitude and duration	*p* = .2231	No
Only duration	*p* = 1.000	No
Only amplitude	*p* = .3247	No
Only f0	*p* = .5258	No

The boxplot in Figure 7 illustrates that the linguistic group varied from the oral-motor group, but not from the control group in their use of the combination of all three parameters. The control group's median was different than the oral-motor group; however, since their ranges overlap individual scores could result in misidentification despite the difference in medians.

Figure 8 presents the frequency of use of acoustic parameter combinations by individual subjects. Comparisons between subjects were complicated by the fact that subjects had different numbers of correct responses (ranging from 2 correct for subject B-2, to all 10 correct for subject C-3). This means that subjects with fewer correct responses had fewer opportunities to select acoustic parameter combinations, i.e. if a subject had only 3 correct responses, the maximum number of combinations they could use was three. By contrast, subjects with many correct responses had more opportunities to use different combinations, and therefore could demonstrate more variety in their production of stress. Figure 9 illustrates the variety of combinations utilized to mark

Figure 7. Boxplot of Group Differences in the Use of All Acoustic Parameters Combined to Mark Stress.

stress. This variety was evident even within-subjects. Control subjects used 3 to 5 different methods to convey stress, linguistic subjects used 2 to 5, while oral-motor subjects used only 2 or 3 methods. Since Figure 8 presents only those responses which were transcribed as having the correct stress pattern, it might be misinterpreted as indicating that control and linguistic subjects demonstrated greater flexibility in choosing acoustic parameters to convey stress. However, it is important to remember than oral-motor subjects had fewer correct responses (B-1 had 3, while B-2 had only 2). When the "incorrect" responses were included also, as in Figure 9, it can be seen that most subjects used a variety of combinations to indicates stress.

Figure 9 presents *all* responses - both those which were transcribed with the target stress pattern and those which were transcribed with other stress patterns. For "incorrect" responses, acoustic parameters (or combinations of parameters) which reflected the *transcribed* stress pattern (rather than the target stress pattern) were considered to have been utilized. Figure 9 demonstrates that most subjects utilized at least four different combinations. Two subjects (A-2 and B-1) employed only the two most popular combinations - all three acoustic parameters, and duration alone - to indicate stress. The fact that most subjects used a variety of combinations of acoustic parameters to indicate stress was expected, based on the findings of other researchers (Kehoe et al., 1995; Schwartz et al., 1996; Konefal & Fokes, 1985; Pollack et al., 1993).

Preferences for Individual Acoustic Parameters

Subjects' *preferences* for using specific acoustic parameters (i.e., amplitude, duration, or f0) to imitate the target stress pattern are presented in Figure 10. Konefal and Fokes (1985) defined preferences as the use of a specific parameter 60% or more of the

time to produce stress (Konefal & Fokes, 1985). Konefal & Fokes studied contrastive stress (not rhythmical patterns) in children with normally developing speech aged two to six; their oldest group was comparable in age to the group in the present study (n=10; mean C.A. 5.4 yrs). They found that 90% of the oldest group and 70% of the 3- and 4-year-old group exhibited preferences for at least two parameters (i.e. f0 was used 60% or more, and amplitude was used 60% or more, but not necessarily in combination), while only 20% of the youngest group showed preferences for two or more parameters. Figure 10 shows that this was true for all but one subject in the present study (8 of 9 subjects, 89%). For this linguistic subject, A-2, only duration was employed 60% of the time.

Generally, duration was the most frequently used acoustic parameter, followed by f0; duration was used in 90% of all correct responses to mark stress, while f0 was used in 73% of all correct responses. Overall, amplitude was only used in 57% of correct responses. In only two cases was amplitude used as often as another parameter. One subject (B-2) used amplitude and duration equally as often; the other subject (B-1) used amplitude and f0 equally as often. These two subjects (B-1 and B-2) were both in the oral-motor group and had the lowest two scores on the rhythm subtest (3 and 2 correct responses, respectively). The fact that amplitude is used less often by normal subjects may indicate that it is a less-robust parameter; in any case, the choice to rely on a parameter which is not often chosen by normal speakers to indicate stress may have negatively affected the accuracy with which oral-motor subjects' responses were perceived. Figure 11 presents group averages for each parameter. The general preference for duration first, followed by f0, is reflected in these group averages; the oral-motor

group, on average, used duration more than the other parameters, but used f0 and amplitude equally as often.

Statistical Analysis of Preferences for Individual Acoustic Parameters

The Median Test was used to test for significant group differences in the use of individual acoustic parameters. The only significant difference was in the use of *amplitude* ($p = 0.0429$). The use of f0 and duration did not differ significantly among groups ($p = 0.5258$ and $p = 1.000$, respectively). The boxplot for the use of amplitude is shown in Figure 12. This demonstrates that the oral-motor group's use of amplitude varied from the linguistic group and control group. The control group's use of amplitude was also different from the linguistic group. This means that the oral-motor group subjects employed amplitude, the least popular parameter for indicating stress, significantly more than the other two groups.

A Closer Examination of "Incorrect" Items

During the *perceptual* scoring of the T-TRIP, items which did not have the correct number of syllables were marked as incorrect per Koike and Asp's (1981) criteria and not analyzed further. In contrast, during the *acoustic analysis* of the 16 selected items, all syllables were analyzed and the number of syllables per item was noted. Table 13 lists the number of added or deleted syllables by subject for each of the 16 T-TRIP items selected for acoustical analysis. Among the 27 "incorrect" responses, 16 were incorrect due to syllable omission and five were incorrect due to syllable addition.

Table 13. Total Number of Syllables Added or Deleted by Subject.

Subjects	**A-1**	**A-2**	**A-3**	**B-1**	**B-2**	**B-3**	**C-1**	**C-2**	**C-3**
Added Syllables	0	1	3	0	0	0	0	1	0
Deleted Syllables	0	1	0	12	12	5	1	0	0

Figure 12. Boxplot of Group Differences in the Use of Amplitude (Alone or in Combination) to Mark Stress.

Syllable Deletion

Syllable deletion was a hallmark of the oral-motor group; as the length and complexity of items increased, this group deleted more syllables in test items. In fact, this group was responsible for 14 of 16 responses with omitted syllables. Oral-motor subjects deleted a total of 29 syllables, while control subjects deleted only one syllable per group. The oral-motor group deleted syllables in 11% of the 3-syllable items (items 3, 4, and 5), in 47% of the 4-syllable items (items 6, 7, 8, 9 and 10), and in 83% of the 6-syllable items (items 11 and 12).

Deletion occurred most frequently for items with initial unstressed syllables. Fifty-six percent (56%) of items with initial unstressed syllables in the target stress pattern had syllables deleted by the oral-motor group, compared to only 30% of items with initial stressed syllables in the target pattern.

Conversion of Iambic Stress to Trochaic Stress

Iambic patterns were converted to trochaic patterns in nine of the "incorrect" responses through syllable deletion, eight times by oral-motor subjects, one time by a control subject, and zero times by linguistic subjects. This probably reflects Allen and Hawkins (1980, p. 240) findings regarding the prevalence of trochaic rather than iambic stress patterns and the tendency for initial unstressed syllables to be deleted by children with developing or disordered speech skills. For example, subject B-1 deleted the first syllable of item 3, thus changing the iambic target stress pattern "uSu" to the simpler trochaic pattern "Su." Subject B-2 changed five of the six iambic patterns to trochaic through syllable deletion (items 4, 7, 8, 9, and 12). However, this subject (B-2) also produced a limited variety of stress patterns, responding with the pattern "Suu" on six of the ten rhythm subtest items. This may indicate that subject B-2 did not perceive the different stress patterns, or that he could not adjust his production to accommodate the complex patterns and therefore simplified all items to the “Suu” response. The graphs of these *Suu* responses were examined for evidence of variability in how these identically perceived items were produced, but no distinct patterns emerged.

One other case of syllable deletion by a subject *not* in the oral-motor group occurred on an item with an initial unstressed syllable (item 12). Acoustic measurement revealed a sixth syllable on this response, which was transcribed by all three judges as having only five syllables. The sixth syllable’s vocalic nucleus was a mere 30 ms in duration, and so may have actually been a burst as the /m/ which closed syllable five was released. Since all judges agreed, only five syllables were counted here, and the response was considered to exemplify initial syllable deletion.

Syllable Addition

Linguistic subjects added a total of four syllables, control subjects added one syllable, and oral-motor subjects added zero syllables. In three cases, syllable addition was due to one subject’s transformation of stimuli to lexical items. The subject (A-3) added an initial unstressed syllable (the lexical item “my”), but preserved the target stress pattern for two of these items (3 “my mom, my mom” and 6 “my mom, my maw-maw”). The subject’s response to item 6 was perceived to have the same stress pattern as his response to item 7; however the “words” used in each response differed - “my mom, my maw-maw” vs. “my ma, my momma”. This raised the question of whether his perception of item 7's stress pattern was identical with his perception of item 6's stress pattern, or whether he varied the lexical content of his responses to reflect the different stress patterns. Since item 7 contains a low-frequency stress-pattern (i.e., beginning with an iambic pair of syllables, followed by a trochaic pair), perhaps the subjects simply wasn't able to create a lexical item with item 7's stress pattern (uSSu). The other two cases of syllable addition (C-2, Item 7; A-2, Item 12) were due to disfluencies. For example, subject C-2 "stuttered" on Item 7, adding a brief unstressed syllable between the two stressed syllables. No disfluencies were produced by an oral-motor subject.

