THE PROSODY-VOICE SCREENING PROFILE (PVSP): PSYCHOMETRIC DATA AND REFERENCE INFORMATION FOR CHILDREN

Phonology Project Technical Report No. 1

Lawrence D. Shriberg

Joan Kwiatkowski

Carmen Rasmussen*

Gregory L. Lof

Jon F. Miller

Spring, 1992

Phonology Project, Waisman Center on Mental Retardation and Human Development,

University of Wisconsin-Madison

*** Madison Metropolitan School District**

Preparation of this report was supported by research grant R01 DC00496 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

ABSTRACT

The Prosody-Voice Screening Profile (PVSP) *is a perceptual procedure to assess a speaker's prosody and voice in conversational speech. The PVSP provides summative and per-utterance data on the appropriateness of a speaker's phrasing, rate, stress, loudness, pitch, and quality. Quantitative information is obtained on 31 types of exclusion codes reflecting paralinguistic status and 31 subtypes of inappropriate prosody and voice. This report is divided into three sections: (a) conceptual and technical rationale for conversation-based prosody-voice assessment, (b) psychometric findings from validity, reliability, and efficiency studies, and (c) prosody-voice reference data for 252 approximately 3-19 year-old children with normal and disordered speech development. Some of the information in this first technical report overlaps information available in more detail in the* PVSP Training Manual. *Other sections present new technical data gathered in 1990-1992.*

A prosody-voice assessment procedure termed *The Prosody-Voice Screening Profile* (hereafter, for convenience, the *PVSP*; Shriberg, Kwiatkowski, & Rasmussen, 1990) was developed in the context of a research program in speech disorders of known and unknown origin. The goal was to develop a psychometrically stable procedure based on the same conversational speech sample used to assess speech production, language production, and intelligibility. The first section of this technical report provides background on prosody and voice assessment, including rationale on conceptual and technical issues and a brief description of the major elements of the procedure. The second section describes findings from psychometric studies to estimate the validity, reliability, and efficiency of the PVSP. A final section provides detailed prosody-voice reference data for children with normal speech-language development and children with speech-language disorders of known and unknown origin.

BACKGROUND

Conceptual and Technical Considerations

Prosody occupies a unique place in the study of normal and deviant communication. Unlike speech, language, fluency, voice, and hearing disorders, which each have their own research literatures and clinical subspecialities, the area of prosody disorders has no recognized subdiscipline. Relevant theories, research, and applied information on prosody are found in many fields, including descriptive linguistics, psycholinguistics, neurolinguistics, developmental linguistics, psychiatry, communication arts, the phonetic sciences, and communicative disorders. Theoretical frameworks and applications include proposals to characterize the underlying organization of prosody in languages and language users, algorithms to deal with prosodic information in speech recognition systems, models of the motor control and phonatory mechanisms subserving prosody in manifest speech, and functional analyses of prosody as a reflection of sociolinguistic mores and affective traits and states. Assessment methods for disordered prosody range from brief check lists, to elaborated scaling tasks, to a variety of instrumental approaches, with increasing availability of dedicated devices and applications software to display and quantify relevant acoustic correlates.

Definitions

Traditional linguistics distinguishes speech from language, with speech divided into segmental and suprasegmental levels of processing. A useful applied distinction between the segmental and suprasegmental (i.e., prosodic) domains of speech is provided by Stevens, Nickerson, and Rollins (1983), who defined prosody as: " . . . those characteristics that span linguistic units longer than a phonetic segment" (p. 35). Stevens and colleagues then specify the three primary linguistic parameters within prosody: " . . . the contour of fundamental frequency versus time, the durations of certain of the speech events and pauses, and the assignment of relative prominence or stress to different syllables" (p. 35).

A conceptual and methodological problem in the assessment of the suprasegmental behaviors described by Stevens et al. (1983) is that such information is always referenced to the speaker's vocal function values. Because suprasegmental elements 'ride' on voice production, a prosody assessment procedure needs to deal directly with vocal function as referenced to normative data in the appropriate ambient community. Thus, in addition to the phrasing, rate, and stress domains of suprasegmentals as defined by Stevens and colleagues, a clinically-relevant prosody assessment procedure must also include information on the perceptual correlates of a speaker's vocal pitch, loudness, and quality. Although the term *prosody* could reasonably be extended to subsume both prosody and these latter characteristics, the hyphenated term *prosody-voice* in the PVSP retains the relevant distinctions in collateral levels of processing within the prosodic domain of phonology. That is, for both conceptual clarity in research and for the use of PVSP information in clinical contexts, the hyphenated term prosody-voice is proposed as an appropriate label for the suprasegmental domain of communication.

Perceptual vs. Instrumental Assessment of Prosody-Voice

For validity, reliability, and efficiency purposes, an instrumental approach to the measurement of prosody-voice is preferable to a purely perceptual approach. Although dedicated instruments and applications software are available for certain measurement tasks in clinical speech pathology, completely objective measurement of all relevant parameters for prosody-voice assessment is currently not a technical option. For example, recognition, quantification, and classification of such diverse prosody-voice behaviors as sound/ syllable repetitions, use of inappropriate lexical, emphatic, and sentential stress, inappropriate intonation in pragmatic contexts, breathiness, nasality, denasality, and other parameters cannot be accomplished by current voice or speech recognition programs. Although emerging algorithms for computer-assisted assessment of these and other parameters have promise for future implementation (cf., Karnell, Scherer, & Fischer, 1991), a comprehensive procedure providing information on all relevant parameters of prosody-voice analysis is currently feasible only if accomplished using the perceptual decisions of a trained examiner (cf., Crystal, 1982; Gelfer, 1988; Hirano, 1981; Laver, 1980; Murry, Brown, & Rothman, 1987).

Screening Prosody-Voice

The rationale for two additional perspectives on the measurement of prosody-voice is also based on current technical limitations on the availability of a comprehensive instrumental approach to assess phrasing, rate, stress, pitch, loudness, and quality. First, if an omnibus prosody-voice assessment procedure can currently be based only on perceptual judgements, the procedure should limit its assessment goals to screening rather than detailed analyses or differential diagnosis. Specifically, the primary goals should be limited to the identification of potential speakers with prosody-voice involvement and perhaps for such uses in intervention as generalization and maintenance probes to monitor progress. For persons identified as having potential prosody-voice involvement, subsequent analyses using more fine-grained diagnostic-assessment tools, including instrumental technologies and other protocols, can provide the required quantitative and qualitative information for specific clinical and research questions.

The second rationale following from current technical limitations is that a screening instrument should provide a profile reflecting pass-fail status for each of the relevant behaviors within the domain. Unlike diagnostic instruments, a screening measure does not need to scale severity of involvement for each prosody-voice characteristic. Rather, consistent with the psychometric goals of other screening instruments, the validity of a screening instrument for prosody-voice should be judged on how well it meets sensitivity/ specificity criteria. The screening instrument should have the requisite sensitivity to detect all involved individuals, at the cost of specificity constraints resulting in over-referral of persons who on subsequent assessment are within the normal range on the construct under test.

Sampling Context

A final consideration in developing a procedure to yield a prosody-voice screening profile on six suprasegmentals reflects a major issue in clinical assessment: what is the appropriate sampling context for prosody-voice assessment? As with all assessment instruments in communicative disorders, the basic choice is whether the data are to reflect a socially valid sample of customary or typical speech, or whether the data reflect verbal behaviors evoked in response to a specific set of test stimuli (cf., Morrison & Shriberg, in press). Conversational speech with a relative or peer, with the sample recorded in a natural setting without the speaker's knowledge, might be considered one endpoint on a sampling dimension. At the other end are controlled stimulus-response tasks, such as recording the speaker repeating a list of nonsense words or short experimental phrases or utterances presented by an unfamiliar examiner. Debate on the value of data from such spontaneous versus controlled or 'formal' assessment contexts and associated scoring permutations is found throughout the assessment literature. The typical consensus position favors an assessment battery including both sampling approaches, followed by careful interpretation of findings.

A conversational speech sample is used in the PVSP procedure because it is the only sampling context that enables an integrated assessment of speech-language-prosody. As described in the following sections, this methodological approach requires the development of many conventions to account for the variety of frequent and infrequent behaviors that occur in unconstrained conversational speech. The increased costs in procedural complexity are offset by gains in clinical and research validity, as well as in the efficiency of accomplishing speech-language-prosody analyses on one conversational speech sample.

Development and Brief Description of the PVSP

History and Research Plan

A protoversion of the PVSP was used in studies describing speech, language, and prosody-voice characteristics of children with speech delays of unknown origin (Shriberg & Kwiatkowski, 1982; Shriberg, Kwiatkowski, Best, Hengst, & Terselic-Weber, 1986). The procedure used in these studies yielded only nominal-level summative decisions based on the percentages of utterances in which performance on each of six suprasegmentals (Phrasing, Rate, Stress, Pitch, Loudness, Quality) was judged inappropriate for the conversational context. Specifically, each of the six suprasegmental categories was estimated as either *normal* (no inappropriate utterances in the sample), *questionable* (10%-15% inappropriate utterances), or *involved* (more than 15% inappropriate utterances).

The original prosody-voice assessment procedure was later elaborated in a study of phonological and social-vocational issues in a group of adults with mental retardation (Shriberg & Widder, 1990). Codes for subtypes of inappropriate prosody-voice behaviors were developed specifically for individuals with mental retardation, including a set of perceptual criteria for each code and summary analyses at the interval-level of measurement. The theoretical and clinical utility of the prosody-voice data obtained from these studies provided the impetus to initiate a research program to develop and psychometrically validate a more comprehensive screening instrument.

The current PVSP was developed over a three-year period that included five sequential research goals: (a) collect and review audiocassettes containing continuous speech samples from 10 speaker groups from different sites in the United States and Canada, including speakers (primarily children) with normal speech development, speech-language delay of unknown origin, dysarthria, dysfluency, apraxia of speech, craniofacial disorders, voice disorders, hearing disorders, emotional disorders, and mental retardation; (b) develop the form and content of the procedure, including conceptual, perceptual, and instrumental support for each procedural guideline, rule, and code; (c) select and validate the audiocassette exemplars and associated instructional text to teach the perceptual skills required to code prosody-voice; (d) conduct initial psychometric studies, including validity, reliability, and efficiency estimates; (e) disseminate preliminary findings for comment (Shriberg, Kwiatkowski, & Rasmussen, 1989a, 1989b); (f) conduct external psychometric studies and disseminate the procedure (Shriberg et al., 1990); and (g) collect and disseminate all relevant technical information on the PVSP, including reference data for speech-normal and speech-disordered children (this technical report).

Brief Description of the PVSP

Data for the PVSP are based on audiocassette recordings of spontaneous conversational speech samples. A set of free speech sampling procedures is followed to insure high quality recordings and linguistically appropriate samples (Shriberg et al., 1990). The paper and pencil formats for scoring and plotting PVSP data are illustrated in the sample case data in Figures 1 and 2. As shown in the left panel in Figure 1 [see Figure 1, page 45], utterances in the speech sample are coded to yield a pass-fail profile reflecting the percent of appropriate utterances for the three prosody and the four voice suprasegmentals. A score of 90% or better meets criteria for *pass*, whereas a score below 90% is considered a *fail*. Evidence presented later suggests the utility of adding a *questionable fail* (80%-89.9%) area to the high end of the *fail* range.

Utterances in the sample are excluded from prosody-voice coding if they meet one or more of 31 exclusion codes as shown in the top section of the right panel in Figure 1. Data on the frequency of occurrence of each exclusion code provide potentially useful information on a number of diagnostic and discourse issues, including evidence of both appropriate and inappropriate paralinguistic behaviors. Also shown in the right panel in Figure 1 are the 31 inappropriate prosody-voice codes from which the summative data in the left panel are derived. In addition to the category for *appropriate* prosody-voice, 15 categories are used to describe inappropriate prosody occurring in an utterance, and 16 categories are used to classify inappropriate voice. Specific information on normative issues and several other considerations used to generate response definitions for each code are provided in a following section on concurrent validity.

As shown in the left panel in Figure 2 [see Figure 2, page 46], coding logs are used to keep track of the sequence of PVSP decisions, with as few as 12 codable utterances required for valid prosody-voice screening of certain speakers. A page for Comments and Recommendations, as illustrated in the right panel in Figure 2, is used to summarize clinical-research findings and recommendations.

The four-panel display in Figure 3 [see Figure 3, page 47], termed a *Prosody-Voice Profile* illustrates the use of PVSP information in a research context. A Prosody-Voice Profile is obtained from a utility program in the PEPPER package running on a VAXstation 3100 (Shriberg, 1986, in submission). The data in this figure were taken from a study of children with normally developing (N) and delayed (D) speech (to be described in a later section). The four panels in the Prosody-Voice Profile display respectively: the summative data on the six suprasegmentals (top left); the percentage of utterances excluded because they met criteria for one or more of the 31 exclusion codes (top right; see Figure 1, right panel for key to all codes); and the percentages of utterances codable for prosody-voice that met criteria for 1 of the 15 inappropriate prosody codes (lower left) and 16 voice codes (lower right). Detailed information for each of the elements shown in these clinical (Figures 1 and 2) and research (Figure 3) examples are presented in the following discussions of validity, reliability, and efficiency studies.

PSYCHOMETRIC STUDIES

Validity Studies

Face, Content, and Consensual Validity

Face, content, and consensual validity for a measure is claimed when the consensus opinion of experts is that the items and subscales are a valid reflection of the content domain for the constructs the measure purports to assess. Under such validity inspection for the PVSP are the following questions (see Figure 1 for all references to PVSP elements): (a) Do the 31 *Exclusion Codes* have face, content, and consensual validity as necessary and sufficient utterance conditions under which prosody-voice should *not* be coded? (b) Do the 31 inappropriate *Prosody-Voice Codes* have face, content, and consensual validity as necessary and sufficient descriptors for the varieties of disordered prosody-voice observed in children with communicative disorders? and

(c) Do the six suprasegmentals divided into Prosody (Phrasing, Rate, Stress) and Voice (Loudness, Pitch, Quality: Laryngeal; Resonance) have face, content, and consensual validity as necessary and sufficient subdomains of suprasegmentals in conversational speech?

Validity data for the three questions were obtained in three stages. First, a review of the prosody-voice literature was undertaken to identify and cross-tabulate the categories used to classify disordered prosody and voice (Shriberg et al., 1989a, 1989b). Research with the two previous versions of the procedure had established the need for many categories; emphasis for the current procedure was to expand the procedures for use with more involved and older speakers. Second, emerging and candidate categories and terms were discussed with colleagues who conduct clinical research in the primary areas of speech, language, fluency, voice, and hearing disorders. Third, the first three authors listened to several hundred speech samples from the 10 normal and disordered speech-language categories listed earlier to attempt to capture all perceptual aspects of voice and prosody. New and modified exclusion and prosody-voice codes were developed as *necessary* until they were *sufficient* to quantify perceptual impressions of inappropriate prosody-voice on all new speech samples. For example, PV32: Nasopharyngeal resonance was developed because no one term currently used in the literature clearly captured this percept. Thus, the claim of face and content validity for the PVSP is based on the survey of the literature, prior work, and the development of the necessary and sufficient codes to describe inappropriate prosody-voice occurring in the extensive audiocassette library. The claim for consensual validity is based on discussions with knowledgeable clinical-research colleagues on the provisional adequacy of the 31 codes for inappropriate prosody-voice.

Criterion Validity

Development of response definitions to identify each of the inappropriate prosody-voice codes was guided by literature sources and, where possible, instrumental verification using audiotaped samples from the library of tapes of speech-normal and speech-delayed children. The several stages of these procedures for each of the relevant suprasegmentals are described in the Appendix.