Lexicalization of Nonsense Stimuli

The hallmark of the linguistic group was *lexicalization* of the nonsense stimuli. *Lexicalization* was defined as the use of real words, for example "my", "mom", "momma", "maw-maw", "mommy" etc. Table 14 lists the use of lexical items by each subject. A total of 17 items were "lexicalized" on the rhythm subtest, equivalent to 18.9% of the 90 total responses. (By contrast, the only two cases of "lexicalization" on the intonation subtest were in response to items 18 and 19, which were not selected for acoustic analysis.) Linguistic subjects lexicalized 14 of the 17 responses (82%), oral-motor subjects lexicalized one (6%), and control subjects lexicalized two (12%). The tendency for lexicalization shown by the linguistic group may have the same cause as their segmental articulation errors: an incorrect conceptual model for the target.

"Incorrectly" Perceived Items with "Correctly" Produced Stress Patterns

Items which had been judged as "incorrect" imitations of the target stress pattern were analyzed according to the perceived (transcribed) stress pattern. Many of the "incorrect" responses had syllables deleted or added (21 of 27 responses), as discussed above. Of the six remaining error responses containing the proper number of syllables, four actually demonstrated the *correct* stress pattern with one or more parameters. This finding indicates that the subjects produced the correct stress pattern with at least one acoustic parameter, but the response was perceived as having a different stress pattern. Two reasons for these “misperceptions” were identified: a lack of acoustic contract, or a conflict between the acoustic parameters. Table 15 contains four items which demonstrate these two patterns.

Table 15. Four “Incorrect” Responses Produced with “Correct” Stress Pattern in one or more Acoustic Parameters.

Subject	**Item**	**Target**	**Acoustic parameter(s) used to produce target stress pattern**
A-1	6	SuSu	Duration (except prolongation of final syllable; 4>3 by 2 ms)
A-2	10	uSuS	f0 (1 > 2 by only .56 ST; 3 > 4 by only .16 ST) (Duration produced as *uSuu*; Response transcribed as *uSuu.*)
B-3	7	uSSu	f0 (1 < 2 by only .71 ST) Amp (1<2; 2<3;3>4 but 4>2. Correct as syllable pairs, not as whole item) Dur (except prolongation of final syllable; 3 < 4 (119 ms < 122 ms))
B-3	8	uuSu	Dur (stress pattern produced correctly; 3 > 1, 2, & 4 by 140 ms) Amp & f0 both produced as SuSu.

Acoustic Contrasts Less Than ‘Just Noticeable Differences’

In two cases (A-2 Item 10; B-3 Item 7), the subjects did not produce large enough contrasts between stressed and unstressed syllables to exceed the *just noticeable differences (JNDs).* *JNDs* were defined as 1 ST, 1 dB, or 10 ms difference between adjacent stressed and unstressed syllables (Hermes & van Gestel, 1991; Lehiste, 1970). For example, subject A-2 produced item 10 with the correct *uSuS* stress pattern using the parameter of f0. However the second (stressed) syllable exceeded the first (unstressed) syllable by only 0.56 semitones (ST); likewise the fourth syllable exceeded the third by only 0.16 ST. The subject’s response was transcribed as *uSuu,* which was the stress pattern strongly manifested with the duration parameter. In general, when acoustic parameters produced conflicting stress patterns, the pattern produced with contrast exceeding the JNDs was perceived by the transcription judges. This applied even to responses which were correctly produced with duration except for pre-pausal lengthening of the final syllable. This was shown in subjects A-1 and B-3's responses in Table 15. In these two cases, the final syllable was supposed to be unstressed but the response exhibited pre-pausal lengthening of the final syllable, resulting in an equal or longer duration than the item's stressed syllables. Since the other acoustic parameters of these responses had "stronger" (albeit incorrect) stress patterns, or correct stress patterns produced without *JNDs*, the items were transcribed as incorrect.

The Case of Conflicting Information

In the fourth case (B-3 Item 8), the target pattern (uuSu) was correctly produced with duration values for all syllables, however amplitude and f0 values presented an incorrect pattern (SuSu). This was an interesting case of conflicting acoustic information. The conflict was reflected in the disagreement among the three transcription judges. The first judge transcribed *uSSu* - which did not agree with the pattern of any acoustic parameter and may have represented a compromise. The second judge transcribed the response as *SuSu* - the pattern presented by the combination of amplitude and f0. The third judge transcribed the response as *SuSu*, but assigned secondary stress to the 1st (unstressed) syllable; this was accepted as equivalent to *uuSu* (the correct pattern). The other two judges did not distinguish between primary and secondary stress for this item. This may have been because most items had two (equally) stressed syllables, or because they were perceiving the two stressed syllables presented with the combination of amplitude and f0. Despite a duration difference in excess of 140 ms between the stressed syllable and the three unstressed, the combination of two parameters presenting the incorrect pattern (amplitude and f0) won out over the single parameter presenting the correct pattern (duration). Although not the focus of the current study, a more-in depth perceptual analysis might reveal a hierarchy of parameters cases of conflicting information like this one. As far as the present study is concerned, the important finding was that some disordered subjects produced conflicting stress patterns among the three acoustic parameters, which were perceived by judges as incorrect or unreliable stress patterns.

Acoustic Results of Intonation Subtest

Although overall performance on the intonation subtest was good for all groups, acoustic analysis of items 20 - 25 provided insight into the variety of acoustic patterns subjects produced while attempting to imitate intonation contours. For each of the 6 intonation items selected for acoustic analysis, the subjects' second attempt was digitized and measured to determine the duration of each syllable's vocalic nucleus, and the degree and direction (or directions) of pitch change within each syllable's vocalic nucleus. These data were graphed to aid in identification of intonation production patterns. Graphs of every subject's response to each item appear as Appendix G. Intra-judge reliability was less than 45% for the intonation items; partly due to the novelty of the task and partly due to inadequate instructions given to the judges. Therefore, comparisons between the narrowband spectrogram of each subject's response and the narrowband spectrogram of the pre-recorded model's production of each item were used to confirm each response's "acoustic correctness" instead of transcription results. Appendix E contains spectrograms of the model' productions of items 20 - 25; Appendix H contains sample narrowband spectrograms of subjects' responses to items 20 - 25 in cases where the acoustic production was questioned. Table 16 summarizes the comparison between the subjects' and model's spectrograms. Only five responses were clearly production errors, either because of syllable deletion (B-3, Items 24 and 25) or because the f0 of an entire syllable moved in the opposite direction from the model (B-3, Item 20; A-1, Item 24; B-2, Item 24); these were labeled as "ERRORS" in Table 16. However, 17 other responses differed slightly from the model's acoustic pattern and could not be labeled as "exact" imitations of the model; these were labeled as "Questionable" (?) In Table 16.

These "questionable" responses were categorized into four main types:

1. Intonation contours which reverse direction *within* a single syllable from rising pitch to falling pitch.

2. Intonation contours which reverse direction *within* a single syllable from falling pitch to rising pitch.

3. Intonation contours which should be level, according to the model, but appear to have a minimal fall.

4. Intonation contours which should rise or fall, according to the model, but appear to be level, or have only a minimal rise or fall.

In order to decide which of these differences were "errors" and which were acceptable variations in production, the responses of the control subjects were examined. Since the three re-scoring judges were unanimously agreed that these responses were correct imitation of the target, the control subjects' productions were accepted as correct, regardless of any variation in acoustic patterns.

Two of the differences in production were thus considered to be within normal limits, regardless of whether they were produced by the control group or the disordered groups. The first ("Intonation contours which reverse direction *within* a single syllable from rising pitch to falling pitch") pattern can be seen in 9 responses, including 3 responses by control subjects. For example, the narrowband spectrogram in Appendix H shows that subject A-3 produced a slight rise in pitch before the expected fall in pitch on the third syllable of Item 24, as did subject B-1 on the fourth syllable of Item 25. This seems to reflect the need to "get set" at an appropriate f0 before beginning the intonation change; in other words by first raising their f0, the subject creates more of a contrast during the pitch fall than would have been possible by beginning from their habitual f0. This "get set" phenomenon occurred in syllables in the 1st, 2nd, and 3rd syllables of an item. It was not always obvious by visual inspection of the graphs, since the graphs included only the vocalic nucleus of each syllable. Items which displayed a pitch rise, then fall during a syllable which should have had falling pitch according to the model were considered to be "correct" productions.

However, the same was not true of the second (and opposite) type of acoustic production difference. Since no control subjects demonstrated this pattern (a *fall* in pitch before a *rise* in pitch during a syllable which should have had *rising* pitch) it was not considered to be an acceptable variation in production. Only one subject produced this pattern, B-1. This subject changed the stimulus "ma ma" to "ma lump" for both items 20 and 21; narrowband spectrograms of these items are included in Appendix H. Although the final syllable of both responses does reflect a noticeable pitch change in the correct direction, the first syllable for both responses demonstrates the same pattern (fall-rise). For item 21, this subject may have been lowering their f0 prior to beginning the expected rise in pitch, “getting set” but in the opposite direction. Since no control subject exhibited this pattern, there is no precedent for acceptance of this variation. For item 20, the “getting set” was clearly in the wrong direction, changing from fall to rise rather than from rise to fall as other subjects demonstrated. The substitution of “lump” for the second syllable in both items was an additional clue that this subject was not imitating the model.

The third difference in production was “intonation contours which should be level, according to the model, but appear to have a minimal fall.” This pattern was demonstrated 4 times, twice by control subjects. Appendix H contains an example of this pattern by subject C-3 in Item 23. This minimal fall pattern was considered to be the effects of reduced amplitude at the end of a syllable, causing a corresponding *declination* in f0. Items which demonstrated this pattern were accepted as correct imitations of the model.

The fourth pattern was not demonstrated by control subjects, and so was considered an unacceptable (i.e., “incorrect”) variation in production. The pattern was “intonation contours which should rise or fall, according to the model, but appear to be level, or have only a minimal rise or fall.” Only one subject demonstrated this pattern (B-2) on three items (20, 21, and 23). Appendix H contains narrowband spectrograms of subject B-2's productions of these items, and also item 22 for comparison. Visual inspection of the spectrograms suggested the use of a restricted pitch range, which might not have exceeded the *just noticeable differences.* The graphs of these items provide further illustration of the “flat” slopes. In order to verify that the pitch change was enough to be perceptible, the amount of pitch rise or fall in semitones (ST) was compared to the *JND* needed for pitch (1 ST). F0 range values are listed in Table 17 for each of B-

2's responses which contained the correct number of syllables.

Table 17. Amount of Pitch Change in Items Judged to Have “Flat” Slopes.

Item	Syllable 1	Syllable 2	Correct ?
20	1.82 ST (Fall)	.92 (Rise)	No, syllable 2<JND
21	2.17 ST (Rise)	1.58 ST (Fall)	Yes, both > JND
22	1.61 ST (Fall)	2.26 ST (Fall)	Yes, both > JND
23	1.91 ST (Fall)*	4.39 (Rise)	Yes, both > JND

*For Item 23, the first syllable should be *level* or *rise*. This *fall* is most likely declination.

Overall, performance on the 6 intonation items selected for acoustic analysis was noticeably better than performance on the 10 items analyzed from the rhythm subtest. A total of 27 responses were judged as errors on the rhythm subtest (30%), while only 8 responses were judged as errors on the intonation subtest (15%).

Summary of Results

Whether examined perceptually or acoustically, the T-TRIP results showed significant differences among subjects with linguistically-related speech disorders, motorically-related speech disorders, and normally developing speech. Perceptual scores provided a general overview of subjects' ability to imitate prosody. The combination of transcription and acoustic analysis revealed more in-depth information about error patterns. In a sense, the T-TRIP perceptual scores equated to a screening, while transcription and acoustic analysis of the recorded responses served as a more descriptive assessment tool.

Perceptual scoring followed Koike and Asp's (1981) criteria, which required only a "yes/no" judgement about whether the subject's response correctly imitated the model's pattern of stress, tempo, or intonation, and contained the correct number of syllables. Even this judgement appeared difficult for T-TRIP administrators to make "on-line". The scores awarded on-line were uniformly lower than the scores made from the tape recorded responses. The discrepancy between on-line scores and scores based on the tape recorded responses raises the question of how reliable on-line judgements were. In contrast, judgements made by three graduate-level clinicians during re-scoring had good inter-judge reliability (90%).

The subtest scores from all three re-scoring judges were averaged and used to determine each subject's subtest and overall T-TRIP scores. The groups were shown to be significantly different in their performance on the rhythm and intonation subtests, and on the overall percentage correct score. More specifically, graphic representations of the groups' scores show that oral-motor subjects’ scores were clearly separated from the other two groups. Linguistic subjects slightly overlapped the score ranges of the control subjects. This overlap implies that the prosody of linguistic subjects is more like normal subjects' prosody than like oral-motor subjects' prosody. Therefore, the T-TRIP has the potential to be used clinically to differentiate between children with oral-motor speech disorders and children with linguistic speech disorders.

The results of the transcription and the acoustic analysis basically confirmed the perceptual findings regarding significant differences among the groups in both rhythm and intonation subtest performance. For the rhythm subtest, the transcriptions provided confirmation that the responses were perceived as having the correct stress pattern. For the intonation subtest, poor inter-judge reliability (45%) on transcription results clearly demonstrated the difficulty of the unfamiliar task of transcribing intonation contours. Therefore, instead of using transcription to judge each response’s *perceptual* “correctness”, acoustic analyses of the model’s and the control subjects’ production were used to judge the *acoustic* "correctness" of intonation subtest items.

Several important error patterns emerged from the detailed examination of rhythm items: linguistic subjects tended to add syllables and to lexicalize items, while oral-motor subjects tended to delete syllables and to convert iambic stress into trochaic. These error patterns would not have been revealed by perceptual scoring alone, since perceptual scoring only produces judgements of responses as correct or incorrect with no explanation of *why* responses were incorrect.

Acoustic measurements permitted an even more detailed look at production patterns of error responses and correct responses alike. Examination of correct productions revealed that subjects used a wide selection of acoustic parameter combinations to indicate stress. This was in keeping with the findings of other researchers regarding the variety of methods for producing stress, even within-subjects (Lehiste, 1970; Pollack et al, 1993; Kehoe et al, 1995) Oral-motor subjects employed the combination of all three acoustic parameters significantly more often than other groups did to indicate stress. Likewise the oral-motor group used the parameter of amplitude (by itself or combined with other parameters) significantly more often than other groups. However, the fact that oral-motor subjects had fewer correct responses than either linguistic or control subjects also meant that they had fewer opportunities to demonstrate their preferences for acoustic parameter use. When incorrect responses were examined to determine which acoustic parameters which had been used to create the *perceived* stress pattern, it became apparent that oral-motor subjects used the same variety of parameters that the other groups did.

Intonation items also exhibited a wide variety of production styles, even by control subjects. Most of these variations appeared to be within normal limits. However, in one particular case, an oral-motor subject produced the correct intonation pattern but did not produce enough pitch change to exceed the *JND* for f0. This also occurred for several rhythm items; in these responses, the contrast between stressed and unstressed syllables did not exceed the *JNDs* for one or more parameters. Other rhythm subtest items were produced with contradictory stress patterns among the acoustic parameters. Items like these, which did not exceed the *JNDs* or contained contradictory stress patterns, were considered to be “incorrect” imitations of the model. The present study did not lend itself to a detailed assessment of the perception of these types of error patterns. The relationship between the use of a particular acoustic pattern and the perception of stress or intonation is an area that could be investigated more thoroughly in the future.

Other responses contained patterns that were just plain "wrong," as when a subject produced a rhythm item with the wrong syllables stressed, or an intonation item with the contour in the wrong direction. These errors were always associated with the disordered speech subjects, and usually with the oral-motor group. Despite the small number of subjects included in the present study, definite differences were evident. This suggests that more extensive research incorporating acoustic analysis and transcription should be undertaken to examine the error patterns of a larger number of responses.

Overall, whether T-TRIP responses were examined by perceptual or acoustic methods, the oral-motor group's imitative prosody ability was significantly different that the other groups'. The clinical implications of this finding are that the T-TRIP has the potential to be used as a screening tool to identify subjects whose imitative prosody abilities difficulties are consistent with oral-motor speech disorders, specifically DVD. Several findings, namely the difficulty of transcribing intonation items and the questionable reliability for on-line scoring of all subtests, suggest that modifications to the design and scoring of the T-TRIP may be warranted.

CHAPTER 4

DISCUSSION

Review of Research Questions

This study utilized the T-TRIP to determine if kindergartners with linguistic speech disorders, oral-motor speech disorders, or normal speech performed differently on imitative prosody tasks (Koike & Asp, 1980). Performance was assessed perceptually with overall and subtest scores on the entire T-TRIP, and acoustically with measurement of individual prosodic variables (f0, amplitude, and duration) on selected T-TRIP items from the rhythm and intonation subtests. The answers to each of the four questions will be reviewed in turn.

Question 1: Group Differences in T-TRIP *Overall Percentage Scores*

With regard to these nine subjects specifically, overall T-TRIP percentage correct scores *did* show a clear separation between the three groups. The separation was clear graphically (see Figure 2), and was shown to be statistically significant at the $p < .05$ level. Based on the limited number of subjects, it is difficult to generalize the results of this study to a larger population of children with disordered speech. Next, group subtest scores were examined to focus on individual aspects of disturbed prosody (i.e., rhythm, tempo, intonation) which may be used to differentiate the groups.

Question 2: Group Differences in T-TRIP *Subtest Scores*

Statistical comparisons of group scores on all three subtests found that *rhythm* and *intonation* subtest scores were significantly different across all groups at the $p < .05$ level, while the *tempo* subtest was not. The disordered groups performed most poorly on the *tempo* subtest and best on the *intonation* subtest, while all groups scored about the same on the *rhythm* subtest as they did on the overall T-TRIP. However, the unequal number of items in each of the T-TRIP subtests made comparisons between them problematic; recall that the *rhythm* subtest contained 14 items, the intonation subtest contained 8 items, and the *tempo* subtest contained only 3 items. Also, the five practice items were all for rhythm subtest items, so there was no practice given for tempo or intonation items. Due to the limited number of tempo subtest items presented with no practice, it seems premature to conclude that this subtest has no potential for differentiating between the groups. However, the present study found no significant group differences for this subtest. Examination of selected rhythm and intonation subtest items through acoustic means provided further information regarding *how* group and individual performance differed on these subtests.

Question 3: Group Differences in Production of Rhythm

Results of acoustic analyses agree with previous studies in that all subjects exhibited variety in their use of acoustic parameters to indicate stressed syllables. For selected rhythm subtest items, the groups were compared on the number and type of acoustic parameter combinations used to imitate stress patterns. Oral-motor subjects were significantly different from linguistic and control subjects in their reliance on the combination of all three acoustic parameters (amplitude, duration, and f0) to mark stress. The oral-motor group also used the parameter of *amplitude* to mark stress significantly more than the other two groups. This cue was shown by other researchers to be infrequently used to indicate stress in normal populations during conversational tasks, presumably because it is less robust than duration or f0 (Crystal, 1969; Lehiste, 1970; Klatt, 1976).