Where possible, instrumental procedures were also used to estimate the criterion validity of over 300 audiotaped exemplars selected to teach the coding procedures. Exemplars were coded by consensus by the first three authors, using well-maintained Dictaphone 2550 audiocassette playback devices. Where the criterion validity of these perceptually-based coding decisions could not be determined by instrumental means, criterion validity was estimated using comparisons with the perceptual decisions of a panel of expert listeners. The following sections describe the methodologies and findings of the instrumental and perceptual studies.

Instrumental Validity Study

Of the six suprasegmentals--Phrasing, Rate, Stress, Loudness, Pitch, and Quality (Laryngeal; Resonance)--instrumental validation of perceptual decisions was feasible for Rate, Stress, Pitch, and two of the descriptors for inappropriate Quality. Two signal processing environments, *CSpeech* (Milenkovic, 1991) and *VOCAL* (1989), were used for the comparative analyses in these four suprasegmental domains. All instrumental measures were accomplished by the fourth author who worked independently of the first three authors and was blind to their consensus perceptual decisions on most exemplars. The Appendix provides technical information on the software and procedures used to digitize, obtain reference values, and compare perceptual to instrumentally-aided decisions. The following sections review the data obtained from procedures described in the Appendix and summarized in Table 1 [see Table 1, page 28]. In Table 1, instrumental confirmations of perceptual decisions are expressed as percentages of confirmations.

1. *Rate*. As shown in the left side of Table 1, perceptual-instrumental comparisons were made on 26 of the 39 (67%) exemplars used in a manual to train listeners to identify inappropriate Rate (Shriberg et al., 1990). As shown on the right side of Table 1, 92% of the Rate comparisons confirmed the perceptual judgements, with 20 (77%) exactly similar and an additional 4 (15%) in the same subclass. Beginning with the Too Slow subclass, the instrumental confirmation data for 10 of the 14 (71%) Too Slow exemplars, PV9: Slow Articulation/Pause Time and PV10: Slow/ Pause Time, supported the perceptual decisions. Two of the failures to confirm the perceptual decisions met criteria for the classification of Too Slow but disagreed at the subclass level of PV9 versus PV10. In one of these exemplars, an equivocal signal associated with audible inspiration/expiration may have influenced the acoustic timing of Articulation Time. Both of the other confirmation failures met instrumental criteria for Too Slow but were perceptually judged as borderline counterexamples. Ten of the 12 (83%) exemplars for PV11: Fast and PV12: Fast/ Acceleration were supported instrumentally.

2. *Stress*. The acoustic output and agreement criteria for Stress judgements are described and illustrated in the Appendix. Concurrent validity was randomly assessed for a randomly selected group of 28 of the 50 (56%) Stress exemplars, with 25 confirmations yielding an overall agreement level of 89%. As shown in Table 1, all three (100%) of the PV13: Multisyllabic Word Stress exemplars were judged by instrumental criteria to fit the response definitions established for the perceptual coding decisions. Four of the four (100%) exemplars for PV14: Reduced/Equal Stress and 18 of the 21 (86%) exemplars for PV15: Excessive/Equal/Misplaced Stress were confirmed by the acoustic data. The remaining three exemplars for PV15 were not actually disconfirmed, as the acoustic data were only useful to support the perceptual correlates of intensity, frequency, and duration involved in the perception of inappropriate stress. For example, the concept of *misplaced stress* requires a linguistic decision on whether the manifest word stress is appropriate in the discourse context. The validity of such perceptual decisions can be supported by acoustic data, but the acoustic signal does not provide a point-to-point representation of the construct of appropriate linguistic stress.

3. *Pitch*. The Appendix describes the procedures used to establish reference data for children's fundamental frequency and those used to assess the concurrent validity of the pitch exemplars used in the PVSP to train listeners to identify inappropriate pitch. As shown in Table 1, 14 of the 14 (100%) pitch exemplars in the PVSP training materials were technically appropriate for concurrent validity comparison with exact confirmations obtained on 10 (71%) exemplars and within-class confirmations for the remaining 4 (29%) exemplars. Two of the exemplars for PV19: Low Pitch/Glottal Fry could not be confirmed instrumentally for the criterion duration and sentence location of glottal fry. Similarly, two of the exemplars for PV21: High Pitch/ Falsetto could not be confirmed instrumentally for

the criterion frequency and duration of falsetto.

4. *Quality*. Although considerable progress has been made on the objective assessment of such disordered voice qualities as breathy, rough, and denasal, considerable disagreement about appropriate methodology is evident in the acoustics and aerodynamics literatures. Therefore, the validity of most of the voice quality exemplars was assessed in the perceptual validity study (to follow). As shown in Table 1, the criterion validity of 25 of the 31 (81%) exemplars for PV26: Break/Shift/Tremulous and PV27: Register Break was assessed instrumentally using the procedures described in the Appendix. Twenty-four of the 25 (96%) perceptually-based codes were exactly confirmed by the instrumental analyses.

Perceptual Criterion Validity Study

Three clinicians-researchers provided concurrent perceptual validity data on the PVSP codes that could not be assessed by instrumental means alone. Each of the judges had extensive experience in hospital-based voice clinics, as well as research credentials in the voice sciences. Each of these off-cite judges was provided with a stimulus tape and a booklet of definitions for the descriptors they were to use to evaluate inappropriate pitch and inappropriate laryngeal and resonance quality. The judges independently progressed through the materials using their preferred audio playback system. For each of the 167 samples on the stimulus tape--136 exemplars of inappropriate pitch and quality and 31 randomly assigned foils (i.e., utterances that did *not* meet PVSP criteria for an exemplar category)--the judges indicated whether or not they agreed with its use as an exemplar. That is, they were asked to decide whether the exemplar met the PVSP criteria for the inappropriate voice code. They also were asked to provide a written rationale for each of their decisions. Results are shown in Table 2 [see Table 2, page 29].

The percentage data in Table 2 indicate that on average, the judges provided criterion validation for approximately four out of every five (78%) exemplars used to teach a type of inappropriate voice pitch, laryngeal quality, or resonance quality. Three classes of disagreements on the exemplars were evident in analysis of the judges' anecdotal comments: (a) they did not hear the percept described as the criteria for the exemplar, (b) they heard the percept, but it was not severe enough to meet their understanding of the criteria for inappropriate voice, and/or (c) they heard the criterial inappropriate voice, but it did not meet a PVSP rule for those codes that required inappropriate voice to occur on at least 50% of the words in the utterance. As shown in Table 2, criterion validation was strongest for the laryngeal quality exemplars (Breathy, Rough, Strained, and Diplophonia; average agreement, 86.4%), less strong for the inappropriate Pitch exemplars (Low/Glottal Fry, Low, High/Falsetto, High; average agreement, 77.1%), and least strong for the resonance quality exemplars (Nasal, Denasal, Nasopharyngeal; average agreement, 74.5%).

Three observations about the data in Table 2 are important to underscore. First, the number of exemplars for some categories was relatively small (i.e., 7 of the 11 categories had fewer than 10 exemplars). Second, among the 136 exemplars, there were only 4 that were not attested as meeting criterion by at least one of the three judges. Finally, there were observable trends among the three judges, with Judge 2 and Judge 3 differing most in agreement with the keyed exemplars. Together with findings from the instrumental validity studies, the data in Table 2 are viewed as support for the criterion validity of the PVSP inappropriate prosody and voice codes. To the degree that judges disagreed with the key and with one another, these findings are consistent with the evident limitations in perceptual judgement that have been reported in all areas of communicative disorders (eg., fluency: Ingham, 1990; Ludlow, 1990; Moore & Perkins, 1990; segmental transcription: Shriberg & Lof, 1991; dysarthria: Sheard, Adams, & Davis, 1991).

Concurrent Validity

In addition to face, content, consensual, and criterion validity, an estimate of a measure's concurrent validity provides important data on the interpretation of scores from a clinical measure. In the present context, there is no comparable prosody-voice measure against which to compare PVSP scores. In lieu of an alternative measure, one estimate of the concurrent validity of the PVSP can be obtained by comparing PVSP data obtained with the present instrument to data generated on different subjects using the previous version of the PVSP.

Figure 4 [see Figure 4, page 48] is a summary of the prosody-voice involvements of each of two study groups. Based on the PVSP's 90% screening cutoff for a *pass*, the filled bars are the percentages of 90 speech-delayed children who failed or were questionable fails on each of the six prosody-voice suprasegmentals, as reported in Shriberg & Kwiatkowski (1982) and Shriberg et al. (1986). Prosody-voice coding of the 90 children in the previous data sets was accomplished by panels of judges using the 1982 protoversion of the PVSP, with *0* = appropriate prosody-voice (*pass*); *1* = slight to pronounced deviation occurring on fewer than 10%-15% of utterances (*questionable fail*); and *2* = slight to pronounced deviation occurring on more than 15% of utterances in the sample (*fail*). The cross-hatched bars are the percentages of a new group of 57 speech-delayed children who failed based on the procedures used in the current PVSP procedure. Coding of the 57 speech-delayed children in the current data set was done by the third author who followed all coding guidelines and procedures in Shriberg et al. (1990).

Compared to data from prior studies using the protoversion of the procedure, current estimates of involvement are reasonably similar for four of the six prosody-voice domains: Phrasing, Stress, Loudness, and Quality. However, clear differences are apparent for the remaining two domains, Rate and Pitch. In the current sample, only one or two (2%) children were at least questionable on these variables, whereas the prior studies indicated that approximately 25% of children were at least questionably involved. Inspection of both sets of data in relation to coding criteria for the original and the revised procedures suggests that these differences cannot be allocated to differences in only subjects or measures. Rather, it appears that both variables may account for the obtained differences. Specifically, the 90 speech-delayed children in the earlier samples may have had more prosody-voice involvement, as indicated by their lower intelligibility scores (cf., Shriberg & Kwiatkowski, in submission); however, the current PVSP procedure has more stringent criteria for rate and pitch judgements. An estimate of the standard error of measurement for each of the six suprasegmentals, based on a dataset of 252 samples (to be described), indicates that measurement error for scale scores ranges from less than 1% to approximately 3%. Thus, with the previous caveats in mind, the generally similar prevalence patterns for prior and current groups of speech-delayed children assessed by different versions of the

PVSP is viewed as providing concurrent validity support for the procedure.

Construct Validity

Finally, support for the construct validity of the PVSP is suggested by the data in Figure 5 [see Figure 5, page 49]. The two trends in this Prosody-Voice Profile are prosody-voice data for a group of 64 children (the 57 children described previously, plus 7 more children) and a group of 14 children with suspected apraxia of speech (Shriberg, Aram, & Kwiatkowski, in preparation). According to the literature on the adult form of acquired apraxia of speech, a hallmark diagnostic characteristic of this clinical entity is prosodic involvement (Kent & Rosenbek, 1982). If the construct validity of a procedure is supported to the degree that it explicates the target construct, to what degree does the PVSP provide useful information on the type and degree of prosody-voice involvement in children with suspected developmental apraxia of speech?

As shown in Figure 5, compared to the speech-delayed children, children with suspected developmental apraxia of speech produced a statistically lower percentage of appropriate utterances (top left panel) and correspondingly higher percentages of inappropriate prosody codes (lower left panel). Substantive discussion of these findings (see key to descriptors, Figure 1) will be reported in a forthcoming paper. In the present context, the descriptive trends and significant prosody-voice differences between speech-delayed children and children with suspected developmental apraxia of speech are viewed as construct validity support for the PVSP procedure. Essentially, the procedure provides a means to explicate the nature of prosodic involvement in children with this putative clinical entity. Similar prosody-voice analyses using the PVSP have been completed on several other clinical groups within whom the constructs of prosody-voice involvement are also of interest, including children with Down syndrome, fluctuating conductive hearing loss associated with early recurrent otitis media with effusion, and psychosocial-affective involvement.

Reliability Studies

Several independent estimates of the reliability of PVSP data were obtained, including information on sampling stability, interjudge and intrajudge reliability of coding, and internal consistency and stability of screening decisions.

Stability of Speech Sampling

The stability of the conversational speech samples has been supported in a number of studies concerned with segmental variables (Shriberg & Lof, 1991; Morrison & Shriberg, in press; Shriberg & Kwiatkowski, 1982, 1985; Shriberg & Widder, 1990). Detailed descriptive and inferential statistical analyses presented in these studies indicate that children and adults produce continuous conversational speech samples that are consistent within and across subjects for measures of utterance productivity, intelligibility, representativeness of canonical, grammatical, and intended segmental forms, and reactivity. Specific to the present context, the intersample and intrasample stability of such distributional characteristics as parts of speech, type/token ratios per minute, number of intelligible words per minute, canonical forms, percentage of occurrence of intended phonemes, and in particular, speech registers (Shriberg & Kwiatkowski, 1985), indicates that conversational speech is robust relative to structural, linguistic, and pragmatic characteristics. The procedures for speech sampling presented in Shriberg et al. (1990) are the outgrowth of methodological work to standardize procedures for conversational speech sampling, including provisions for identifying affective states in which a speaker's speech sample would be invalid for the purposes of prosody-voice assessment.

Intrajudge and Interjudge Agreement

Initial Studies

The three domains in which examiner reliability is relevant in the PVSP procedure are: (a) the application of rules for segmentation (i.e., parsing the conversational speech sample into utterance length units), (b) the application of rules and perceptual decisions for excluding utterances from prosody-voice coding (exclusion codes), and (c) the application of rules and perceptual decisions for determining appropriate prosody-voice and classifying inappropriate prosody-voice behaviors. The reliability of computational and clerical tasks required to derive percentage scores was not considered relevant to estimate. Formatting procedures were designed to maximize accurate and efficient recording of codes, percentage calculations, and summative scoring.

During the different stages of PVSP development, interjudge and intrajudge agreement estimates on utterance segmentation, exclusion coding, and prosody-voice coding within and among the first three authors ranged from point-to-point percentages of agreement in the low 70%s to 100% agreement (Shriberg et al., 1989a). The general pattern of the disagreements was predictably related to the cognitive and/or perceptual difficulty of the coding task in relation to the status of the utterance on the dimension being judged. Agreement was lowest when the exemplar represented a midway point on the dimension being judged and certain segmentation, exclusion, and prosody-voice rules and codes were routinely associated with lower interjudge and intrajudge agreement.

Examiner Reliability Study I

The first of two examiner reliability studies was conducted with three university students who responded to a notice for a short-term research job. The students--one masters-level and two undergraduates--were hired on an hourly basis to learn the procedure on their own from the training manual and audiotapes, and when ready, to take three tests to assess their understanding and reliability of exclusion and prosody-voice coding. No other selection criteria were used; the three judges were essentially the first three persons to respond to the job vacancy notice. They proceeded independently to learn the PVSP procedures (see the following Efficiency section for more detail). They then each took a 30-utterance test, received brief feedback on the nature of their disagreements with the key, and then took a second and third 30-utterance test without further feedback. The three tests had been equated for overall difficulty.

In keeping with the need to estimate individual reliability percentages for 'easy' versus 'hard' perceptual decisions (cf., Diedrich & Bangert, 1976; Kearns, 1990; Kearns & Simmons, 1988), two percentages of agreement were calculated for each of the judges: agreement on whether an item was *appropriate* or *inappropriate* (i.e., without regard to the exact *inappropriate* code) and exact agreement on the code used for all utterances keyed as *inappropriate*. The overall agreement data, agreement on *appropriate* and *inappropriate*, are taken to reflect the reliability of screening decisions (i.e., the summative percentages used for *pass*/*fail* decisions). Agreement on only those utterances keyed as *inappropriate* was presumed to reflect the outcome for the more difficult perceptual decision. Table 3 [see Table 3, page 30] is a summary of the agreement data for all judges on each of the three tests. It should be kept in mind that the number of tokens keyed as *inappropriate* was almost half the number keyed as *appropriate*. The three judges averaged low to mid 80% agreement with the key on both agreement criteria, with Judge 1 having more difficulty learning the task (see later Efficiency section).

For the most stringent test of reliability--an estimate of agreement with the key reached by 'average' independent learners using the training materials--the agreement data in Table 4 [see Table 4, page 31] are the averaged performance of all three judges over all three tests. The two levels of agreement are the same as those described for Table 3. These estimates yielded overall interjudge agreement figures ranging from approximately 77% to 96%, with mean agreement on the exact codes for just the utterances keyed as *inappropriate* ranging from approximately 69% to 100%. The pattern of percentages is generally consistent with the criterion validity data, with reliable use of the quantitative and perceptual criteria for inappropriate pitch and quality among the more difficult PVSP codes to acquire. Again, this reliability estimate reflects averaged performance over three tests of the first three students who applied for a job requiring they learn a set of guidelines and perceptual skills entirely on their own.

Examiner Reliability Study II

The intrajudge and interjudge consistency of PVSP scores for a large sample of children with speech-language disorders of known and unknown origin was estimated in a second examiner reliability study. One of the authors had independently scored 24-utterance speech samples from 28 children with normally-developing speech and speech delays of known and unknown origin. The former group consisted of 14 3-19 year-old children (M = 6 years, 3 months) randomly selected from a reference database for language acquisition (Miller, 1990) and from control subjects used in other studies. The 14 3-15 year-old speech disordered children were randomly selected from a database of subjects in a variety of studies; they included children with speech delays of unknown origin and speech delays associated with mental retardation, psychosocial involvement, early recurrent otitis media, and submucosal clefts of the palate. All tapes had been prosody-voice coded by one of the authors (see later Efficiency section for information on procedures). For the current purposes, the randomly selected sample of 28 tapes represented 11% of a database of 252 samples. The author and the masters-level student who had participated in Study I individually scored all the transcripts using a master transcript to enable utterance-by-utterance comparison. Reliability analyses for exclusion coding and prosody-voice coding are summarized respectively in Table 5 and Table 6 [see Tables 5 and 6, pages 32-33].

Table 5 is a summary of three increasing levels of precision of intrajudge and interjudge agreement on exclusion coding. The first block of agreement percentages indicates relatively high (81%-100%) intrajudge and interjudge agreement on the assignment of utterances to either an exclusion code or a prosody-voice code. The middle block of agreement data also indicates good (76.5%-100%) intrajudge and interjudge agreement on assignment of an utterance to one of the four categories of exclusion codes (i.e., indicating agreement that an exclusion was assigned a Content/Context, Environment, Register, or State code). Finally, when percentaged on the basis of exact agreement on 1 of the 31 Exclusion Codes, agreement again ranged from 76.5% to 100%. Mean agreement percentages for all six estimates were above 90%.

Table 6 is a summary of the intrajudge and interjudge percentage of agreement on the six prosody-voice domains, divided into Screening Agreement and Exact Agreement. The three estimates of the reliability of screening decisions reflect the consequences of prosody-voice coding on the summative percentage scores used for a *pass*/*fail* decision on each of the six suprasegmentals. The three sets of percentages indicate agreement on utterances considered *appropriate* (referenced to the comparison transcript or judge), *inappropriate*, and an overall *appropriate* plus *inappropriate* estimate. Both intrajudge and interjudge agreement on *appropriate* ranged from approximately 86% to 100% across the summative screening variables. Intrajudge and interjudge agreements for utterances judged *inappropriate* by the standard judge (i.e., one of the authors) on the first listening ranged from approximately 22% to 97%. The total number of tokens and number of tokens per child sample was relatively small for many of these comparisons (e.g., only nine utterances for inappropriate pitch). Anecdotal comments by the standard judge indicate that there may have been some 'drift' or 'decay' (O'Leary & Kent, 1972) in the response definitions used for some suprasegmentals. Some of the samples had been scored more than one year earlier, at a time when response definitions and the audiocassette training tapes had not been fully developed. As shown in the Overall data for Screening Agreement, point-to-point percentages ranged from approximately 80% to 99%. The ranges of Exact Agreement on each of the 31 *inappropriate* prosody-voice codes, as shown in the second set of intrajudge and interjudge data in Table 6, was from approximately 47% to 100%. Overall Exact Agreement, reflecting agreement on utterances judged *appropriate* and the exact *inappropriate* code, ranged from approximately 74% to 99% across the six suprasegmentals.

As indicated in both these Screening Agreement and the Exact Agreement estimates, intrajudge and interjudge reliability for the use of the 31 prosody-voice codes may range from acceptable to unacceptable for certain utterances, children, codes, examiners, and lengths of time. PVSP screening decisions and assignment of an inappropriate code were particularly difficult for pitch and laryngeal quality. These findings are consistent with criterion validity results, reflecting the particular difficulties in making unaided perceptual decisions in these vocal function domains for speakers who have marginal involvement. As with all assessment instruments, it is important to obtain frequent reliability checks on PVSP decisions and to undertake recalibration as needed.

Internal Consistency of PVSP Scores and PVSP Screening Outcomes

Initial Studies

The internal consistency of the PVSP was first assessed in a study using the first and second 12 utterances in 24-utterance samples from 64 children with developmental phonological disorders. Using the 90% *pass* criterion for each suprasegmental, results indicated that 80%-100% of the retest decisions were similar across the six suprasegmentals, providing the child received a *pass* based on the first 12 utterances. There were too few subjects receiving a *fail* on the first 12 utterances to adequately assess the stability of failing scores on these few occurrences. Pending confirmation in the studies described below, the findings from the initial studies generated the interim procedural guideline that a 12-utterance sample is sufficient for speakers who have a clear *pass* on all six suprasegmentals, or who consistently produce the same, readily-coded inappropriate prosody-voice in each utterance. However, a 24-utterance sample is suggested for speakers with low utterance productivity or speakers whose prosody-voice is inconsistently or marginally inappropriate on any one of the six suprasegmentals.

Large Group Study

A second estimate of the internal consistency of PVSP scores was obtained using two samples of scores randomly selected from the database of 252 PVSP samples described above. Table 7 [see Table 7, page 34] is a summary of the internal consistency results based on PVSP scores and Table 8 [see Table 8, page 35] is a summary of internal consistency findings for screening outcome decisions. The data in both tables were obtained from the same two randomly chosen PVSP transcripts from 40 subjects in each of the two speech status groups referred to previously--the 115 speech-normal children and 137 speech-delayed children. The transcripts were split into odd and even utterances, with summary percentages (Table 7) and screening outcomes (Table 8) for each of the six suprasegmentals calculated from the 12 utterances in each half.

As shown in Table 7, all of the part-whole Spearman Rho coefficients are statistically significant at the .01 alpha level, with the absolute magnitude of the Rho values ranging from .66 to 1.00 (M = .88) across suprasegmentals and between the two speech status groups. The split-half coefficients were attenuated by the large number of tied scores which, on inspection, greatly affected the obtained coefficients for Phrasing, Rate, Stress, and Loudness. The screening outcome findings summarized in Table 8 form a similar pattern to the findings in Table 7. Across the six suprasegmentals and the two speech status groups, the internal consistency of *pass* and *fail* scores are higher for the part-whole comparisons. The internal consistency of the *fail* decisions is considerably less stable than that of the *pass* decisions, but there were too few samples available for appropriate comparisons.

The data in Table 7 and Table 8 are viewed as supporting the internal reliability of the PVSP, including the guidelines for obtaining 12-utterance versus 24-utterance samples developed in the internal consistency study. As inferred from both Table 7 and Table 8, a 12-utterance sample is stable (and very efficient) for speakers whose non-involvement yields a clear *pass*, whereas speakers whose inconsistent involvement yields a *fail* based on 12 random utterances are more reliably tested with a 24-utterance PVSP sample.

Efficiency Studies

Efficiency issues in assessment are concerned with the time and effort needed to (a) accurately learn a procedure, (b) reliably administer a procedure, (c) validly score a procedure, and (d) insightfully interpret the results from a procedure. To date, PVSP data on the first three of these four efficiency domains has been collected in several studies.

Efficiency Data on Learning the PVSP

Information concerning the learning process during the acquisition of the conceptual and perceptual skills required in PVSP scoring was obtained in three studies: (a) an initial study involving two clinical instructors, (b) an independent learning study involving a group of three students who learned the procedure entirely from the training manual and audiotapes, and (c) a group learners study involving a class of students who learned the procedure with the assistance of classroom instruction.

Initial Study

Information on processes involved in learning the PVSP was first assessed in a field test of a preliminary version of the procedure. Two experienced clinical instructors volunteered to learn the materials from a preliminary version of the text and take three perceptual tests that assessed learning of exclusion and prosody-voice coding. The quantitative results and their anecdotal comments on the training materials provided invaluable information on modifications in form and content needed to efficiently teach the procedure. Essentially, they experienced difficulty with the individual response definitions for the many codes, suggesting the need for training formats that would assist the reader in acquiring and retaining the cognitive concepts and perceptual skills.

Independent Learning Group Study

The independent learners were the three students--one masters-level and two undergraduates--whose reliability data were described previously. They were hired on an hourly basis to independently learn the procedure entirely from the training materials. When they felt they were appropriately prepared, they took tests to assess their understanding and interjudge agreement with the scoring key. The independent learners also kept information on the number of hours needed to learn the materials, with the decision left completely up to them about when they felt ready to take the first test. Their logs indicated a mean total training time of 15 hours 7 minutes to learn the procedure (range = 10 hr 15 min to 18 hr 45 min). Their session lengths averaged 1 hour 33 minutes (range = 56 min to 2 hr 26 min), and they reported spending an average of 16 sessions (range = 15 to 18 sessions) to learn the procedure completely on their own. In all cases, the masters-level student took the least time to learn the procedure, and the least experienced undergraduate student took the most time. Interestingly, these overall training data correspond to typical times and sessions used in semester-based laboratory work, such as learning phonetic transcription.

Group Learners Study

Informal assessment of the learning process for PVSP was completed by a class of 25 undergraduate students who learned the procedure in the context of a quarter-semester class in phonological disorders (P. Hargrove, personal communication). Over a period of nine weeks, approximately one hour per week was spent in class discussion of conceptual and procedural issues and group listening to the training tapes, with students optionally spending additional and independent time reading the training manual and listening to the training tapes. Students each completed an analysis of the prosody of one speaker.

The teacher's anecdotal report of the process and outcome of student learning can be summarized as follows: (a) they eventually grasped each element of the procedure, although the complexity of the task was initially perceived as somewhat daunting; (b) they most readily learned the utterance segmentation tasks; (c) they had some difficulty reaching agreement on some of the exclusion codes; and (d) they had the most overall difficulty learning to code Loudness and Pitch, with disagreements on specific inappropriate prosody-voice codes distributed across the 31 codes. These findings reflect the limitations in perceptual evaluation of voice, as observed previously in both the validity and the reliability data. More generally, they are reminiscent of the course of learning to phonetically transcribe speech, wherein certain response classes are extremely difficult for some, but not all, students to learn. As with other clinical skills, learning to assess prosody-voice requires a substantial commitment of time.

Efficiency Data on Administering the PVSP

As the PVSP data are taken from the same conversational speech samples used for various forms of speech assessment (e.g., severity of involvement, error pattern, intelligibility index) and language assessment (e.g., structural stage, discourse analysis), the approximately 10 minutes needed to obtain a spontaneous conversational speech sample is considered an efficient use of both a subject's and an examiner's time. That is, no special skills, stimuli, or associated time demands beyond those needed for conversational speech-language sampling are required to obtain a suitable sample for a PVSP analysis. As described above, the training manual provides specific guidelines to insure that the samples are valid and efficient for PVSP scoring.

Efficiency Data on Scoring the PVSP

Initial Study

An initial study of efficiency issues was completed based on data provided by one of the authors who independently completed PVSP analyses on 57 3-5 year-old children with moderate to severe speech disorders of unknown origin (Shriberg et al., 1989a, 1989b). The conversational speech samples had been gathered prior to the development of the PVSP and associated guidelines for obtaining samples for efficient PVSP coding. Results indicated that approximately 50% of these children's utterances were excluded from prosody-voice coding (range = 15% to 85% of utterances). Inspection of the exclusion codes indicated that 44% of these utterances met criteria for one of three codes: C7: Only One Word (35%), C12: Too Many Unintelligible Words (5%), and R2: Narrative Register (4%). The remaining 56% of excluded utterances were spread across the other exclusion codes. These findings indicate the importance of sampling procedures for the efficiency of PVSP scoring. Specifically, the examiner can maximize efficiency of subsequent PVSP scoring by using speech sampling techniques that yield relatively low frequencies of one-word responses and narrative registers. The total times required for exclusion coding and prosody-voice coding averaged 1.1 minute per utterance (range = .5 min to 1.5 min). Assuming approximately one minute per utterance for glossing/segmenting utterances, the total scoring time for a 12-utterance sample would be approximately 25 minutes.

Large Group Study

Reference data from a large group study of the time needed to score the PVSP and other efficiency questions are presented in Table 9 [see Table 9, page 36]. These data were taken from time logs kept by one of the authors who completed PVSPs on the previously described database of 252 speech samples from 11 subgroups of children with normal and disordered speech acquisition. For the present purposes, children in the 11 subgroups were divided into two large groups. A group of 137 approximately 3-19 year-old children (Mean = 6 years, 3 months; SD = 4.0 years) had speech disorders of unknown origin and disorders associated with risk factors and suspected etiologies, including early recurrent otitis media, mental retardation, suspected apraxia of speech, unilateral brain lesion, psychosocial involvement, and submucosal clefts. A group of 115 approximately 3-18 year-old speech-normal children (Mean = 5 years, 5 months; SD = 2 years, 11 months) was comprised of 71 3-5 year-old children sampled from a language database (Miller, 1990) and 44 children used as control samples in the speech disorders studies. The actual number of samples used for the different calculations in Table 9 ranged from 71 to 252.

As shown in the first row of Table 9, the average sample for both groups took approximately 42 minutes to score (SD = approximately 25 min). The speech-disordered children required twice as much time (64.7 min) as the speech-normal children (28.8 min). These figures reflect only the time needed to segment and code the samples, as glosses were already available from prior phonetic transcription.

The second row in Table 9 provides data on the total number of utterances that had to be coded to meet the criteria of 24 utterances eligible for prosody-voice coding, including the three codable warm-up utterances required by the PVSP procedure (Shriberg et al., 1990). Across the speech-normal and speech-disordered groups an average of approximately 52.7 utterances (SD = 22.4) were needed, including the average of five utterances that were needed to obtain the three codable warm-up utterances. The average number of utterances required was approximately 25% more for the speech disordered group (59.9) compared to the speech-normal children (44.1), representing approximately twice the number of prosody-voice codable utterances (24), plus the three codable warm-up utterances. These data are in good agreement with data from the initial study.

If excluded utterances in a sample are considered to be theoretically or clinically uninteresting, the procedure would be considered fairly inefficient based on this 2:1 ratio of obtained-to-codable utterances for both speech-normal and speech-delayed children. However, as described earlier, many of the exclusion codes provide information about a speaker's paralinguistic performance and thus provide important information in their own right on pragmatics and discourse. It is for this reason that the excluded utterances have been divided into the 31 types, and data on their reliability and frequency of occurrence has been carefully assembled. For example, using an earlier form of the PVSP, the prosody-voice of some adults with mental retardation was characterized as including high frequencies of some socially inappropriate behaviors (Shriberg & Widder, 1990; see also discussion of interpretation of PVSP results in Shriberg et al., 1990). The reference data in the third section of this report provides information on the relative occurrence of each of the exclusion codes in speech-normal and speech-delayed children. Additional detail at the level of groups based on risk factors and suspected etiological origin will be reported in subsequent studies. Thus, although the average time required to score a PVSP may be lengthy, the additional time required to code the exact bases for the exclusions is not viewed as inefficient. Times could be shortened for certain clinical-research tasks, however, if exclusion coding was elected to be handled as a simple binary decision.