Both disordered speech groups were more likely than control subjects to use an insufficient range of separation (i.e. acoustic contrasts less than the *JNDs)* between stressed and unstressed syllables. This may point to perceptual or developmental difficulties. However, these two groups also performed differently in other respects, not measured acoustically. The oral-motor group demonstrated more syllable deletion, and less syllable addition, lexicalization, or disfluencies than the other two groups. The linguistic group demonstrated more syllable addition through disfluency and lexicalization than the other two groups. These distinct error patterns support the theory that the disordered subjects were sampled from populations with breakdowns at different levels of the speech production process.

Question 4: Group Differences in Production of Intonation

Performance on the intonation subtest was hard to assess. Acoustic results were not able to be cross-validated through transcription as they were on the rhythm subtest, due to the poor reliability obtained between the two trained phoneticians. Instead, visual comparisons between the narrowband spectrograph of each subject's response and the narrowband spectrograph of the pre-recorded model's production of each item were used to confirm each response's "acoustic correctness." As on the rhythm subtest, subjects exhibited variation in production of intonation contours. Obvious errors included syllable deletion and reversing the pitch contour of syllables. However, other slight variations in production were noted which were not necessarily errors. Since control subjects were perceived to have correctly imitated all intonation items during perceptual scoring, these variations were evaluated for "correctness" based on their use by control subjects.

Patterns which were demonstrated by the control subjects were accepted as within normal limits when produced by any subject, regardless of diagnostic group. These patterns included falling pitch in syllables that were produced with level pitch by the model (declination), and a rise-fall pitch contour in syllables which were given falling pitch by the model ("getting set"). Patterns which were not exhibited by control subjects were considered to be "incorrect" imitations of the model. These included fall-rise pitch contours in syllables which were given falling or rising pitch by the model, and level or nearly level pitch in syllables which clearly had rising or falling pitch in the model. Both "incorrect" patterns were exhibited by oral-motor subjects.

For one subject in the oral-motor group, the magnitude of the pitch contours (slopes) was small. This subjects was the same one who was used an insufficient range of separation (i.e. less than the *Just Noticeable Differences)* between stressed and unstressed syllables in the rhythm subtest. Additionally, he did not increase f0 without an accompanying increase in amplitude during rhythm subtest items, suggesting that he was either not able to perceive or produce f0 changes. Overall performance on the intonation subtest was good, with the few errors mainly found in the oral-motor group.

Current Results Compared to Expected Results

The results of the present study can be compared to previous studies focusing on the use of stress and intonation by normal and speech disordered children. In general, the findings of the present study support the findings of other researchers regarding stress patterns, but were inconclusive regarding intonation patterns. The results of this study may also help to answer questions posed by other researchers and explain seemingly contradictory findings.

"Normal" Acoustic Correlates of Stress

The current study complements the findings of other studies which investigated the acoustic correlates of stress in normal children. Like Pollock et al. (1993) and Kehoe et al. (1995), the present study found that children often combine all three acoustic parameters to mark stress. Oral-motor, linguistic, and control subjects used this combination an average of 45% of the time to mark stress in correct responses. Duration was the most popular single parameter for marking stress in the present study (90% of correct responses). Similarly, Pollock et al.'s found that all three age groups (2-, 3-, and 4-year-olds) used duration to differentiate between stressed and unstressed syllables. Subjects in the present study also demonstrated variability in their choice of acoustic parameters to mark stress in correct items, rather than using the same pattern of parameters in all responses. This variability in the use of acoustic parameters to mark stress has been found in previous studies with children (Pollock et al., Kehoe et al.) and with adults (Lehiste, 1970).

All subjects in the present study - normal and disordered - marked stress in the expected way for responses which were judged to be correct. However, the context of this study (imitated reduplicated nonsense syllables in a formal assessment task) was quite different from the context of previous studies (contrastive stress in novel 2-syllable word pairs, or monosyllabic and disyllabic word-pairs elicited during structured play) (Pollock et al., 1993; Kehoe et al., 1995). Since the subjects in the present study produced stress in the expected manner, despite the different context, the validity of using this type of task (i.e., the T-TRIP's imitated reduplicated nonsense syllables) for assessment was increased.

Disordered Subjects Exhibit Error Patterns Similar to Young Subjects

Error patterns observed in disordered subjects in the current study were consistent with error patterns demonstrated by younger subjects in previous studies. For example, Kehoe et al. (1995) found that items which were judged by listeners to have "equal" or "level" stress had less contrast between stressed and unstressed syllables when measured acoustically than items which had been judged to have "reliable" stress. This parallels the current finding of perceptually "incorrect" responses which acoustically demonstrated the "correct" stress pattern, but with contrasts between stressed and unstressed syllables less than the required JNDs. The current study also found conflicting stress patterns across the three acoustic parameters in at least one "incorrect" response. Pollock et al., (1993) found similar conflicts between f0 and amplitude to be associated with "unreliable" listener transcriptions. Since syllable deletion errors tended to occur on initial unstressed syllables, Pollock et al. associated this phenomenon with a "trochaic bias" by young subjects. The present study also found that deleted syllables were associated with initial unstressed syllables, although most cases of syllable deletion were found in disordered subjects.

The key difference between the current study and previous studies was the *age* of the children making these errors: the children studied previously were much younger (2-year-olds in Pollock et al. (1993); 18- to 30-month-olds in Kehoe et al. (1995)) than the 5-year-olds in the current study. In young children, these errors would be considered developmental, and attributed in part to the inexperienced speakers' need to simplify productions to match the capabilities of their immature systems. Pollock et al. suggested that 2-year-olds "have the production ability to use the acoustic parameters to mark stress, even though they often stress the wrong syllable" and that difficulties with perception may have played a central role in errors produced by young, normal speakers. Likewise, the errors of the subjects in the current study may actually be developmental in nature. However, the similarities between errors produced by younger subjects' and disordered subjects' does not necessarily indicate that the cause of the errors is the same. To the contrary, since the current study found unique error patterns to be associated with each disordered group, different causes for each groups' errors are implied.

In any case, more research into the hypothesized role of perception in errors of stress production is clearly needed. This includes investigations of whether subjects perceive differences in stress patterns, and also investigations of which acoustic parameters listeners attend to when they identify stress patterns. Pollock et al (1993) stated that listeners appeared to rely more on f0 and amplitude than on duration, despite the fact that 2-year-olds exhibited poor control of these parameters while all subjects exhibited reliable use of duration to indicate stress. Transcriptions of items in the current study were examined for evidence of increased sensitivity to certain parameters, but no clear patterns emerged. Research which systematically examines the effects of reduced contrasts between stressed and unstressed syllables, and of contradictory stress patterns presented by different acoustic parameters on listener's perception of stress is needed. It is possible that the "clues" which subjects are providing about stress patterns are "clues" to which listeners are not sensitive.

Problems Encountered in Interpreting the Intonation Findings

Overall, the T-TRIP's intonation subtest did not lend itself to comparisons with other studies. First, the task itself was particularly unnatural. As mentioned before, English is not a "tone language" and so does not specify semantic or phonological meaning through the use of word-level contrasts in tone (Lehiste, 1970). Instead, English contains sentence-level intonation contrasts which specify the syntactic function of utterances. Since T-TRIP intonation items are one, two or four syllables in length, they are approximately the size of one or two words at the most, rather than a whole utterance. The T-TRIP thus appears to be testing the ability of subjects to imitate word-level intonation contrasts; since word-level intonation is not phonemic in English, this calls the content validity of the intonation subtest into question.

The use of an unnatural task also negatively affected the ease of scoring, whether transcription or on-line "correct/incorrect" judgements were used. Interestingly, the subjects generally were able to *imitate* the intonation patterns, while the judges reported difficulty *transcribing* the intonation patterns. This was another indication that English rarely requires close attention to contrasts between adjacent intonation contours on such a small scale.

Previous studies of children's productions of intonation patterns suggested that rising contours were more difficult for both young normal subjects (Loeb & Allen, 1993) and hearing impaired subjects of all ages (Most & Yael, 1994). However, Crary and Tallman (1993) did not find significant differences between the production of rising contours and falling contours by speech disordered children. In the present study, no distinct patterns of difficulty with a particular intonation contour emerged, aside from "getting set" which occurred primarily on *falling* contours. Other error-patterns identified by Crary and Tallman included the use of longer response durations and more restricted f0 range by disordered subjects. In the present study, response durations were not remarkably different for normal and disordered subjects, and only one subject demonstrated the use of restricted f0 range. This lack of agreement between findings in the present study and previous studies does not undermine the previous studies' results; rather it suggests that the tasks were simply not comparable. The intonation subtest of the T-TRIP featured only eight items (six of which were selected for acoustic analysis), each with very different intonation patterns, and no practice items. In future studies, the design of this subtest might well need to be changed to include items which more closely reflect the way intonation contours are used in English.

In Support of Inappropriate Stress as a Diagnostic Marker for DVD

In general, the results of the current study support Shriberg et al.'s (1997b, 1997c) hypothesis that inappropriate stress is a diagnostic marker for oral-motor speech disorders, such as DVD. Shriberg et al. used the PVSP (Shriberg et al, 1990) to assess multiple aspects of prosody and voice, namely *phrasing, stress (lexical/phrasal), rate, loudness, pitch, laryngeal quality* and *laryngeal resonance* during conversational tasks (1997b; 1997c). Subjects with suspected DVD were found to have a significantly increased occurrence of *inappropriate lexical or phrasal stress* when compared to their peers diagnosed with speech delay (Shriberg et al., 1997b; Shriberg et al., 1997c). Recall that speech delay, or SD, is roughly equivalent to phonological disorders, or in the terms of the current study "linguistically-based speech disorders." So, like the current study, Shriberg et al. (1997b, 1997c) found that oral-motor subjects had significantly more errors in stress production than linguistic subjects.