The third row in Table 9 provides information on a third procedural convention in the PVSP that impacts efficiency--the requirement that at least 50% of the utterances in a sample be four or more words in length. As shown in the third row, an average of approximately 71% of utterances of four or more words in length were actually included in the samples, with similar percentages for both speech-status groups. As indicated later in the reference data (Table 10; see Table 10, pages 37-38), only approximately 7% of subjects had frequencies of occurrence of 5% or more of this exclusion code (i.e., had utterances that had to be excluded due solely to this requirement). This loss in efficiency is considered reasonable in relation to the validity concerns requiring that judgements of prosody-voice require utterances of varying lengths.

The final rows in Table 9 provide an informal estimate of the relative difficulty of scoring PVSP samples. The coder used a four-category system to indicate if the samples were particularly difficult to score and the perceived source of the difficulty. As shown, across both groups approximately 87% of the samples were scored without comment, with external factors (e.g., inadequate speech sample, poor audio signal, interfering tape noises) accounting for 7.6% of the remaining tapes and the speaker's severity of involvement accounting for an additional 4.3% of the difficult samples. These data differ somewhat for each of the speech status groups, with proportionally more speech-delayed children's tapes experienced as more difficult to score due to both external factors and severity of involvement. In addition to supporting the expectation that it takes more time and effort to code disordered compared to normal prosody-voice, these efficiency findings also underscore the value of obtaining good speech samples and high quality audio recordings.

Overall, these efficiency figures agree with the data from the initial studies, indicating an average of two minutes coding time per utterance. Thus, anything that reduces the overall number of utterances needed for a valid PVSP sample reduces the time needed for scoring. The time needed for glossing differs considerably, depending on such factors as the child's moment-to-moment intelligibility and the examiner's familiarity with the child. As with all such data, efficiency reflects interactions among the relative involvement of the child, the skills of the examiner who obtains the sample, and the skills of the prosody-voice judge. Our experience suggests that with practice, total times decrease substantially.

REFERENCE DATA

Reference data for 252 children with normally-developing speech-language and speech disorders of unknown and known origin are provided in Tables 10, 11, and 12. The data are taken from the same samples as described in the prior section, including 115 speech-normal children and 137 children with speech delays of both known and unknown origin.

Table 10 provides general reference data for the occurrence of the 31 PVSP exclusion codes in the speech samples. The left side of Table 10 provides group central tendency and distributional data for the proportional occurrence of each exclusion code in the two subgroups, speech-normal and speech-disordered, and combined statistics. The right side of the table provides data on the percentage of children in each of three categories of exclusion code occurrence: 0% occurrence of the code; 1-4.9% occurrence; 5% or greater occurrence. As shown in the left side of Table 10, each of the exclusion codes occurs at least once in at least one speech sample, thus supporting the content validity of the 31 exclusion codes. The most frequently occurring codes are C7: Only one word (approximately 25% of total utterances), R2: Narrative register (approximately 7% of total utterances), and C12: Too many unintelligibles (approximately 4% of total utterances). These figures are in good agreement with those found in the initial efficiency studies.

Of particular interest in Table 10 is the similarity in the occurrence of exclusion codes in the speech-normal and speech-delayed children. The distributions for each group are comparable for most of the three distributional statistics in the proportional occurrence data (mean, SD, range) and in each of the percentage categories in the percentage of children data (percentage of children with 0% occurrence of the code, 1-4.9% occurrence, and 5% or greater occurrence of the exclusion code). In addition to its use as reference data, the similarity in the two groups in Table 10 provides additional support for the stability of conversational speech samples. For example, approximately 29% of both speech-normal and speech-disordered children respond with "I don't know" (C3) on approximately 1%-5% of their utterances. Most interesting from a clinical-research perspective are the values in the range column in the left side of Table 10, which indicate that some individual children have extremely high percentages of occurrence of certain exclusion codes. Aside from explanations due to technical or speech sampling constraints, the correlates of such high rates of behaviors within any of the four classes of exclusion codes are interesting to pursue in their own right.

Table 11 [see Table 11, pages 39-40] provides data on the 31 inappropriate prosody-voice codes. The formats for data presentation are the same as those used in Table 10. Overall, the data for the speech-normal and speech-disordered children are generally similar. What differences are observed generally indicate more involvement for the speech-disordered children, particularly on several stress codes and several resonance codes. Again, the range data provide the most provocative findings, with individual children in both groups having high occurrence of some inappropriate prosody-voice scores. These data are purposefully representative of an undifferentiated group of children with normal and disordered speech, providing reference data against which to compare suprasegmental involvement of well-defined subgroups (e.g., Shriberg & Widder, 1990); see also Figure 5.

Table 12 [see Table 12, page 41] provides data for the seven summary-level suprasegmentals that comprise the PVSP profile. The left section provides group central tendency and dispersion data, and the right side provides the percentages of children who scored 90% or above (*pass*), 80%-90% (*questionable fail*), or below 80% (*fail*) on this screening measure. The speech-normal and speech-delayed children are clearly different on several of the summary suprasegmentals, although inferential tests to assess the statistical significance of differences were not deemed appropriate to compute for these reference data (see Figure 3 for a related comparison). Inspection of each of the cells in Table 12 and comparison of data for the speech-normal versus speech-disordered children generates a number of hypotheses about prosody-voice in normal and disordered speech development. In the present context, the technical focus of the data is primarily toward the use of the PVSP as a screening device for clinical research.

The data in Table 12 support the decision to locate the cutoff levels at 90% for clear *pass* and 80% - 90% for *questionable fail*. As indicated in the percentage ranges for all variables, children acquiring speech normally may have PVSP scores below these two cutoff points. The goal of a screening instrument is to adjust the rates of false positives and false negatives such that no child with a potential problem passes the screen. Follow-up inspection of the prosody-voice codes and instrumental analysis can determine the clinical severity of the problem. For example, a *questionable fail* or *fail* on laryngeal or resonance quality could readily be explained by transient lower or upper respiratory involvement (see the boxes for these state variables in Figure 1). Similarly, low appropriate phrasing scores could be due to the high rates of repetitions or revisions that are associated with periods of language expansion or particular affective states.

CONCLUSION

These psychometric and initial empirical data on the prosody-voice characteristics of children suggest that there is much to be learned about this domain of communicative disorders. As much or perhaps more than any of the other areas of speech-language processing, prosody-voice function involves trait and state variables that play a vital role in a speaker's perceived communicative competence. Perceptual screening procedures such as the instrument described in this technical report represent first generation approaches, ideally to be followed by instrumentally-aided technologies for diagnostic assessment needs in research and clinical practice.

ACKNOWLEDGMENTS

This work was supported by a grant from the Public Health Service, National Institute on Deafness and Other Communication Disorders, DC00496. We wish to acknowledge the collegiality and competent assistance of a very large number of persons who made crucial contributions at the several stages of this project. Space limitations permit only the following alphabetized list: Dorothy Aram, Russell Attoe, Barri Babow, Diane Bless, Dolores Boyd, Eugene Buder, Karen Carlson, Maria Cavicchio, Barbara Ekelman, Patricia Engebose, Stanley Ewanowski, Florence Filley, Karen Forrest, Leslie Glaze, Sandy Grafton, Patricia Hargrove, Kirsteen Harrison, Anne Heintzelman, Linda Hesketh, Rebecca Hinke, Megan Hodge, Sara Hoffman, Doris Kistler, Anne Kleckner, Karen Koerber, Danna Koschkee, Mark Leddy, Jane Loncke, Paul Milenkovic, Linda Milosky, Kellie Moran, Lois Nelson, Bruce Orchard, Amparo Ortiz, Linda Rammage, Dorothy Rorick, Margaret Rosin, Hye-Kyeung Seung, John Strei, Robin Susser, Edie Swift, Helen Thielke, Catherine Trost-Steffen, Gary Weismer, Dennis White, Carol Widder, Frank Wilson, and Wayne Zimmerman.

REFERENCES

Baken, R. J. (1987). *Clinical measurement of speech and voice*. Boston, MA: Little, Brown and Company.

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

Diedrich, W. M., & Bangert, J. (1976). Training speech clinicians in recording and analysis of articulatory behavior. Washington, DC: US Office of Education Grant No. OEG-0-70-1689 and OEG-0-71-1689.

Forrest, K., & Rockman, B. K. (1988). Acoustic and perceptual analysis of word-initial stop consonants in phonologically disordered children. *Journal of Speech and Hearing Research, 31*, 449-459.

Gelfer, M. P. (1988). Perceptual attributes of voice: Development and use of rating scales. *Journal of Voice, 2,* 320-326.

Hirano, M. (1981). *Clinical examination of voice*. New York: Springer-Verlag.

Ingham, R. J. (1990). Theoretical, methodological, and ethical issues in treatment efficacy research: Stuttering therapy as a case study. In L. B. Olswang, C. K. Thompson, S. F. Warren, & N. J. Minghetti (Eds.), *Treatment efficacy research in communication disorders* (pp. 15-29). Rockville, MD: American Speech-Language-Hearing Foundation.

Karnell, M. P., Scherer, R. S., & Fischer, L. B. (1991). Comparison of acoustic voice perturbation measures among three independent voice laboratories. *Journal of Speech and Hearing Research, 34*, 781-790.

Kearns, K. J. (1990). Reliability of procedures and measures. In L. B. Olswang, C. K. Thompson, S. F. Warren, & N. J. Minghetti (Eds.), *Treatment efficacy research in communication disorders* (pp. 79-90). Rockville, MD: American Speech-Language-Hearing Foundation.

Kearns, K. P., & Simmons, N. N. (1988). Interobserver reliability and perceptual ratings: More than meets the ear. *Journal of Speech and Hearing Research, 31,* 131-136.

Kent, R. D., & Rosenbek, J. C. (1982). Prosodic disturbance and neurologic lesion. *Brain and Language*, *15*, 259-291.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Ludlow, C. L. (1990). Research procedures for measuring stuttering severity. In J. A. Cooper (Ed.), *Research needs in stuttering: Roadblocks and future directions* (pp. 26-31). *ASHA Reports*, No. 18.

Milenkovic, P. (1987). Least mean square measures of voice perturbation. *Journal of Speech and Hearing Research, 30,* 529-538.

Milenkovic, P. (1991). *CSpeech User's Manual, Version 3.1*. Dept. of Electrical and Computer Engineering, University of Wisconsin-Madison.

Miller, J. F. (1990). SALT Reference Data Base Project. Language Analysis Laboratory, Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison.

Moore, S. E., & Perkins, W. H. (1990). Validity and reliability of judgements of authentic and simulated stuttering. *Journal of Speech and Hearing Disorders, 55,* 383-391.

Morrison, J. A., & Shriberg, L. D. (in press). Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research*.

Murry, T., Brown, W. S., & Rothman, H. (1987). Judgements of voice quality and preference: Acoustic interpretations. *Journal of Voice*, *1,* 252-257.

O'Leary, K. D., & Kent, R. (1972). Behavior modification for social action: Research tactics and problems. In L. A. Hammerlynck, L. C. Handy, & E. J. Mash (Eds.), *Behavior change: Methodology, concepts, and practice*. Champaign, IL: Research Press.

Sheard, C., Adams, R. D., & Davis, P. J. (1991). Reliability and agreement of ratings of ataxic dysarthric speech samples with varying intelligibility. *Journal of Speech and Hearing Research, 34*, 285-293.

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

Shriberg, L. D. (in submission). Four new speech-prosody measures for genetics and other subgroup research in developmental phonological disorders.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (in preparation). Segmental and suprasegmental characteristics of children with developmental apraxia of speech.

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonologic disorders I: A diagnostic classification system. *Journal of Speech and Hearing Disorders, 47*, 226-241.

Shriberg, L. D., & Kwiatkowski, J. (1985). Continuous speech sampling for phonologic analyses of speech-delayed children. *Journal of Speech and Hearing Disorders, 50*, 323-334.

Shriberg, L. D., & Kwiatkowski, J. (in submission). A descriptive profile and subgroup studies in developmental phonological disorders.

Shriberg, L. D., Kwiatkowski, J., Best, S., Hengst, J., & Terselic-Weber, B. (1986). Characteristics of children with speech delays of unknown origin. *Journal of Speech and Hearing Disorders, 51*, 140-161.

Shriberg, L. D., Kwiatkowski, J., & Rasmussen, C. (1989a, November). *The Prosody-Voice Screening Profile (PVSP): I. Description and psychometric studies*. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, St. Louis, MO.

Shriberg, L. D., Kwiatkowski, J., & Rasmussen, C. (1989b, November). *The Prosody-Voice Screening Profile (PVSP): II. Reference data and construct validity*. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, St. Louis, MO.

Shriberg, L. D., Kwiatkowski, J., & Rasmussen, C. (1990). *The Prosody-Voice Screening Profile*. Tucson, AZ: Communication Skill Builders.

Shriberg, L. D., & Lof, G. L. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics, 5*, 225-279.

Shriberg, L. D., & Widder, C. J. (1990). Speech and prosody characteristics of adults with mental retardation. *Journal of Speech and Hearing Research, 33*, 627-653.

Stevens, K. N., Nickerson, R. S., & Rollins, A. M. (1983). Suprasegmental and postural aspects of speech production and their effect on articulatory skills and intelligibility. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.), *Speech of the hearing impaired: Research, training and personnel preparation* (pp. 35-51). Baltimore, MD: University Park Press.

VOCAL: Program to edit and analyze verbal utterances. (1989). Research Computing Facility, Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison.

APPENDIX

DEVELOPMENT OF RESPONSE DEFINITIONS AND CRITERION VALIDITY PROCEDURES

Instrumental Measures

CSpeech (Milenkovic, 1991), operating on an IBM-PC platform, was used for information on fundamental frequency, jitter, shimmer, duration, and a waveform display. A Marantz PMD221 3-head audiocassette recorder was used as the input device. Speech signals were preamplified and low-pass filtered using an eight-pole Butterworth filter (Model 901F1, Frequency Device, Inc., Haverhill, MA) with a 10K Hz cutoff frequency and subsequently sampled at 20K Hz using an analog-to-digital converter with 12 bits of numeric resolution (Labmaster, Scientific Solutions, Solon, OH). Measurements of fundamental frequency, jitter, and shimmer use Henke's FPRD algorithm (cf., Forrest & Rockman, 1988; see also Milenkovic, 1987).

VOCAL, a waveform editing and analyzing software package running in the Harris/800 minicomputer environment, provided displays of the fundamental frequency contour and amplitude intensity displays of the peak voltage envelope. A Sony TCM-5000 3-head audiocassette recorder was used as the input device. The speech samples were digitized at 20K Hz after processing through a low-pass seven-pole elliptical filter by a DigiSound 16 analog-to-digital converter (Micro Technology Unlimited, Raleigh, NC) with a conservative cutoff of 9.8K Hz. Fundamental frequency as a function of time is calculated as the reciprocal of the period between successive glottal pulses using Henke's FPRD algorithm. The intensity contour uses the method of the first difference, rectified and smoothed, and provides a peak voltage tracing for qualitative evaluation.

Response Definitions for Rate

To develop the response definitions for the four inappropriate rate codes and to assess the concurrent validity of the exemplars (see Table 1), acoustic rate characteristics were examined for a sample of 21 utterances. From among the hundreds of recorded utterances collected during the development of the PVSP, the 21 samples met the following criteria: (a) spoken by a child, (b) contained a minimum of four syllables, (c) did not meet any of the exclusion code criteria affecting rate (i.e., respiratory involvement, overtalk, reading, singing, repetitions, interfering noise, character register, narrative register, whisper, belch, cough, hiccup, laugh, or yawn), and (d) had not been tentatively categorized as inappropriate for rate by the first three authors. The fourth author obtained the following measures for each of the 21 utterances, using the original transcripts and the cursers and expanded waveform displays in *CSpeech* for segmentation and duration:

1. *Total Syllables*: Number of syllables in the utterance glossed and segmented by the examiner.
2. *Utterance Length*: Duration of utterance (in milliseconds) from onset to offset of waveform.
3. *Syllables Per Second*: Total Syllables divided by Utterance Length.
4. *Articulation Time*: Duration of the continuous energy portions of the utterance, minus breaks in the energy pattern within and between words.

5. *Pause Time*: Utterance Length minus Articulation Time.

6. *Syllables Per Articulation Time*: Total syllables divided by Articulation Time.

7. *Percent Pause Time*: Pause Time divided by Utterance Length.

8. *Percent Articulation Time*: 100 minus Percent Pause Time.

9. *Average Syllable Time*: Articulation Time divided by Total Syllables.

Table A1 [see Table A1, page 42] is a summary of descriptive data for the five rate measures. Relative to the 2-4 syllables per second generally considered normal rate for young children (Baken, 1987) and used for the perceptual decisions, these averaged instrumental data suggested that syllable rates below 2.2 syllables per second (syll/s) would be the more appropriate cutoff point for too slow and above 3.6 syll/s for too fast. A rate of 2.2 syll/s reflects a point two standard deviations below the average of 3.7 syll/s shown in Table A1, plus 10% for measurement error (i.e., 3.7 - (2 X .83) + 10% = 2.24 syll/s). Similarly, a rate of 3.6 syll/s reflects a 10% margin for measurement error (i.e., 4 syll/s, less 10%). The values for Percent Articulation Time and Average Syllable Time in Table 1 were used to divide slow utterances into PV9: Slow Articulation /Pause Time or PV10: Slow/Pause Time. If fewer than 50% of syllables in the utterance were 300 ms or longer (i.e., mean (248 ms) + one standard deviation (48.2)), the rate was determined to be PV10, that is, too slow due to pause time only. Instrumental analysis could not provide the information needed to differentiate PV11: Fast from PV12: Fast/Acceleration (i.e., fast speech containing at least one train of increasingly shorter syllables).

Criterion Validation for Stress

The waveform, pitch envelope, and intensity envelope routines of *VOCAL* were used to assess the concurrent validity of the stress exemplars. To establish a reference database for the acoustic correlates of appropriate stress, waveforms and pitch-intensity envelopes for 20 utterances perceived by the fourth author as grossly appropriate in stress were randomly selected from the audiocassette database. Visual examination of the glosses and acoustic records for these reference samples were thus available for comparison to exemplars selected for three individual PVSP inappropriate stress codes--PV13: Multisyllabic Word Stress, PV14: Reduced/Equal Stress, and PV15: Excessive/Equal/Misplaced Stress.

Criterion validation of 30 exemplars was undertaken. The 30 selected exemplars were free of involvement in other suprasegmentals that would obviate or complicate acoustic analysis of stress. Criterion validation of one of the three inappropriate stress codes in Table 1 was considered confirmed if visual examination of the acoustic displays indicated one or more of the following correlates: (a) abrupt increase in intensity, (b) extended duration, (c) extreme pitch alteration, (d) visible breaks in the pitch and/or intensity envelopes, (e) equality of the pitch and intensity peaks per syllable, and (f) inappropriate number or duration of pauses.

Figure A1 [see Figure A1, page 50] is an example of the *VOCAL* output used for criterion validation of an exemplar for PV13: Multisyllabic Word Stress. The upper left-hand panel is the waveform trace of the clinician-child interaction. The clinician asks, "[I wonder] if there are any other toys you have that are fun?" (250 to 2750 ms), and the child responds, "My little ponies." (3200 to 4800 ms). The lower left-hand panel is the frequency tracing of the utterance, with an abrupt upward frequency shift occurring on the second syllable of "ponies" (4200 ms). The two right-hand panels are waveform and frequency analysis expansions of the left-panels for the portion in which the inappropriate word stress occurs. Note in the lower right panel the increased intensity of this syllable (peak volts) and the rapid upward fundamental frequency change.

Figure A2 [see Figure A2, page 51] is an example of the *VOCAL* output used for criterion validation of a PVSP exemplar for PV14: Reduced/Equal Stress. The child was perceived to say "Alice and I watch it" with reduced/equal stress evident in pitch, loudness, and duration. This is confirmed in the *VOCAL* waveform (top panel) and frequency/intensity information (lower panel), indicating nearly flat frequency tracing of the first four syllables (175-1450 ms) with little variation in frequency, intensity, and duration.

The four panels in Figure A3 [see Figure A3, page 52] provide an example of the *VOCAL* output used for criterion validation of a PVSP exemplar for PV15: Excessive/ Equal/ Misplaced Stress (also coded PV9: Slow Articulation/Pause Time). The two left panels are the *VOCAL* waveform and frequency/ intensity trace for a 17-syllable utterance with an elapsed time of approximately 18 seconds. The waveform, frequency contours, and intensity envelopes all show sharp onsets, lack of coarticulation from syllable to syllable, and limited variation in intensity, duration, and frequency on each syllable. As shown in the right two panels, expansion of the segments occurring from 7000-11000 ms in the left panels clearly reveals the syllable-by-syllable pattern of speech production, with approximately equal-length pauses between syllables.

Response Definitions for Pitch

The pitch extraction routines of *CSpeech* were used to establish central tendencies for fundamental frequency in children and to assess the concurrent validity of the pitch exemplars. Syllable-level data on fundamental frequency was obtained at the midpoint of the vowel. The fourth author randomly selected 21 utterances from the audiocassette library meeting the following criteria: (a) spoken by a child, (b) contained a minimum of four syllables, (c) did not meet any of the exclusion code criteria affecting pitch (i.e., respiratory involvement, overtalk, reading, singing, repetitions, interfering noise, character register, narrative register, whisper, belch, cough, hiccup, laugh, or yawn), and (d) had not been judged abnormal for pitch.

Table A2 [see Table A2, page 43] is a summary of the fundamental frequency data derived from the total of 141 syllables in the 21-utterance sample. The mean fundamental frequency for the sample of syllables was 268 Hz (SD = 46 Hz). Based on these empirical findings the response definition for PV20: Low Pitch was set at 50% or more syllables in the utterance lower than one standard deviation below the reference mean (i.e., 268 Hz - 46 Hz = 222 Hz). Utterances meeting this criteria that also include 50% or more syllables below the 80 Hz cutoff established for fry register (Baken, 1987) were given the code PV19: Low Pitch/Glottal Fry. The code for PV22: High Pitch was set at the criterion of 50% or more syllables above 314 Hz (i.e., 268 Hz + 46 Hz = 314 Hz). Utterances meeting this criteria that also include 50% or more syllables approximately 500 Hz or above were assigned the code PV21: High Pitch/Falsetto.

Criterion Validation for Quality

Outputs from *CSpeech* were used for criterion validity assessment of two PVSP quality codes, PV26: Break/Shift/Tremulous and PV27: Register Break. Figure A4 [see Figure A4, page 53] is a *CSpeech* waveform (upper tracing) and a frequency envelope (lower tracing) illustrating a frequency shift sufficient to meet PVSP criteria for PV26 (as distinguished from pitch shifts associated with PV13: Multisyllabic Word Stress). The initial 638 ms of the utterance ". . . form a rainbow" has a steady pitch with a midpoint F_o of 179 Hz. There is an obvious upward pitch shift beginning at "bow" which averages 350 Hz.

Figure A5 [see Figure A5, page 54] is a CSpeech display that illustrates the acoustic confirmation of PV 27: Register Break. The first three words of the utterance, "Nope, they're not dead" has pitch levels of approximately 207 Hz. On the final word, an upward shift of approximately 278 Hz occurs, with this adolescent hearing-impaired boy's voice breaking into a falsetto range of approximately 485 Hz.

Table 1. Summary of findings from the instrumental criterion validity study.

Supra-segmental	Prosody-Voice Code	Exemplars: Number Trained	Exemplars: Number Sampled	Exemplars: % Sampled	Confirmations: Exact Number	Confirmations: Exact %	Confirmations: Within Class Number	Confirmations: Within Class %	Confirmations: Total
Rate		39	26	67%	20	77%	4	15%	92%
	Class: Too Slow	24	14	58%	10	71%	4	29%	100%
	PV9: Slow Articulation/ Pause Time	19	10	53%	8	80%	2	20%	100%
	PV10: Slow/Pause Time	5	4	80%	2	50%	2	50%	100%
	Class: Too Fast	15	12	80%	10	83%	0	--	83%
	PV11: Fast	6	5	83%	4	80%	0	--	80%
	PV12: Fast/Acceleration	9	7	78%	6	86%	0	--	86%
Stress		50	28	56%	25	89%	0	--	89%
	PV13: Multisyllabic Word Stress	3	3	100%	3	100%	0	--	100%
	PV14: Reduced/Equal Stress	8	4	50%	4	100%	0	--	100%
	PV15: Excessive/Equal/ Misplaced Stress	39	21	54%	18	86%	0	--	89%
Pitch		14	14	100%	10	71%	4	29%	100%
	Class: Low Pitch	6	6	100%	4	67%	2	33%	100%
	PV19: Low Pitch/Glottal Fry	4	4	100%	2	50%	2	50%	100%
	PV20: Low Pitch	2	2	100%	2	100%	0	--	100%
	Class: High Pitch	8	8	100%	6	75%	2	25%	100%
	PV21: High Pitch/Falsetto	5	5	100%	3	60%	2	40%	100%
	PV22: High Pitch	3	3	100%	3	100%	0	--	100%
Quality		33	25	81%	24	96%	0	--	96%
	PV26: Break/Shift/Tremulous	31	23	74%	22	96%	0	--	96%
	PV27: Register Break	2	2	100%	2	100%	0	--	100%

Table 2. Summary of findings from the perceptual criterion validity study.

Supra-segmental	PV Code	Number of Utterances			Percentage of Agreement[a]							
					Judge 1		Judge 2		Judge 3		Mean Judges	
		Exemplars	Foils	Total	E[b]	E+F[b]	E	E+F	E	E+F	E	E+F
Pitch		19	2	21	72.1	77.1	74.6	77.1	72.1	77.1	72.9	77.1
	PV19: Low Pitch/ Glottal Fry	4	0	4	75.0	75.0	75.0	75.0	75.0	75.0	75.0	75.0
	PV20: Low Pitch	3	1	4	50.0	66.7	100	100	50.0	66.7	66.7	77.8
	PV21: High Pitch/ Falsetto	6	1	7	80.0	83.3	40.0	50.0	80.0	83.3	66.7	72.2
	PV22: High Pitch	6	0	6	83.3	83.3	83.3	83.3	83.3	83.3	83.3	83.3
Quality		58	16	74	97.7	97.0	86.1	85.1	73.3	77.0	85.7	86.4
	PV23: Breathy	9	1	10	100	100	62.5	66.7	75.0	77.8	79.2	81.5
	PV24: Rough	33	11	44	90.9	87.9	81.8	84.8	68.2	69.7	80.3	80.8
	PV25: Strained	7	1	8	100	100	100	100	66.7	71.4	88.9	90.5
	PV28: Diplophonia	9	3	12	100	100	100	88.9	83.3	88.9	94.4	92.6
Resonance		59	13	72	81.4	74.1	73.5	74.8	68.7	74.7	74.5	74.5
	PV30: Nasal	19	5	24	71.4	68.4	57.1	57.9	64.3	73.7	64.3	66.7
	PV31: Denasal	17	5	22	83.3	76.5	83.3	88.2	66.7	76.5	77.8	80.4
	PV32: Naso-pharyngeal	23	3	26	89.5	77.3	80.0	78.3	75.0	73.9	81.5	76.5
Totals		136	31	167	83.9	83.5	78.5	79.4	71.6	76.4	78.0	79.8

[a] All values are percentages
[b] Exemplars (E); Exemplars plus Foils (E+F)

Table 3. Summary of the point-to-point percentage of agreement data for the three judges for Prosody-Voice coding in Interjudge Reliability Study I. See text for explanation of the two stringency conditions.

	Agreement on Appropriate and Inappropriate[a]				Exact agreement on Inappropriate[b]			
Judge	Test 1	Test 2	Test 3	Mean	Test 1	Test 2	Test 3	Mean
1	78.3%	88.0%	87.4%	84.6%	94.5%	73.8%	62.1%	76.8%
2	84.5%	86.6%	88.9%	86.7%	77.5%	85.7%	89.0%	84.1%
3	84.6%	84.8%	87.1%	85.5%	86.1%	84.6%	80.8%	83.8%
Mean	82.5%	86.5%	87.8%	85.6%	80.0%	81.4%	77.3%	81.5%

[a] Based on 25-30 exemplars per test.

[b] Based on 2-12 exemplars per test.

Table 4. Mean point-to-point interjudge agreement for exclusion and prosody-voice coding in Examiner Reliability Study I.

Variable	Agreement on Appropriate and Inappropriate	Exact agreement on Inappropriate
Exclusion Codes	87.9%	83.8%
Prosody-Voice Codes	86.5%	81.8%
Prosody	85.7%	77.2%
Phrasing	96.2%	69.1%
Rate	81.6%	77.4%
Stress	79.4%	85.2%
Voice	87.2%	86.4%
Loudness	92.6%	100%
Pitch	89.1%	77.8%
Quality	79.8%	81.3%
Laryngeal	82.1%	76.4%
Resonance	77.4%	86.1%

Table 5. Intrajudge and interjudge exclusion coding reliability based on 28 conversational speech samples assessed in Examiner Reliability Study II.

	Agreement on Exclusion Code vs. Prosody-Voice Code			Agreement on Class of the Exclusion Code			Agreement on the Exact Exclusion Code		
	M	SD	Range	M	SD	Range	M	SD	Range
Intrajudge	96.8%	3.6	85.7%-100%	95.4%	6.8	76.5%-100%	91.7%	5.3	84.2%-100%
Interjudge	91.7%	4.7	81.0%-100%	95.9%	5.8	81.0%-100%	90.6%	6.8	76.5%-100%

Table 6. Intrajudge and interjudge prosody-voice coding reliability in Examiner Reliability Study II.