The present study also found that oral-motor subjects scored significantly lower on the T-TRIP intonation subtest when responses were assessed perceptually. The PVSP does not include assessment of intonation specifically; intonation is encompassed by the variables of *pitch, phrasing,* and *laryngeal quality* instead. No significant findings emerged with any of those variables (Shriberg et al., 1997b, 1997c), so this may contradict the results of the current study. However, as mentioned above, the T-TRIP's intonation items were difficult to compare with other studies in general because of their unnaturalness. The discrepancy between the T-TRIP results and the PVSP results is probably a further example of how unlike conversational speech the T-TRIP tasks are.

The T-TRIP *Revisited - Shadden et al.(1980) vs. Henry (1990)*

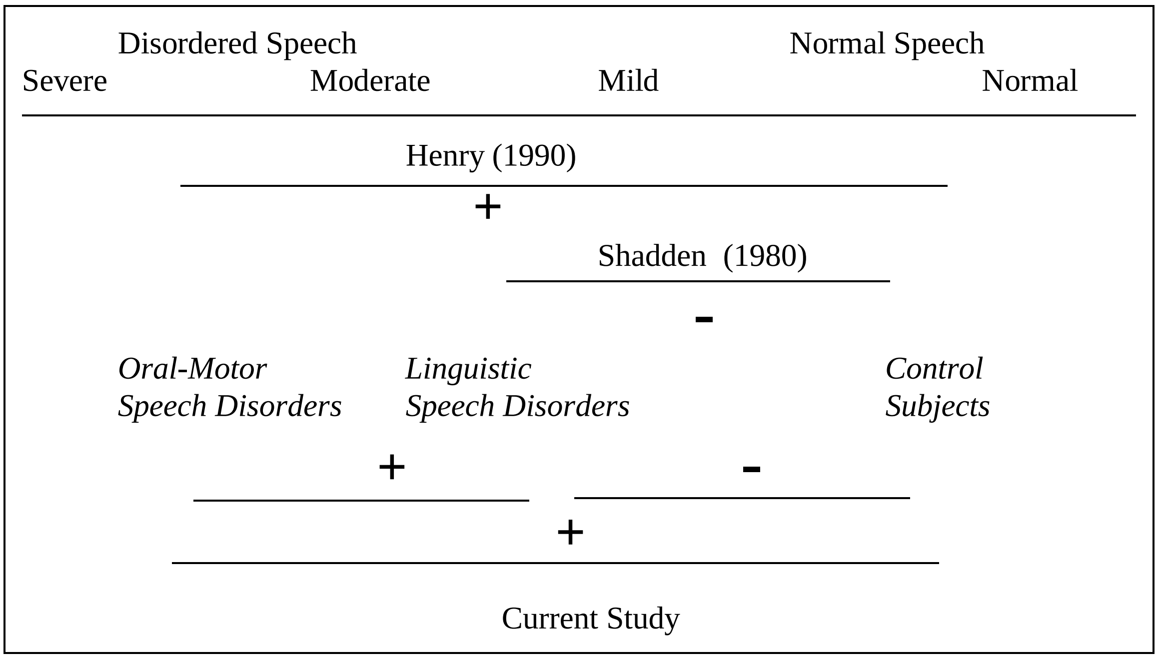
Two previous studies which used the T-TRIP to investigate prosody in children with speech disorders resulted in contradictory findings. Shadden et al (1980) compared the performance of 5-year-old children "with adequate and inadequate speech articulation" on all 25 T-TRIP items. Shadden et al. failed to find significant group differences in either subtest or overall scores, and suggested uncontrolled length of time in therapy as one possible explanation. Henry (1990) compared the performance of children aged 3- to 5-years-old with severe articulation disorders and with normal articulation on the T-TRIP's first two subtests - Rhythm and Tempo. She found that disordered subjects in all age groups scored significantly less well than normal children, and that rhythm items were more difficult than tempo items. Henry suggested that the discrepancy between her results and Shadden et al.'s might be due to more severe speech disorders in her subjects.

The current study sought to resolve this contradiction by investigating a third hypothesis: that the two studies had sampled different populations, both with moderate to severe speech disorders, but caused by breakdowns at different levels of speech production planning. Recall that according to Levelt (1989), segmental aspects of speech are generated separately from suprasegmental aspects of speech. Developmental research suggests that segmental and suprasegmental aspects of speech develop at different times and rates (Crystal, 1979). Therefore, breakdowns at different levels of speech production may result in the same effect on *segmental* aspects of speech (i.e. unintelligibility), but produce different effects on *suprasegmental* aspects of speech, creating a continuum of prosodic abilities along which subjects with speech disorders and normal subjects are distributed.

In order to test this hypothesis, two groups of speech disordered children were included in the present study. Disordered subjects were matched for age and severity of speech disorder, then categorized as having an 'oral-motor' speech disorder or 'linguistic' speech disorder. As expected, the 'linguistic' group performed similarly to the control group of subjects, while the 'oral-motor' group performance was different from that of the other two groups. This result suggests that Shadden et al. (1980) and Henry (1990) did indeed test different populations - Shadden et al.'s subjects may have been representative of the linguistic disordered group, while Henry's may have been oral-motor subjects. Figure 13 demonstrates how the results of the current study explain Shadden et al. (1980) and Henry (1990)'s seemingly contradictory results.

The current study may also contain evidence of the effect of differing amounts and types of therapy, as suggested by Shadden et al. (1980). Shadden et al. (1980) theorized that several characteristics of therapy may facilitate better performance during assessment of suprasegmental aspects of speech. For example, as a result of the therapeutic training process, subjects may gain experience at listening to and reproducing speech stimuli, especially speech with exaggerated intonation and stress patterns which focuses attention on target sounds (Shadden et al., 1980). This may introduce a practice-effect which could positively influence the subjects' ability to imitate both segmental and suprasegmental aspects of speech, as in the T-TRIP.

Figure 13. Continuum of Prosodic Abilities for Speech Disordered and Control Subjects.
+ = *significant between group differences*; - = *no significant between group differences*



For example, the oral-motor subject with the highest rhythm subtest score was the only subject who was not a member of the kindergarten class from which the other subjects were drawn. This subject had completed several semesters of speech therapy at the University of South Florida's Communication Disorders Clinic. His score may have been higher than that of other oral-motor group members because of the uncontrolled variables of 'length of therapy' or 'type of therapy.' He may have participated in therapy tasks which differed from those of the other subjects, and which gave him an edge at imitation tasks. Interestingly, this subject also had the worst score on the Limb and Oral Praxis Test, indicating that imitative limb and oral praxis ability and imitative prosodic ability did not correlate in his case.

Current Results and Levelt's Model of Speech Production

The error-patterns displayed by linguistic subjects (syllable addition) and oral-motor subjects (syllable deletion) may be explained in terms of Levelt's model of speech production. Linguistic subjects made errors on the segmental aspects of rhythm subtest items, while oral-motor subjects made errors on the metrical aspects. This suggests that the two groups have breakdowns within different processing components. Figure 14 shows Levelt's model of speech production and the proposed sites for breakdowns.

Linguistic subjects added syllables to four rhythm subtest items - three times because of lexicalization and one time because of a disfluency. Eleven other items were lexicalized by the linguistic group, but without syllable additions. These eleven items retained their target stress patterns, but contained altered segmental patterns. For example, Item 4 was changed from "ma MA ma" to "my MOMma". The altered segmental content points to selection of the incorrect phonemes during segmental spellout. Segmental spellout fills the empty frames generated by the morphological and metrical spellout procedures with the phonemic content of the word (Levelt, 1989). In the case of lexicalization errors, the "wrong" phonemic content was selected. This could have been caused by a breakdown in the segmental spellout procedure itself, or by a breakdown at an earlier stage if the surface structure which formed the input for the segmental spellout was faulty. Since the T-TRIP tests imitation, not spontaneous speech, this surface structure must be generated by the subject's perception of the target, bypassing the first level of speech production - the Conceptualizer. Regardless of whether the breakdown was in the

formulation of surface structure or in the segmental spellout procedures, the fact that the metrical pattern was retained for these 11 items shows that linguistic subjects had no breakdowns within the Prosody Generator.

In the three cases where lexicalization resulted in syllable addition, the additional syllable disrupted the stress pattern of the item. For example, item 3 “MA ma MA” was lexicalized to “my MOM, my MOM.” This error indicates that an extra frame was generated during the segmental and metrical spellout procedure. Since the surface structure specifies the morphemes to be used, this strengthens the theory of faulty surface structure leading to a breakdown at the level of the metrical spellout.

The two cases of syllable addition resulting from disfluency are more difficult to explain. Levelt does not address the issue of disfluencies (1989). These syllable repetitions occurred once for a control subject and once for a linguistic subject. They did not appear to be associated with lexicalization of the target. Again, a syllable addition error indicates that an extra frame was generated in the surface structure. Although the cause of such an addition cannot be determined, the addition occurred *before* the metrical frames were fed into the Prosody Generator, indicating that the subjects had no breakdowns within the Prosody Generator itself.

Conversely, oral-motor subjects had specific difficulties with the metrical aspects of the stimuli. The only case of lexicalization (a segmental “error”) by a subject from this group was also accompanied by an alteration of the target stress pattern (a metrical “error”). Within the Prosody Generator, the phonological structure of a phrase is reorganized to make coarticulation easier, and then “syllabified” into pronounceable syllables. The “free” parameters for each successive syllable (duration, stress, pitch, and inter-syllabic pauses) are also set by the Prosody Generator (Levelt, 1989). The oral-motor group's tendency to delete syllables may have resulted from breakdowns during syllabification (in the Prosody Generator), or earlier, if too few address frames for syllables were created during metrical spellout procedures. As mentioned above, faulty perception or faulty surface structure could have also been responsible for deleted syllables. These error-sites cannot be ruled-out. However, the oral-motor group's other errors generally involved aspects of speech programmed by the Prosody Generator. Examples include incorrect stress patterns, insufficient contrasts in pitch, amplitude or duration between successive syllables, and conflicts in the stress patterns exhibited by different acoustic parameters. The fact that these errors were all on aspects of speech handled by the Prosody Generator suggests that breakdowns occurred at that level. This fact also undermines the possibility that the breakdown occurs *later* in the process, i.e. during retrieval of articulatory programs from the syllabary. Admittedly these conclusions are based on a limited number of errors in a small group of oral-motor subjects and therefore require more investigation. Nevertheless, the Prosody Generator is certainly implicated as the site of errors for this group.

Although Levelt's model of speech production was based on speech errors by normal adult speakers and not disordered children, it helps to pinpoint possible breakdown sites for both segmental and suprasegmental errors. In the case of linguistic subjects, it is probable that a faulty "frame" was input into the intact Prosody Generator. This resulted in correct stress patterns but "incorrect" phonemic patterns. For oral-motor subjects, the opposite most likely occurred. The target segmental information emerged intact, but the faulty Prosody Generator degraded the metrical content of the target. The results of the current study strengthen Levelt's hypothesis that segmental and suprasegmental aspects of speech are produced separately.

Other Factors Possibly Influencing Results

The findings of the present study, both acoustic and perceptual, may have been influenced by differences in the voice/resonance characteristics or recording quality of subjects in the disordered speech groups. Two subjects, both in the oral-motor group, had distinct voice/resonance quality differences which were apparent during the T-TRIP and may have negatively biased scoring. The female subject in the oral-motor group had a harsh/hoarse quality to her voice; it is not known whether this was her typical voice quality or was the temporary result of an illness. A second oral-motor subject had a hyponasal resonance. These differences may have negatively influenced their perceptual scores; specifically, "disordered voices" may have unconsciously been associated with "disordered speech" by listeners, resulting in lowered performance expectations. By adding *normal voice/resonance quality* to the inclusion criteria, this extraneous variable could be eliminated. However, some subjects may have both voice disorders and oral-motor speech disorders, and some may have temporary reductions in voice quality due to illnesses or allergies. To allow inclusion of these important subjects, future studies could include subjects with normal speech but disordered voices as a way to more closely examine the effect of this variable.

Different noise levels on the recordings used for re-scoring, transcription, and acoustic analysis may have been another source of bias. Since these recordings were made "in the field" at an elementary school, control of recording conditions was not ideal and varied from day to day. Three recordings had higher noise floors than the others. The recording of subject A-2 from the linguistic group had a noise floor measured at approximately 63.71 dB, while the recording of oral-motor subjects B-1 and B-2 had noise floors measuring 49.91 dB and 51.0 dB. By comparison, the recordings of other subjects had noise floors between 45 and 48 dB. The lowest noise levels were on the recordings of control subjects, who were all tested the same day. The poor noise-to-signal ratio created by higher noise levels may have reduced the apparent contrast between stressed and unstressed syllables, and reduced the apparent range of intonation changes, both perceptually and acoustically. This may have negatively influenced the scores of the disordered subjects. In future studies, care should be taken to minimize noise differences by recording all subjects on the same day, or more closely monitoring the noise levels in the background.

An alternative explanation for the phenomenon of lexicalization may be the reduction of unstressed syllable vowels (i.e. to the vowel schwa) as a segmental means of marking stress (Lehiste, 1970). However, if that were the case then vowel change would have occurred on unstressed syllables only. No such pattern emerged to support this explanation. In this study, vowel reduction was considered to be a segmental means of marking stress and so was not investigated systematically. Future studies may incorporate measurement of vowel formants to further examine the use of vowel reduction as a means of marking unstressed syllables.

Improving the Assessment Instrument

The results of this preliminary investigation suggest that the assessment of imitative prosody in non-meaningful contexts may aid in the differential diagnosis of oral-motor speech disorders, such as DVD. However, the current version of the <u>T-TRIP</u> is not an ideal tool for assessment of prosodic skills, either perceptually or acoustically. Several areas for improvement were identified.

The design of the T-TRIP would be strengthened by adding more *tempo* and *intonation* items so that all subtests had an equal number of items for ease of comparison. Practice *tempo* and *intonation* items should also be added before the respective subtests to aid subjects in changing behavioral sets. Another improvement would be the use of an alerting tone or other means to signal the conclusion of the model, and to discourage subjects from responding in the middle of the model.

In the present study, reliability of acoustic measurement was compromised due to the difficulty in establishing a cut-off between the continuant phoneme /m/ and the following vowel. The use of a stop-plosive instead of the nasal phoneme /m/ would provide a readily identifiable acoustic landmark (i.e. the burst) for measurement of syllable duration. The bilabial voiced plosive /b/ is one possible alternative. The use of "ma ma" as the stimulus also prompted some subjects to change the nonsense syllables into real words, for example "Mommy, maw maw, my mom etc." However, other segmental combinations may also result in lexicalization.. For example, subjects may use the target syllable as a familiar term for a relatives (i.e. grandparents are often affectionately called 'Baba' while parents may be known as 'Mama', 'Papa', or 'Dada' etc.). Still other subjects may be members of cultures (i.e. Mandarin, among others) where the reduplicated syllable pattern "mama" has derogatory connotations, depending on its intonation contour. Accordingly, it may be necessary to have multiple alternatives available.

Scoring reliability would be increased through systematic training of examiners and perceptual judges. Prosody is a difficult aspect of communication to assess, in part due to its paralinguistic nature, and in part due to the lack of formal training in the area. As this study showed, even experienced listeners with training in phonetic transcription encounter difficulty in the transcription of rhythm and intonation patterns. The poor inter-judge reliability obtained between on-line judges and re-scoring judges testifies to the fact that 'familiarity' with the T-TRIP's administration does not guarantee the ability to score subjects' responses reliably. Not surprisingly, the demands of simultaneously administering the test, dealing with equipment and subjects, and scoring responses increase the difficulty of the task and decrease the reliability of scoring. One way to increase scoring reliability would be to use training tapes, such as those used by Shriberg et al. (1990) to train transcriptionists on the PVSP. Inter- and intra-judge reliability could be established on recorded samples from normal and disordered subjects before on-line scoring was attempted. Routinely audio-taping T-TRIP administration would enable verification of on-line scores, while video-taping would allow examination of non-verbal indications of stress or intonation, i.e. a subject may raise her eyebrows or nod her head each time she stresses a syllable. These audio and video tapes would in turn increase the availability of training and research materials.

Directions for Future Research

Clearly, in order to generalize the results of this study, more subjects need to be tested. Replication studies would also benefit from including assessments of subjects' prosody in different contexts. Assessment of prosody in conversation in the same population would add to the knowledge of whether the T-TRIP results are truly representative of the prosodic deficits that subjects demonstrate in more natural speech.

Contrastive stress drills, featuring either lexical or phrasal stress, may present a middle-ground between the contrived imitative tasks and difficult-to-control conversational tasks.

An assessment of *subjects'* ability to perceive differences in prosody would also add valuable information to future studies. Computerized routines could by adapted to present T-TRIP stimuli in pairs for subjects to judge as 'same/different', or in groups of three for 'odd one out' judgements. Research into perceptual skills may indicate that perception and production abilities correlate in linguistic subjects, but not in oral-motor subjects. Or no relationship may be identified, suggesting that the errors are truly production errors. A third possibility is that subjects are *producing* correct patterns, but in such a fashion that listeners are not *perceiving* them. More research into *listener's* perception of reduced contrasts between stressed and unstressed syllables, and of contradictory stress patterns presented by different acoustic parameters is needed before this possibility can be ruled out. Pollock et al. (1993) suggested that synthesized- or altered-speech be used to further investigate which acoustic parameters listeners attend to when they identify stress patterns.

Investigations of treatment effects would provide a more-clinical direction for future research. Shadden et al. (1980) suggested incorporating two groups of disordered subjects, one before treatment, and one after, to test for treatment effects. A study incorporating repeated-measures with the same group might also yield information about the T-TRIP's sensitivity to small changes in prosody as a result of treatment.