PV Code	Screening Agreement									Exact Agreement					
	Appropriate			Inappropriate			Overall			Inappropriate			Overall		
	n Samples	n Utterances	Average	n Samples	n Utterances	Average	n Samples	n Utterances	Average	n Samples	n Utterances	Average	n Samples	n Utterances	Average
Intrajudge															
Phrasing	28	535	97.4%	25	102	94.1%	28	634	96.2%	25	96	82.3%	28	633	94.0%
Rate	28	597	98.5%	12	37	45.9%	28	634	95.4%	8	17	88.2%	25	634	95.1%
Stress	28	547	97.3%	20	79	34.2%	28	628	90.6%	14	27	92.6%	28	634	89.4%
Loudness	28	580	97.8%	11	54	79.6%	28	634	96.2%	9	43	100%	28	634	96.2%
Pitch	28	623	99.7%	3	9	44.4%	28	634	98.9%	2	4	100%	28	634	98.9%
Quality															
Laryngeal	25	496	92.1%	24	137	63.5%	28	634	85.3%	18	87	90.8%	28	634	85.0%
Resonance	27	553	89.2%	13	80	92.5%	28	634	89.4%	10	74	100%	28	634	89.6%
Interjudge															
Phrasing	28	529	96.6%	24	103	82.5%	28	632	94.1%	23	87	60.9%	28	632	89.1%
Rate	28	592	93.8%	12	35	65.7%	28	632	91.6%	11	23	95.7%	28	632	89.7%
Stress	28	533	85.7%	22	93	51.5%	28	632	80.1%	17	50	84.0%	28	632	77.2%
Loudness	28	579	97.2%	10	53	54.7%	28	632	93.7%	6	29	100%	28	632	93.7%
Pitch	28	632	97.1%	3	9	22.2%	28	632	96.0%	2	2	100%	28	632	96.0%
Quality															
Laryngeal	25	476	89.3%	22	146	52.1%	28	632	80.9%	14	77	46.8%	28	632	74.2%
Resonance	26	512	91.2%	12	117	96.6%	28	632	91.6%	9	113	100%	28	632	91.9%

Table 7. Internal consistency of PVSP scores.[a]

Variable	Split-Half Coefficients: Speech-Normal *D*	*p*	Speech-Disordered *D*	*p*	All *D*	*p*	Part-Whole Coefficients: Speech-Normal *D*	*p*	Speech-Disordered *D*	*p*	All *D*	*p*
Prosody												
Phrasing	.409	.0107	.197	.2194	.301	.0074*	.834	.0001*	.799	.0001*	.816	.0001*
Rate	.415	.0095*	.519	.0012*	.534	.0001*	.785	.0001*	.866	.0001*	.873	.0001*
Stress	.048	.7629	.695	.0001*	.514	.0001*	.579	.0003*	.927	.0001*	.847	.0001*
Voice												
Loudness	.591	.0002*	.620	.0001*	.605	.0001*	.873	.0001*	.842	.0001*	.857	.0001*
Pitch	1.00	.0001*	.827	.0001*	.822	.0001*	1.00	.0001*	1.00	.0001*	1.00	.0001*
Quality												
Laryngeal	.611	.0001*	.778	.0001*	.707	.0001*	.859	.0001*	.899	.0001*	.881	.0001*
Resonance	.724	.0001*	.903	.0001*	.832	.0001*	.884	.0001*	.995	.0001*	.957	.0001*

[a]All coefficients are Spearman Rho (*D*) values corrected for ties (Siegel and Castellan, 1988) based on sample sizes of 40 speech-normal and 40 speech-disordered children.
* $p < .01$

Table 8. Internal consistency of PVSP screening decision outcomes.[a]

	Split-Half Coefficients												Part-Whole Coefficients											
	Speech-Normal				Speech-Disordered				All				Speech-Normal				Speech-Disordered				All			
	Pass		Fail		Pass		Fail		Pass		Fail		Pass		Fail		Pass		Fail		Pass		Fail	
Variable	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Prosody																								
Phrasing	19	78.9	21	52.4	25	64.0	15	26.7	44	70.5	36	41.7	19	78.9	21	90.5	25	80.0	15	80.0	44	79.5	36	86.1
Rate	35	97.1	5	40.0	38	97.4	2	50.0	73	97.3	7	42.9	35	97.4	5	60.0	38	97.4	2	50.0	73	97.3	7	57.1
Stress	39	97.4	1	0	28	96.5	12	75.0	67	97.0	13	69.2	39	100	1	0	28	96.5	12	83.3	67	98.5	13	76.9
Voice																								
Loudness	37	97.3	3	66.7	32	96.9	8	50.0	69	97.1	11	54.5	37	100	3	66.7	32	100	8	75.0	69	100	11	72.7
Pitch	40	100	--	--	40	100	--	--	80	100	--	--	40	100	--	--	40	100	--	--	80	100	--	--
Quality																								
Laryngeal	29	100	11	54.5	22	81.8	18	83.3	51	92.2	29	72.4	29	100	11	54.5	22	90.9	18	94.4	51	96.1	29	79.3
Resonance	34	94.1	6	33.3	31	96.8	9	77.8	65	95.4	15	60.0	34	100	6	50.0	31	96.8	9	100	65	98.5	15	80.0

[a]All table entries are the percentage of subjects who have the same screening decision outcome based on either computation. There were 40 subjects in each speech-status group.

Table 9. Efficiency data for scoring the PVSP.

	Speech-Normal					Speech-Disordered					All				
Variable	n	M	SD	Range	Total	n	M	SD	Range	Total	n	M	SD	Range	Total
Time required to score PVSP (minutes)	114	28.8	14.7	15.0-90.0		66	64.7	21.5	20.0-140.0		180	42.0	24.6	15.0-140.0	
Number of utterances required to obtain 24 codeable utterances including 3 codeable warmups	115	44.1	12.3	25.0-96.0		137	59.9	26.2	21.0-196.0		252	52.7	22.4	21.0-196.0	
Percentage of utterances containing 4 or more words	115	76.1%	13.1%	38.5%-100%		137	66.6%	15.6%	14.3%-95.8%		252	70.9%	15.6%	14.9%-100%	
Percentage of samples perceived as difficult to score and source of difficulty:															
No comment	106				93.0%	54				76.1%	160				86.5%
Difficult due to external factors	7				6.1%	7				9.9%	14				7.6%
Difficult due to severity of involvement	1				0.9%	7				9.9%	8				4.3%
Difficult due to both factors above	0				0%	3				4.2%	3				1.6%

Table 10. Reference data for the frequency of occurrence of the 31 Exclusion Codes in 115 speech-normal children and 137 speech-delayed children.

Code	Proportional Occurrence[a]									Percentage of Children[b]								
	Speech-normal			Speech-delayed			Both			Speech-normal			Speech-delayed			Both		
	M	SD	Range	M	SD	Range	M	SD	Range	0%	1-4.9%	$\geq$5%	0%	1-4.9%	$\geq$5%	0%	1-4.9%	$\geq$5%
Content/Context																		
C1: Automatic sequential	0.07	0.39	0-2.3	0.09	0.40	0-2.4	0.08	0.39	0-2.4	96.5	3.5	0	94.9	5.1	0	95.6	4.4	0
C2: Back channel/aside	0.17	0.63	0-3.0	0.25	1.15	0-10.4	0.21	0.95	0-10.4	93.0	7.0	0	92.7	5.8	1.5	92.9	6.4	0.8
C3: I don't know	1.36	2.19	0-9.3	1.07	2.05	0-13.2	1.20	2.12	0-13.2	61.7	27.8	10.4	65.7	30.7	3.7	63.9	29.4	6.8
C4: Imitation	0	--	--	0.19	0.78	0-5.0	0.10	0.58	0-5.0	100	0	0	93.4	5.8	0.7	96.4	3.2	0.4
C5: Interruption/overtalk	1.81	2.54	0-14.3	1.50	2.45	0-15.9	1.64	2.49	0-15.9	52.2	33.9	13.9	56.2	38.0	5.8	54.4	36.1	9.5
C6: Not 4(+) words	0.48	3.43	0-33.3	1.86	6.52	0-47.4	1.23	5.37	0-47.4	96.5	0.9	2.6	83.9	5.1	11.0	89.7	3.2	7.1
C7: Only one word	21.65	11.92	0-48.3	27.09	13.36	0-63.6	24.60	12.99	0-63.6	2.6	3.5	93.9	2.2	0.7	97.1	2.4	2.0	95.6
C8: Only person's name	0.56	1.54	0-8.6	0.89	1.87	0-10.0	0.74	1.73	0-10.0	85.2	10.4	4.4	71.5	23.4	5.1	77.8	17.5	4.8
C9: Reading	0.02	0.20	0-2.1	0.01	0.06	0-0.8	0.01	0.14	0-2.1	99.1	0.9	0	99.3	0.7	0	99.2	0.8	0
C10: Singing	0.11	0.65	0-5.0	0.13	0.72	0-6.1	0.12	0.69	0-6.1	96.5	2.6	0.9	94.9	4.4	0.7	95.6	3.6	0.8
C11: Second repetition	0.02	0.22	0-2.3	0.10	0.60	0-6.1	0.06	0.47	0-6.1	99.1	0.9	0	96.4	2.9	0.7	97.6	2.0	0.4
C12: Too many unintelligibles	2.17	3.30	0-19.6	6.22	6.68	0-43.7	4.35	5.77	0-43.7	53.9	30.4	15.7	23.4	31.4	45.3	37.3	31.0	31.8
Environment																		
E1: Interfering noise	0.44	1.03	0-6.5	0.73	1.90	0-13.6	0.60	1.57	0-13.6	81.7	17.4	0.9	75.2	21.9	2.9	78.2	19.8	2.0
E2: Recorder wow/flutter	0.07	0.44	0-3.5	0.56	5.43	0-63.2	0.33	4.01	0-63.2	97.4	2.6	0	95.6	2.9	1.5	96.4	2.8	0.8
E3: Too close to microphone	0.02	0.18	0-1.9	0.41	1.11	0-9.0	0.23	0.85	0-9.0	99.1	0.9	0	81.0	17.5	1.5	89.3	9.9	0.8
E4: Too far from microphone	0.49	1.76	0-12.5	1.37	3.08	0-27.3	0.97	2.59	0-27.3	88.7	8.7	2.6	62.8	29.2	8.0	74.6	19.8	5.6
Register																		
R1: Character register	0.36	2.02	0-20.0	0.94	5.84	0-64.4	0.67	4.52	0-64.4	93.9	4.4	1.7	89.1	8.0	2.9	91.3	6.4	2.4
R2: Narrative register	5.47	13.41	0-55.8	7.57	14.06	0-70.6	6.61	13.78	0-70.6	80.0	2.6	17.4	58.4	11.7	29.9	68.3	7.5	24.2
R3: Negative register	0.10	1.07	0-11.5	0.22	0.88	0-6.3	0.17	0.97	0-11.5	99.1	0	0.9	92.0	7.3	0.7	95.2	4.0	0.8
R4: Sound effects	0.27	1.04	0-7.0	0.22	0.80	0-5.6	0.24	0.91	0-7.0	92.2	6.1	1.7	90.5	8.0	1.5	91.3	7.1	1.6
R5: Whisper	0.65	3.59	0-35.8	0.26	1.24	0-12.5	0.44	2.59	0-35.8	89.6	8.7	1.7	91.2	8.0	0.7	90.5	8.3	1.2

Table 10. Continued.

	Proportional Occurrence[a]									Percentage of Children[b]								
	Speech-normal			Speech-delayed			Both			Speech-normal			Speech-delayed			Both		
Code	M	SD	Range	M	SD	Range	M	SD	Range	0%	1-4.9%	≥5%	0%	1-4.9%	≥5%	0%	1-4.9%	≥5%
States																		
S1: Belch	0	--	--	0.03	0.28	0-2.8	0.02	0.21	0-2.8	100	0	0	98.5	1.5	0	99.2	0.8	0
S2: Cough/throat clear	0.02	0.21	0-2.2	0.11	0.58	0-4.4	0.07	0.45	0-4.4	99.1	0.9	0	96.4	3.7	0	97.6	2.4	0
S3: Food in mouth	0.04	0.32	0-2.7	0.02	0.19	0-2.3	0.03	0.26	0-2.7	98.3	1.7	0	99.3	0.7	0	98.8	1.2	0
S4: Hiccup	0.11	0.83	0-8.3	0	--	--	0.05	0.57	0-8.3	97.4	1.7	0.9	100	0	0	98.8	0.8	0.4
S5: Laugh	1.45	4.04	0-25.9	1.26	2.78	0-15.5	1.34	3.40	0-25.9	75.7	16.5	7.8	74.5	14.6	11.0	75.0	15.5	9.5
S6: Lip smack	0.03	0.30	0-3.2	0.03	0.38	0-4.4	0.03	0.35	0-4.4	99.1	0.9	0	99.3	0.7	0	99.2	0.8	0
S7: Body movement	0.12	0.61	0-4.6	0.03	0.22	0-1.7	0.07	0.45	0-4.6	95.7	4.4	0	97.8	2.2	0	96.8	3.2	0
S8: Sneeze	0	--	--	0.04	0.40	0-4.4	0.02	0.29	0-4.4	100	0	0	98.5	1.5	0	99.2	0.8	0
S9: Telegraphic	0.11	0.57	0-4.2	0.42	1.17	0-8.1	0.28	0.95	0-8.1	95.7	4.4	0	83.2	15.3	1.5	88.9	10.3	0.8
S10: Yawn	0.04	0.31	0-2.5	0.08	0.44	0-4.1	0.06	0.39	0-4.1	98.3	1.7	0	95.6	4.4	0	96.8	3.2	0

[a] Entries are the percentages of code occurrence based on each child's total number of utterances in the speech sample.

[b] For each code, children were categorized into one of three groups; the table entries indicate the percentage of children in each group: children who had 0% of their utterances meeting criteria for the exclusion code, 1%-4.9% of their utterances meeting criteria for the exclusion code, or 5% or more of their utterances meeting criteria for the exclusion code.

Table 11. Reference data for the frequency of occurrence of the 31 inappropriate Prosody-Voice Codes in 115 speech-normal children and 137 speech-delayed children.

	Proportional Occurrence[a]									Percentage of Children[b]								
	Speech-normal			Speech-delayed			Both			Speech-normal			Speech-delayed			Both		
Code	M	SD	Range	M	SD	Range	M	SD	Range	0%	1-4.9%	≥5%	0%	1-4.9%	≥5%	0%	1-4.9%	≥5%
Phrasing																		
PV2: Sound/syllable repetition	1.57	2.13	0-11.1	1.83	2.66	0-16.1	1.71	2.43	0-16.1	54.8	38.3	7.0	45.3	46.0	8.8	49.6	42.5	7.9
PV3: Word repetition	2.55	3.10	0-20.0	1.73	3.24	0-26.5	2.11	3.19	0-26.5	39.1	42.6	18.3	54.0	38.0	8.0	47.2	40.1	12.7
PV4: Sound/syllable and word repetition	0.22	1.20	0-10.5	0.33	1.35	0-9.8	0.28	1.28	0-10.5	94.8	3.5	1.7	90.5	7.3	2.2	92.5	5.6	2.0
PV5: More than one word repetition	0.25	0.91	0-6.7	0.40	1.22	0-8.9	0.33	1.09	0-8.9	91.3	7.8	0.9	85.4	12.4	2.2	88.1	10.3	1.6
PV6: One word revision	1.39	2.02	0-10.0	0.50	1.05	0-5.5	0.91	1.63	0-10.0	58.3	36.5	5.2	77.4	21.9	0.7	68.7	28.6	2.8
PV7: More than one word revision	0.19	0.68	0-3.4	0.07	0.38	0-2.8	0.13	0.54	0-3.4	92.2	7.8	0	96.4	3.7	0	94.4	5.6	0
PV8: Repetition and revision	0.65	1.70	0-11.4	0.21	0.71	0-4.4	0.41	1.28	0-11.4	80.0	18.3	1.7	90.5	9.5	0	85.7	13.5	0.8
Rate																		
PV9: Slow articulation/ pause time	0.04	0.46	0-4.9	0.71	3.53	0-34.9	0.40	2.64	0-34.9	99.1	0.90	0	86.9	10.2	2.9	92.5	6.0	1.6
PV10: Slow/pause time	0.11	0.61	0-5.0	0.17	0.59	0-2.9	0.14	0.60	0-5.0	96.5	2.6	0.9	91.2	8.8	0	93.7	6.0	0.4
PV11: Fast	1.55	6.47	0-56.3	0.59	2.57	0-20.4	1.03	4.78	0-56.3	85.2	8.7	6.1	86.9	10.2	2.9	86.1	9.5	4.4
PV12: Fast/acceleration	0.06	0.45	0-3.7	0	--	--	0.03	0.30	0-3.7	98.3	1.7	0	100	0	0	99.2	0.8	0
Stress																		
PV13: Multisyllabic word stress	0.06	0.50	0-4.8	0	--	--	0.03	0.34	0-4.8	98.3	1.7	0	100	0	0	99.2	0.8	0
PV14: Reduced/equal stress	0.02	0.23	0-2.5	0.48	2.46	0-21.8	0.27	1.83	0-21.8	99.1	0.9	0	89.8	8.0	2.2	94.1	4.8	1.2
PV15: Excessive/equal/ misplaced stress	0.85	1.74	0-10.0	5.03	9.75	0-50.0	3.12	7.57	0-50.0	74.8	20.9	4.4	48.9	28.5	22.6	60.7	25.0	14.3
PV16: Multiple stress features	0	--	--	0	--	--	0	--	--	100	0	0	100	0	0	100	0	0