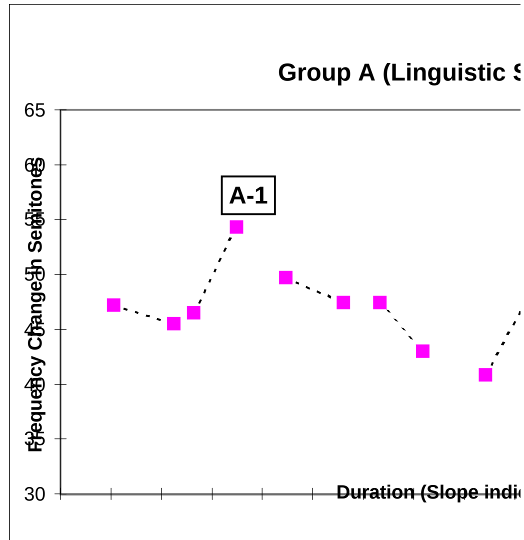
Clinical Implications

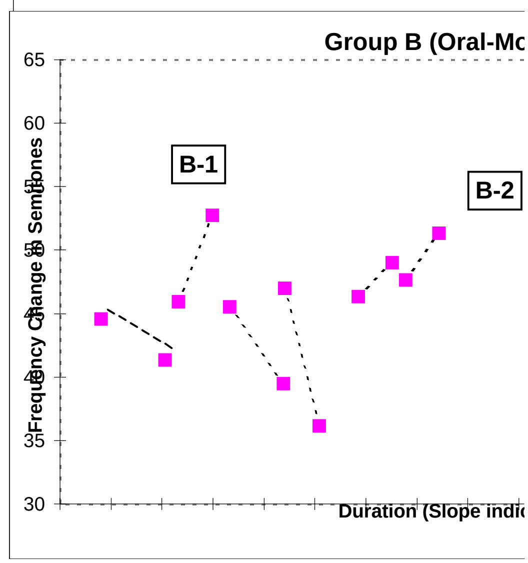
The results of this study have clinical implications both for the assessment of prosody disorders, and for their treatment. With adequate training to establish reliability, perceptual T-TRIP scores may be used by clinicians to identify prosody disorders in the areas of rhythm, or intonation. (Since performance on the current tempo subtest was not shown to be significantly different among groups, the results of this subtest may not be a valid reflection of prosodic ability.) The effect of any identified prosody disorder on intelligibility should of course be established with more natural tasks, i.e. conversational samples.

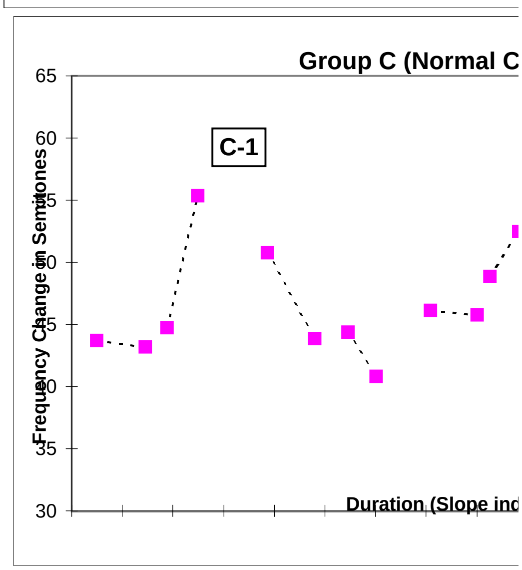
Clinicians may administer the T-TRIP to aid in the differential diagnosis of oral-motor speech disorders, such as DVD. The current study found that performance on the T-TRIP differentiated two groups of subjects, both with moderate to severe speech disorders. Since the group with motorically-based speech disorders scored significantly lower than the group with linguistically-based speech disorders, the T-TRIP thus appears to have potential to aid in differential diagnoses of speech disorders as oral-motor or linguistic in nature. Exact cut-off scores have not been established. It is important to remember that this study is a preliminary one involving a limited number of subjects and therefore generalizations cannot be made without further testing to verify the results of the current study.

However, close examination of the error patterns of speech disordered subjects on T-TRIP rhythm and intonation items may provide further support for a provisional diagnosis or oral-motor speech disorder or linguistic speech disorder. For example, linguistic subjects in the current study lexicalized the stimuli to a much greater extent than other groups, while oral-motor subjects deleted syllables (in particular initial unstressed syllables) more often. These error patterns are not captured by the "correct/incorrect" judgements recommended by Koike and Asp (1981). Detailed transcription of the segmental and suprasegmental patterns is required to obtain this information. Once specific error patterns have been identified, they may be used to establish treatment goals and targeted during treatment activities.

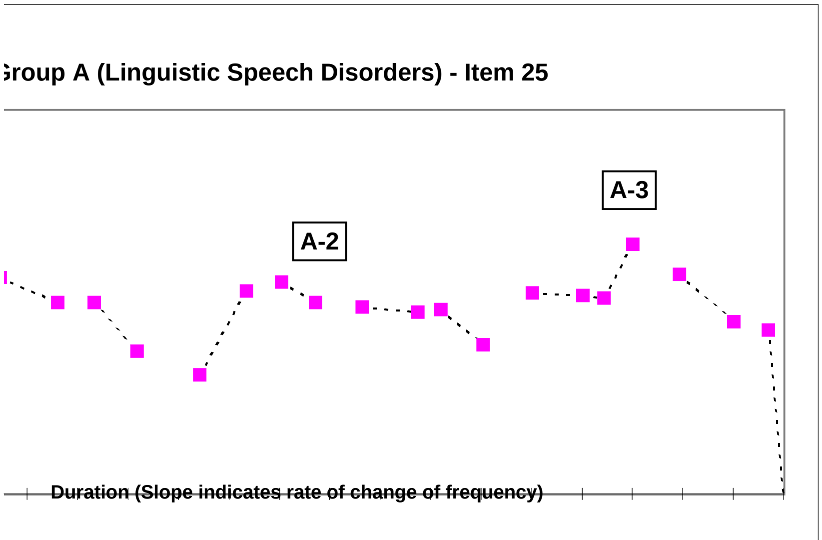
25 dur	25 Hz	25 dur	25 st
262	250	262	47.21
561	227	561	45.54
659	240	659	46.51
873	377	873	54.33
1115	289	1115	49.72
1400	253	1400	47.42
1581	253	1581	47.42
1795	196	1795	43
2106	173	2106	40.84
2336	269	2336	48.48
2510	282	2510	49.3
2680	253	2680	47.42
2911	247	2911	47.01
3185	241	3185	46.58
3300	244	3300	46.79
3510	203	3510	43.61
3755	266	3755	48.29
4003	263	4003	48.09
4110	259	4110	47.83
4253	344	4253	52.74
4484	294	4484	50.02
4754	229	4754	45.7
4925	219	4925	44.92
4996	92	4996	29.91
202	215	202	44.6
			#NUM!
			#NUM!
514	178	514	41.33
582	232	582	45.92
746	344	746	52.74
832	227	832	45.54
1095	160	1095	39.49
1102	247	1102	47.01
1271	132	1271	36.16
1463	238	1463	46.36
1630	277	1630	48.99
1695	256	1695	47.63
1857	317	1857	51.33
2399	285	2399	49.48
2581	263	2581	48.09
2664	259	2664	47.83
2783	227	2783	45.54
3118	168	3118	40.33
3632	290	3632	49.78
3848	230	3848	45.77
4227	156	4227	39.05
123	204	123	43.69
362	198	362	43.18
471	217	471	44.76
623	400	623	55.35
967	307	967	50.77
1200	206	1200	43.86
1364	212	1364	44.36

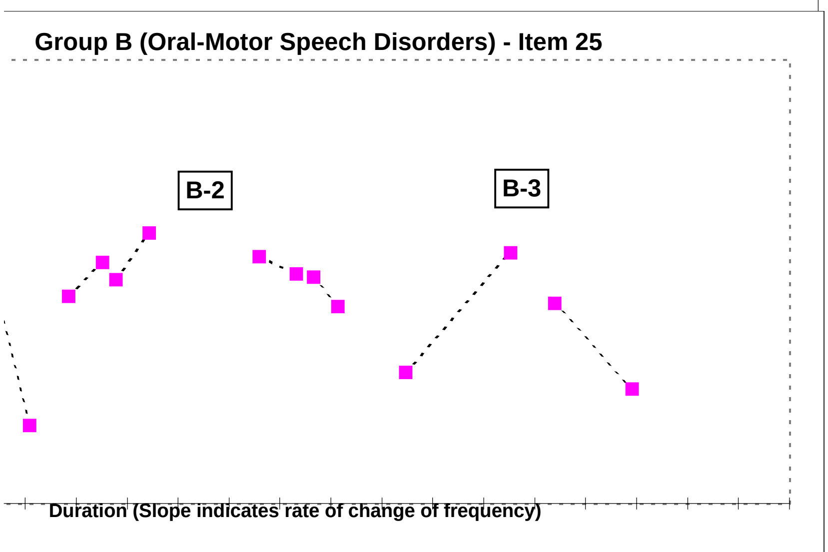


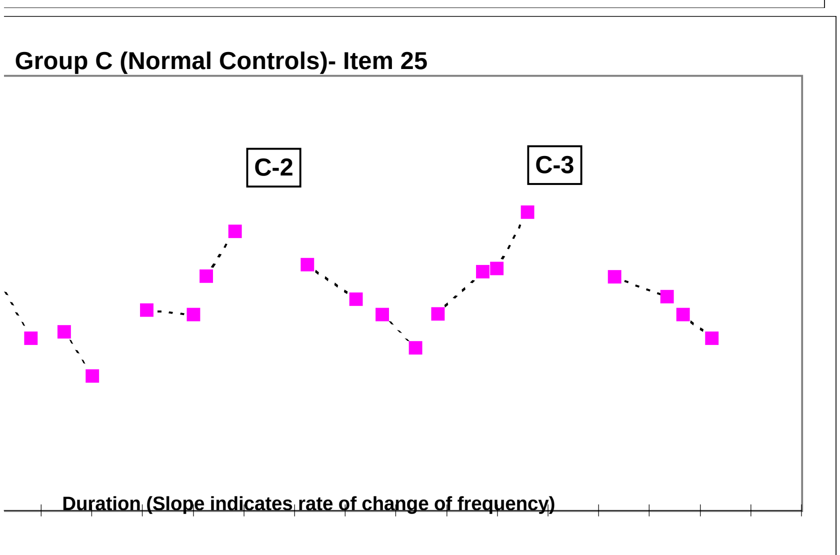




1501	173	1501	40.84
1772	235	1772	46.14
2001	230	2001	45.77
2064	275	2064	48.86
2207	339	2207	52.49
2563	290	2563	49.78
2804	247	2804	47.01
2932	230	2932	45.77
3098	197	3098	43.09
3207	231	3207	45.85
3429	281	3429	49.24
3496	285	3496	49.48
3648	370	3648	54
4078	274	4078	48.8
4336	250	4336	47.21
4416	230	4416	45.77
4557	206	4557	43.86


roup A (Linguistic Speech Disorders) - Item 25
A-2
A-3
Duration (Slope indicates rate of change of frequency)


Group B (Oral-Motor Speech Disorders) - Item 25
B-2
B-3
Duration (Slope indicates rate of change of frequency)


Group C (Normal Controls)- Item 25
C-2
C-3
Duration (Slope indicates rate of change of frequency)

REFERENCES

Allen, G.D. (1983). Linguistic experience modifies lexical stress perception. Journal of Child Language, 10, 535-549.