Table 11. Continued

Code	Proportional Occurrence[a]									Percentage of Children[b]								
	Speech-normal			Speech-delayed			Both			Speech-normal			Speech-delayed			Both		
	M	SD	Range	M	SD	Range	M	SD	Range	0%	1-4.9%	≥5%	0%	1-4.9%	≥5%	0%	1-4.9%	≥5%
Loudness																		
PV17: Soft	1.06	4.27	0-35.0	2.77	7.54	0-60.0	1.99	6.31	0-60.0	85.2	9.6	5.2	65.7	19.0	15.3	74.6	14.7	10.7
PV18: Loud	0.05	0.37	0-3.1	0.91	3.23	0-27.9	0.52	2.43	0-27.9	98.3	1.7	0	82.5	11.7	5.8	89.7	7.1	3.1
Pitch																		
PV19: Low pitch/ glottal fry	0.27	1.32	0-12.3	0.28	1.44	0-12.7	0.27	1.38	0-12.7	93.9	5.2	0.9	94.2	3.7	2.2	94.1	4.4	1.6
PV20: Low pitch	0	--	--	0.04	0.42	0-4.9	0.02	0.31	0-4.9	100	0	0	99.3	0.7	0	99.6	0.4	0
PV21: High pitch/falsetto	0	--	--	0	--	--	0	--	--	100	0	0	100	0	0	100	0	0
PV22: High pitch	0	--	--	0.04	0.42	0-4.8	0.02	0.31	0-4.8	100	0	0	98.5	1.5	0	99.2	0.8	0
Quality																		
Laryngeal features																		
PV23: Breathy	0.12	0.58	0-3.9	0.32	1.25	0-10.4	0.23	1.00	0-10.4	95.7	4.4	0	89.8	8.0	2.2	92.5	6.4	1.2
PV24: Rough	4.96	14.13	0-81.5	8.23	14.78	0-71.0	6.74	14.55	0-81.5	66.1	16.5	17.4	48.2	20.4	31.4	56.4	18.7	25.0
PV25: Strained	0.18	0.63	0-2.9	0.30	1.79	0-19.1	0.25	1.39	0-19.1	92.2	7.8	0	94.2	4.4	1.5	93.3	6.0	0.8
PV26: Break/shift/ tremulous	0.80	1.75	0-7.9	1.04	1.87	0-11.8	0.93	1.82	0-11.8	78.3	15.7	6.1	65.0	29.9	5.1	71.0	23.4	5.6
PV27: Register break	0	--	--	0	--	--	0	--	--	100	0	0	100	0	0	100	0	0
PV28: Diplophonia	0	--	--	0	--	--	0	--	--	100	0	0	100	0	0	100	0	0
PV29: Multiple laryngeal features	0.19	0.90	0-6.7	0.95	4.29	0-40.0	0.60	3.24	0-40.0	94.8	3.5	1.7	86.1	10.2	3.7	90.1	7.1	2.8
Resonance features																		
PV30: Nasal	0.18	1.35	0-12.9	0.60	2.50	0-25.4	0.41	2.06	0-25.4	97.4	0.9	1.7	85.4	11.7	2.9	90.9	6.8	2.4
PV31: Denasal	2.89	8.98	0-48.3	2.44	8.13	0-55.6	2.65	8.52	0-55.6	81.7	6.1	12.2	80.3	8.0	11.7	81.0	7.1	11.9
PV32: Nasopharyngeal	0	--	--	4.68	15.21	0-70.6	2.55	11.44	0-70.6	100	0	0	90.5	0	9.5	94.8	0	5.2

[a] Entries are the percentages of code occurrence based on each child's total number of utterances in the speech sample.

[b] For each code, children were categorized into one of three groups; the table entries indicate the percentage of children in each group: children who had 0% of their utterances meeting criteria for the exclusion code, 1%-4.9% of their utterances meeting criteria for the exclusion code, or 5% or more of their utterances meeting criteria for the exclusion code.

Table 12. Reference data for the appropriateness of the six PVSP suprasegmentals in 115 speech-normal children and 137 speech-delayed children.

	Proportional Occurrence[a]									Percentage of Children[b]								
	Speech-normal			Speech-delayed			Both			Speech-normal			Speech-delayed			Both		
Code	M	SD	Range	M	SD	Range	M	SD	Range	≥90%	80-89.9%	<80%	≥90%	80-89.9%	<80%	≥90%	80-89.9%	<80%
Prosody																		
Phrasing	88.20	10.20	29.2-100	88.88	11.47	29.2-100	88.57	10.90	29.2-100	53.9	28.7	17.4	58.4	27.0	14.6	56.4	27.8	15.9
Rate	97.34	9.09	25.0-100	96.53	10.01	8.3-100	96.90	9.59	8.3-100	93.0	2.6	4.4	93.4	2.2	4.4	93.3	2.4	4.4
Stress	98.14	3.78	77.8-100	87.26	21.45	0-100	92.22	16.89	0-100	96.5	1.7	1.7	70.8	8.8	20.4	82.5	5.6	11.9
Voice																		
Loudness	96.25	15.14	0-100	89.73	19.73	0-100	92.70	18.05	0-100	94.8	0.9	4.4	73.0	13.9	13.1	82.9	7.9	9.1
Pitch	99.53	2.91	70.8-100	99.16	3.86	66.7-100	99.33	3.46	66.7-100	99.1	0	0.9	97.1	2.2	0.7	98.0	1.2	0.8
Quality	83.51	26.20	0-100	60.78	36.64	0-100	71.15	34.17	0-100	64.4	11.3	24.4	29.9	16.1	54.0	45.6	13.9	40.5
Laryngeal	89.20	21.70	0-100	75.34	31.41	0-100	81.66	28.22	0-100	77.4	7.8	14.8	49.6	16.1	34.3	62.3	12.3	25.4
Resonance	94.30	15.66	8.3-100	83.38	31.95	0-100	88.36	26.35	0-100	87.8	3.5	8.7	73.7	5.1	21.2	80.2	4.4	15.5

[a] Entries are the percentage of utterances coded as appropriate, based on each child's total number of utterances eligible for prosody-voice coding.

[b] Percentage of children who had 90% or more appropriate utterances (pass), from 80% to 89.9% appropriate utterances (questionable fail), or below 80% appropriate utterances (fail).

Table A1. Reference data for Rate codes.[a]

Rate Measure	M	SD	Range
Syllables Per Second	3.7	.83	2.6-6.0
Syllables Per Articulation Time	4.2	.97	2.8-7.2
Percent Articulation Time(%)	87.4	12.5	64-100
Percent Pause Time(%)	12.6	12.5	0-36
Average Syllable Time (ms)	248.0	48.2	138.5-353.5

[a] Based on 21 conversational speech samples.

Table A2. Reference data for Pitch and Quality codes.[a]

Variable	Number of Syllables	M	SD	Range
Fundamental Frequency (Hz)	141	267.90	46.10	198.0-443.3
Jitter (Hz)	141	0.027	0.014	0.010-0.088
Shimmer (%)	141	3.230	1.710	1.21-12.48
Signal-to-Noise Ratio (dB)	140	23.81	4.540	12.74-33.31

[a] Based on 21 conversational speech samples.

FIGURE CAPTIONS

Figure 1. Sample of the first and second pages of the *Prosody-Voice Screening Profile (PVSP)* scoring form. The left panel illustrates the summative PVSP profile. The right panel includes the key to the 31 PVSP exclusion codes and the 31 inappropriate prosody-voice codes. Portions of the original handwritten text and tallies are not legible in this reduced illustration.

Figure 2. Sample of the third pages and fourth pages of the *Prosody-Voice Screening Profile (PVSP)* scoring form. The left panel includes the grids used to tally the utterance-by-utterance occurrence of Exclusion and Prosody-Voice codes. The right panel illustrates the emphasis in the PVSP on interpretation of the quantitative screening information. Portions of the original handwritten text and tallies are not legible in this reduced illustration.

Figure 3. Sample Prosody-Voice Profile illustrating research use of data from the *Prosody-Voice Screening Profile (PVSP)*. The open circles are profiles for a group of 71 3-5 year-old children with normally developing (N) speech (Miller, 1990). The filled circles are the mean PVSP profiles for a group of 57 3-5 year-old children with delayed speech (D) of unknown origin (Shriberg & Kwiatkowski, in submission).

Figure 4. *Prosody-Voice Screening Profile (PVSP)* findings for two samples of children with speech delays of unknown origin. The filled bars are the percentages of 90 children who had *fails* or *questionable fails* on the protoversion of the PVSP (Shriberg & Kwiatkowski, 1982; Shriberg, Kwiatkowski, Best, Hengst, & Terselic-Weber, 1986). The cross-hatched bars are the percentages of *fails* and *questionable fails* for a recent group of 57 children assessed with the PVSP (Shriberg & Kwiatkowski, in submission).

Figure 5. *Prosody-Voice Screening Profile (PVSP)* findings for two groups of speech-involved children. The filled circles are the Prosody-Voice Profile profiles for 64 3-5 year-old children with speech delays of unknown origin. The open circles are profiles for 14 5-15 year-old children with suspected developmental apraxia of speech (Shriberg, Aram, and Kwiatkowski, in preparation).

Figure A1. *VOCAL* output used for criterion validation of a PVSP exemplar for PV13: Multisyllabic Word Stress. See text for explanation of the four panels.

Figure A2. *VOCAL* output used for criterion validation of a PVSP exemplar for PV14: Reduced/Equal Stress. See text for explanation of the two panels.

Figure A3. *VOCAL* output used for criterion validation of a PVSP exemplar for PV15: Excessive/Equal/Misplaced Stress. See text for explanation of the four panels.

Figure A4. *CSpeech* output used for criterion validity of a PVSP exemplar for PV26: Break/Shift/Tremulous. See text for explanation.

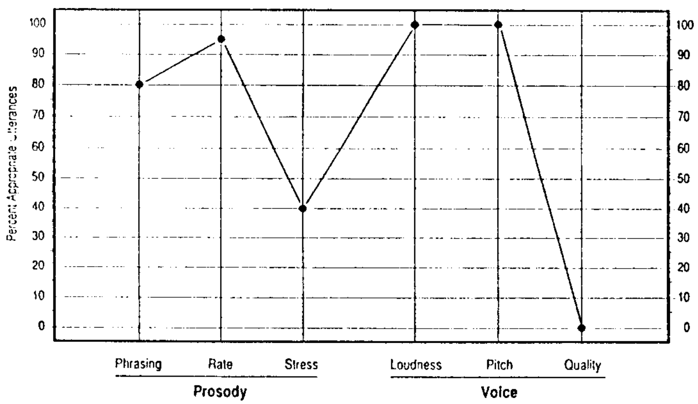
Figure A5. *CSpeech* output used for criterion validity of a PVSP exemplar for PV27: Register Break. See text for explanation.

Prosody-Voice Screening Profile (PVSP) Scoring Form

Identification

Name C	Tape Label CMRO-C1	Telegraphic Speech / Respiratory Involvement
Age 12;10 M ✓ F		**Screening Outcome**
Sample Date 7 30 90	Side A ___ B ✓	Pass / Fail
Sample Type Free Conversation	**Counter Number**	Phrasing 80% / 20%
Examiner RG	Beginning of Sample 184	Rate 95% / 5%
Scoring Date 7-31-90	First Coded Utterance 187	Stress 40% / 60%
Scorer RG	Last Coded Utterance 223	Loudness 100% / 0%
		Pitch 100% / 0%
		Quality 0% / 100%

Profile



Exclusion Codes

Content/Context	Environment	Register	States
C1 Automatic Sequential	E1 Interfering Noise	R1 Character Register	S1 Belch
C2 Back Channel/Aside	E2 Recorder Wow/Flutter	R2 Narrative Register	S2 Cough/Throat Clear
C3 I Don't Know	E3 Too Close to Microphone	R3 Negative Register	S3 Food in Mouth
C4 Imitation	E4 Too Far from Microphone	R4 Sound Effects	S4 Hiccup
C5 Interruption/Overtalk		R5 Whisper	S5 Laugh
C6 Not 4 (+) Words			S6 Lip Smack
C7 Only One Word			S7 Body Movement
C8 Only Person's Name			S8 Sneeze
C9 Reading			S9 Telegraphic
C10 Singing			S10 Yawn
C11 Second Repetition			
C12 Too Many Unintelligibles			

Prosody-Voice Codes

Prosody

Phrasing	Rate	Stress
1 Appropriate	1 Appropriate	1 Appropriate
2 Sound/Syllable Repetition II	9 Slow Articulation/Pause Time	13 Multisyllabic Word Stress
3 Word Repetition	10 Slow/Pause Time	14 Reduced/Equal Stress I
4 Sound/Syllable and Word Repetition	11 Fast	15 Excessive/Equal/Misplaced Stress
5 More than One Word Repetition	12 Fast/Acceleration	16 Multiple Stress Features
6 One Word Revision I		
7 More than One Word Revision		
8 Repetition and Revision I		

Voice

Loudness	Pitch	Quality: Laryngeal Features	Quality: Resonance Features
1 Appropriate	1 Appropriate	1 Appropriate	1 Appropriate
17 Soft	19 Low Pitch/Glottal Fry	23 Breathy	30 Nasal
18 Loud	20 Low Pitch	24 Rough	31 Denasal
	21 High Pitch/Falsetto	25 Strained	32 Nasopharyngeal
	22 High Pitch	26 Break/Shift/Tremulous	
		27 Register Break	
		28 Diplophonia	
		29 Multiple Laryngeal Features	

Figure 1

Sequential Utterance Log

1 [illegible]	2 [illegible]	3 C[illegible]	4 1	5 2	6 3	7 C7	8 4	9 5	10 6	11 7	12 8	13 9	14 C7	15 C3	16 10	17 C7	18 C7	19 11	20 12
21 C[illegible]	22 C7	23 13	24 14	25 15	26 16	27 C7	28 C7	29 C7	30 C7	31 17	32 C7 C5	33 C7	34 18	35 19	36 20	37	38	39	40
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