Allen, G. D., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In G.H. Yeni-Komshian, J.F. Kavanagh, & C. A. Ferguson (Eds.), Child Phonology: Volume 1- Production (pp.227-256). New York: Academic Press.

Baken, R. J. (1987). Clinical Measurement of Speech and Voice. Boston: Allyn & Bacon.

Baltaxe, C. A. M., & Simmons III, J. Q. (1985). Prosodic development in normal and autistic children. In E. Schopler, & G. B. Mesibov (Eds.), Communication Problems in Autism (pp.95-125). New York: Plenum Press.

Blakely, R. (1980). Screening Test for Developmental Apraxia of Speech. Tigard, OR: CC Publications.

Crary, M.A. (1993). Developmental Motor Speech Disorders. San Diego: Singular Publishing Group, Inc.

Crary, M. A., & Tallman, V. L. (1993). Production of linguistic prosody by normal and speech disordered children. Journal of Communication Disorders, 26, 245-262.

Crystal, D. (1969). Prosodic systems and intonation in English. Cambridge: University Press.

Crystal, D. (1979). Prosodic development. In P. Fletcher, & M. Garman (Eds.), Language Acquisition (pp. 33-48). Cambridge: University Press.

Crystal, D. (1982). Profiling linguistic disability. London: Edward Arnold.

Dewey, D., Roy, E.A., Square-Storer, P.A., & Hayden, D. (1988). Limb and oral praxic abilities of children with verbal sequencing deficits. Developmental Medicine and Child Neurology, 30, 743-751.

Edwards, M. (1973). Developmental verbal dyspraxia. British Journal of Disorders of Communication, 8 (1), 64-70.

Furrow, D., Podrouzek, W., & Moore, C. (1990). The acoustic analysis of children's use of prosody in assertive and directive contexts. First Language., 10 (28: 1), 37-49.

Haak, N. J., & Darling, R. M. (1993). Imitation of suprasegmental patterns by non-native speakers of English. American Journal of Speech-Language Pathology, 2, 47-50.

Hargrove, P. M., & McGarr, N. S. (1994). Prosody management of communication disorders. San Diego: Singular Publishing Group, Inc.

Hargrove, P. M., Roetzel, K., & Hoodin, R. B. (1989). Modifying the prosody of a language-impaired child. Language, Speech and Hearing Services in the Schools, 20 (3), 245-258.

Hermes, D. J., & van Gestel, J. C. (1991). The frequency scale of speech intonation. Journal of the Acoustical Society of America, 90, 97-102.

Henry, C. E. (1990). The development of oral diadochokinesia and non-linguistic rhythmic skills in normal and speech-disordered young children. Clinical Linguistics and Phonetics, 49 (2), 121-137.

Hodge, M. M. (1994). Assessment of children with developmental apraxia of speech: A rationale. Clinics in Communication Disorders, 4 (2), 91-101.

Hodson, B.W. (1986). The Assessment of Phonological Processes-Revised. Danville, IL: Interstate Printers & Publishers, Inc.

Hoffman, P. R., Schuckers, G. H., & Daniloff, R. G. (1989). Children's phonetic Disorders: Theory and treatment. Boston: Little, Brown and Company.

Hornby, P. A., & Hass, W.A. (1970). Use of contrastive stress by preschool children. Journal of Speech and Hearing Research, 13, 387-394.

Kehoe, M., Stoel-Gammon, C., & Buder, E.H. (1995). Acoustic correlates of stress in young children's speech. Journal of Speech and Hearing Research, 38, 338-350.

Kent, R. D., & Rosenbek, J.C. (1983). Acoustic patterns of apraxia of speech. Journal of Speech and Hearing Research, 26, 231-249.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America, 59, 1208-1221.

Koike, K. J., & Asp, C. W. (1981). Tennessee test of rhythm and intonation patterns. Journal of Speech and Hearing Disorders, 46, 81-86.

Konefal, J. A., & Fokes, J. (1985). Acoustic analysis of children's conversational repairs. Journal of Phonetics, 13 (1), 97-105.

Lee, G. (1971). Northwestern Syntax Screening Test. Chicago: Northwestern University Press.

Lehiste, I. (1970). Suprasegmentals. Cambridge, MA: MIT Press.

Lenneberg, E. (1967). Biological foundations of language. New York: Wiley.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. Cambridge: M.I.T. Press.

Levelt, W. J. M. (1993). Timing in speech production with special reference to word form encoding. Annals of New York Academy of Science, 682, 283-295.

Levelt, W. J. M. (1994). The skill of speaking. In P. Bertelson, P. Eelen, & G. D'Ydewalle (Eds.), International perspectives on psychological science: Leading themes. Hove (UK): Lawrence Erlbaum Associates.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? Cognition, 50, 239-269.

Lieberman, P. (1966). Intonation, perception, and language. Cambridge: M.I.T. Press.

Loeb, D., & Allen, G. D. (1993). Preschoolers' imitation of intonation contours. Journal of Speech & Hearing Research, 36, 4-13.

Love, R. J. (1992). Childhood Motor Speech Disability. NY: Macmillan Publishing Company.

Menyuk, P., & Bernholtz, N. (1969). Prosodic features of children's language. Quarterly progress report of research laboratory of electronics, 93, 216-219.

Most, T., & Yael, F. (1994). The effects of age and hearing loss on tasks of perception and production of intonation. Volta Review, 96, 137-149.

Murphy, A. L., McGarr, N., & Bell-Berti, F, (1990). Acoustic analysis of stress contrasts produced by hearing-impaired children. Volta Review, 92, 80-91.

Pindzola, R. (1987). VAP: A voice assessment protocol for children and adults. Tulsa, OK: Modern Education.

Pollock, K. E., Brammer, D. M., & Hageman, C. F. (1993). An acoustic analysis of young children's production of word stress. Journal of Phonetics, 21 (3), 193-203.

Prins, D., Hubbard, C. P., & Krause, M. (1991). Syllabic stress and the occurrence of stuttering. Journal of Speech and Hearing Research, 34, 1011-1016.

Rosenbek, J. C., & Wertz, R. T. (1972). A review of 50 cases of developmental apraxia of speech. Language, Speech, and Hearing Services in the Schools, 3, 23-33.

Schwartz, R. G., Petinou, K., Goffman, L., Lazowski, G., & Cartusciello, C. (1996). Young children's production of syllable stress: An acoustic analysis. Journal of the Acoustical Society of America, 99, 3192-3200.

Scukanec, G. P., & Watson, M. M. (1995). Early prosodic development: A case study. Infant-Toddler Intervention: The Transdisciplinary Journal, 5 (2), 131-140.

Semel, E., Wiig, E. H., & Kwiatowski, J. (1987). Clinical Evaluation of Language Fundamentals-Preschool. San Antonio, TX: Psychological Corp.

Shadden, B. B., Asp, C. W., Tonkovich, J. D., & Mason, D. (1980). Imitation of suprasegmental patterns by five-year old children with adequate and inadequate articulation. Journal of Speech and Hearing Disorders, 45, 390-400.

Shriberg, L., Kwiatowski, J., & Rasmussen, C. (1990). Prosody-Voice Screening Profile. Tucson: Communication Skill Builders.

Shriberg, L.D. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. Journal of Speech and Hearing Research, 36, 105-140.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997a). Developmental apraxia of speech: I. Descriptive and theoretical perspectives. Journal of Speech, Language and Hearing Research, 40, 273-285.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997b). Developmental apraxia of speech: II. Toward a diagnostic marker. Journal of Speech, Language and Hearing Research, 40, 286-312.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997c). Developmental apraxia of speech: III. A subtype marked by stress. Journal of Speech, Language and Hearing Research, 40, 313-337.

Templin, M. (1953). Norms on a screening test of articulation for ages three through eight. Journal of Speech and Hearing Disorders, 18, 323-331.

Templin, M., & Darley, F., (1960) The Templin-Darley Tests of Articulation. Iowa

City: University of Iowa Bureau of Educational Research and Service.

Velleman, S. L., Huntley, R., & Lasker, J. (1991). Is it DVD or is it phonological disorder? ASHA Convention, Atlanta.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. Journal of Memory and Language, 34, 311-334.

Wingfield, A., Lombardi, L., & Sokol, S. (1984). Prosodic features and the intelligibility of accelerated speech: Syntactic versus periodic segmentation. Journal of Speech and Hearing Research, 27, 128-134.

Zimmerman, I.L., Steiner, V.G., & Pond, R.E. (1992). Preschool Language Scale - 3rd Edition. New York: Psychological Corporation.