Prosody-Voice Coding Log

Utterance Data			Codes													Notes
			Prosody						Voice							
			Phrasing		Rate		Stress		Loudness		Pitch		Quality			
Coded Utterance Number	4(+) Words (✓)	Counter Number	Approp (1)	Not Approp (code)	Approp (1)	Not Approp (code)	Approp (1)	Not Approp (code)	Approp (1)	Not Approp (code)	Approp (1)	Not Approp (code)	Approp (1)	Laryngeal (code)	Resonance (code)	
1	✓	187	1		1			15	1		1				32	misplaced stress
2	✓		1		1			15	1		1				32	
3	✓			6	1			15	1		1				32	
4			1		1		1		1		1				32	
5	✓			8	1			15	1		1			26	32	tremulous "swing"
6	✓		1		1	(11)	1		1		1				32	
7	✓		1		1		1		1		1				32	
8				2	1		1		1		1				32	% *
9			1		1		1		1		1				32	emphatic stress "it"
10	✓		1			11		15	1		1				32	
11	✓			2	1			15	1		1				32	
12	✓		1		1			15	1		1				32	
13	✓		1		1		1		1		1			26	32	voice break "and"
14			1		1		1	(15)	1		1				32	
15	✓		1		1			14	1		1				32	
16			1		1		1		1		1				32	
17	✓		1		1			15	1		1			(26)	32	borderline tremulous "then you"
18	✓		1		1			15	1		1				32	
19	✓		1		1			15	1		1				32	
20			1		1			15	1		1				32	
21																
22																
23																
24																
25																

	Phrasing	Rate	Stress	Loudness	Pitch	Quality
Numerator / Denominator	16/20	19/20	8/20	20/20	20/20	0/20
Percent Appropriate	80	95	40	100	100	0

Comments

Key to symbol codes in Notes
% Did not code prolongation on 's' because sound was also repeated
* Indeterminate vocal behavior in gloss marked by *

Summary

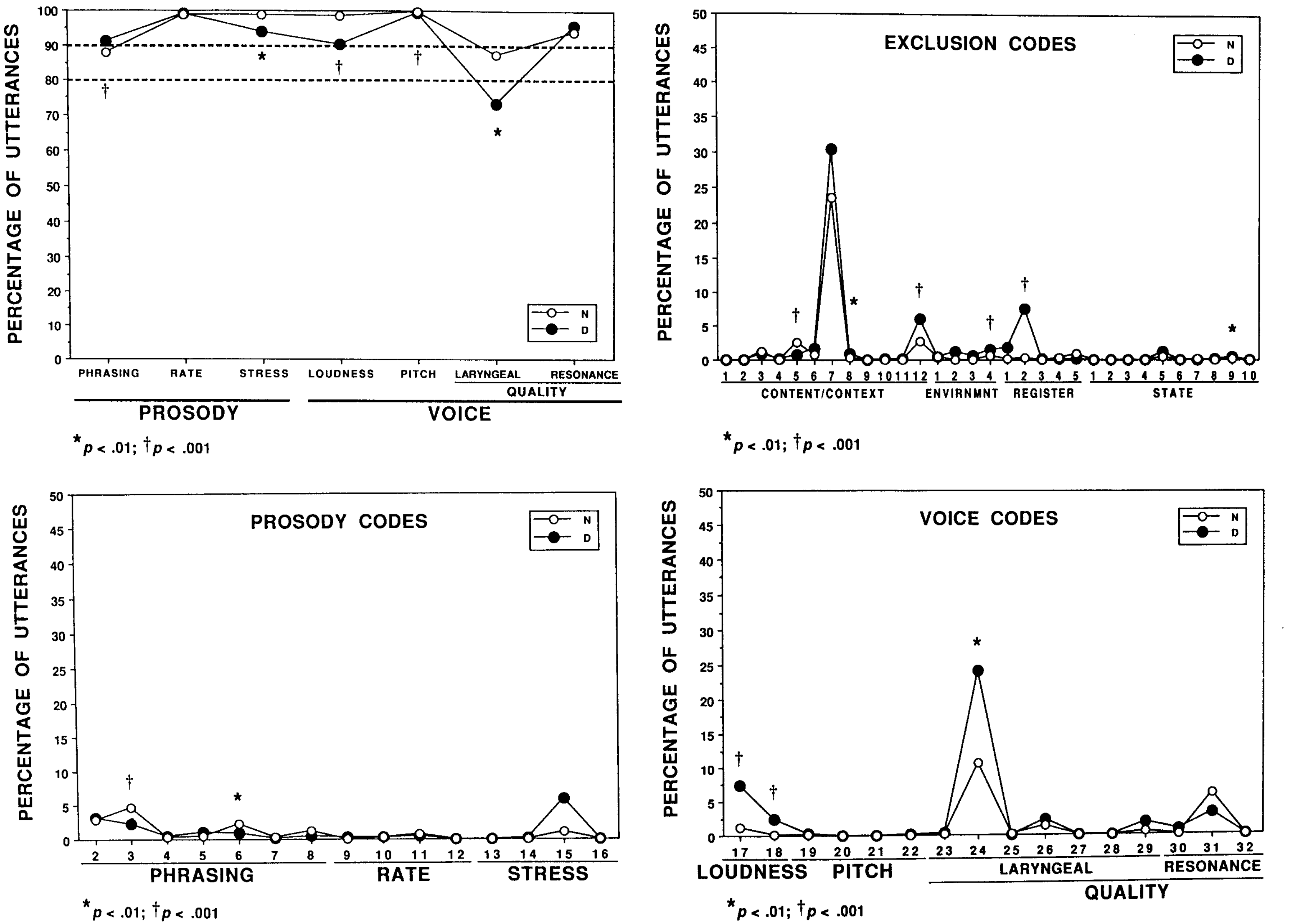
Exclusions
Infrequent exclusion of utterances due to sampling context variables. All utterances were sufficiently intelligible for PVSP coding in spite of the severity of the speech and language problems

Prosody-Voice
Phrasing: Frequent repetitions and revisions may be associated with word retrieval problems.
Rate: Occasional fast rate suggests that rate may be a problem in other speech contexts.
Stress: Frequent excessive/equal and also misplaced stress reflects difficulty maintaining and contrasting primary stress with secondary and tertiary stress.
Loudness and Pitch: Consistently appropriate.
Quality: Voice breaks and pitch shifts (unassociated with normal pubescent voice change) suggest problems coordinating the larynx with other components of speech sound production. Resonance was consistently nasopharyngeal

Recommendations

Inappropriate Phrasing, Rate, and Stress appear to be associated with language variables. In addition, inappropriate Rate and Stress and voice breaks/pitch shifts appear associated with speech motor constraints. Current speech and language management targets include accurate production and sequencing of sounds within multisyllabic words, vocabulary development, and verbal expression of abstract ideas. In addition to continuing work on these targets, diagnostic teaching is recommended to (1) determine how readily stress can be modified, and (2) identify techniques to facilitate stress modification and self monitoring

Figure 2



PERCENTAGE OF UTTERANCES
100
90
80
70
60
50
40
30
20
10
0
N
D
PHRASING
RATE
STRESS
LOUDNESS
PITCH
LARYNGEAL
RESONANCE
QUALITY
PROSODY
VOICE
*p < .01; †p < .001
EXCLUSION CODES
PERCENTAGE OF UTTERANCES
50
45
40
35
30
25
20
15
10
5
0
N
D
1 2 3 4 5 6 7 8 9 10 11 12 1 2 3 4 1 2 3 4 5 1 2 3 4 5 6 7 8 9 10
CONTENT/CONTEXT
ENVIRNMNT
REGISTER
STATE
*p < .01; †p < .001
PROSODY CODES
PERCENTAGE OF UTTERANCES
50
45
40
35
30
25
20
15
10
5
0
N
D
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
PHRASING
RATE
STRESS
*p < .01; †p < .001
VOICE CODES
PERCENTAGE OF UTTERANCES
50
45
40
35
30
25
20
15
10
5
0
N
D
17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
LOUDNESS
PITCH
LARYNGEAL
RESONANCE
QUALITY
*p < .01; †p < .001

Figure 3

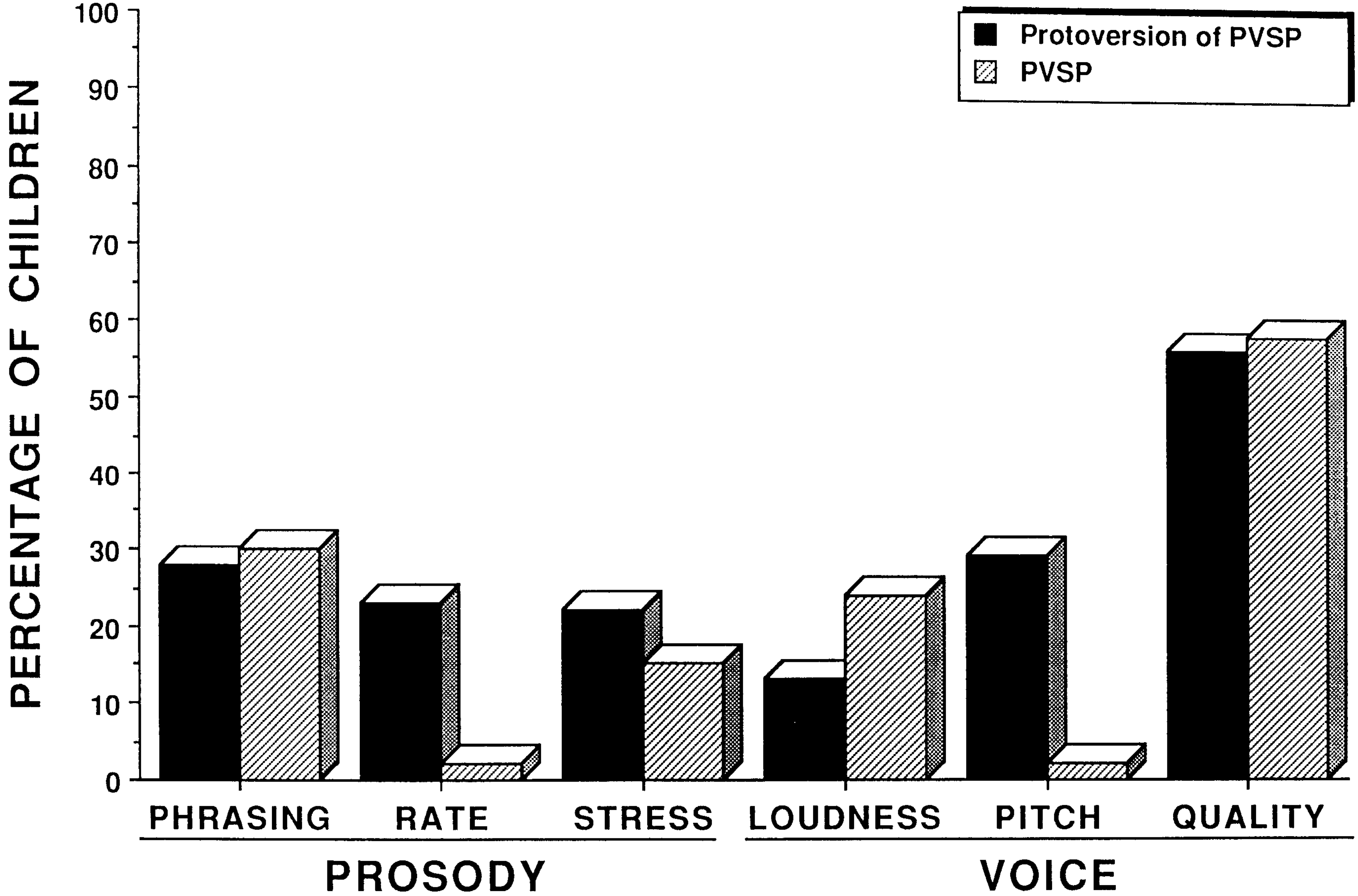

PERCENTAGE OF CHILDREN
100
90
80
70
60
50
40
30
20
10
0
Protoversion of PVSP
PVSP
PHRASING
RATE
STRESS
LOUDNESS
PITCH
QUALITY
PROSODY
VOICE

Figure 4

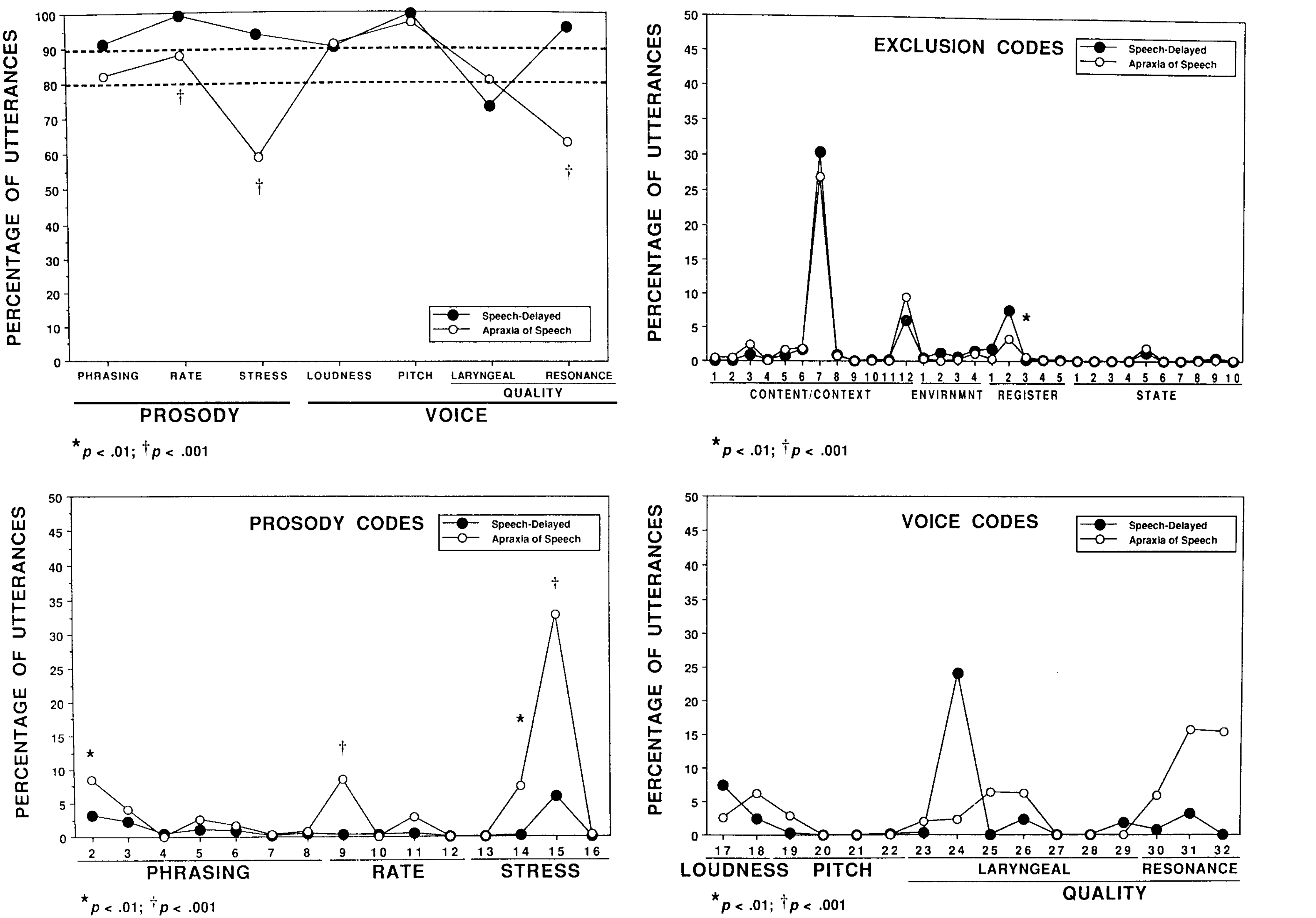

PERCENTAGE OF UTTERANCES
Speech-Delayed
Apraxia of Speech
PHRASING
RATE
STRESS
LOUDNESS
PITCH
LARYNGEAL
RESONANCE
QUALITY
PROSODY
VOICE
*p < .01; †p < .001
EXCLUSION CODES
CONTENT/CONTEXT
ENVIRNMNT
REGISTER
STATE
PROSODY CODES
VOICE CODES
LOUDNESS
PITCH

Figure 5

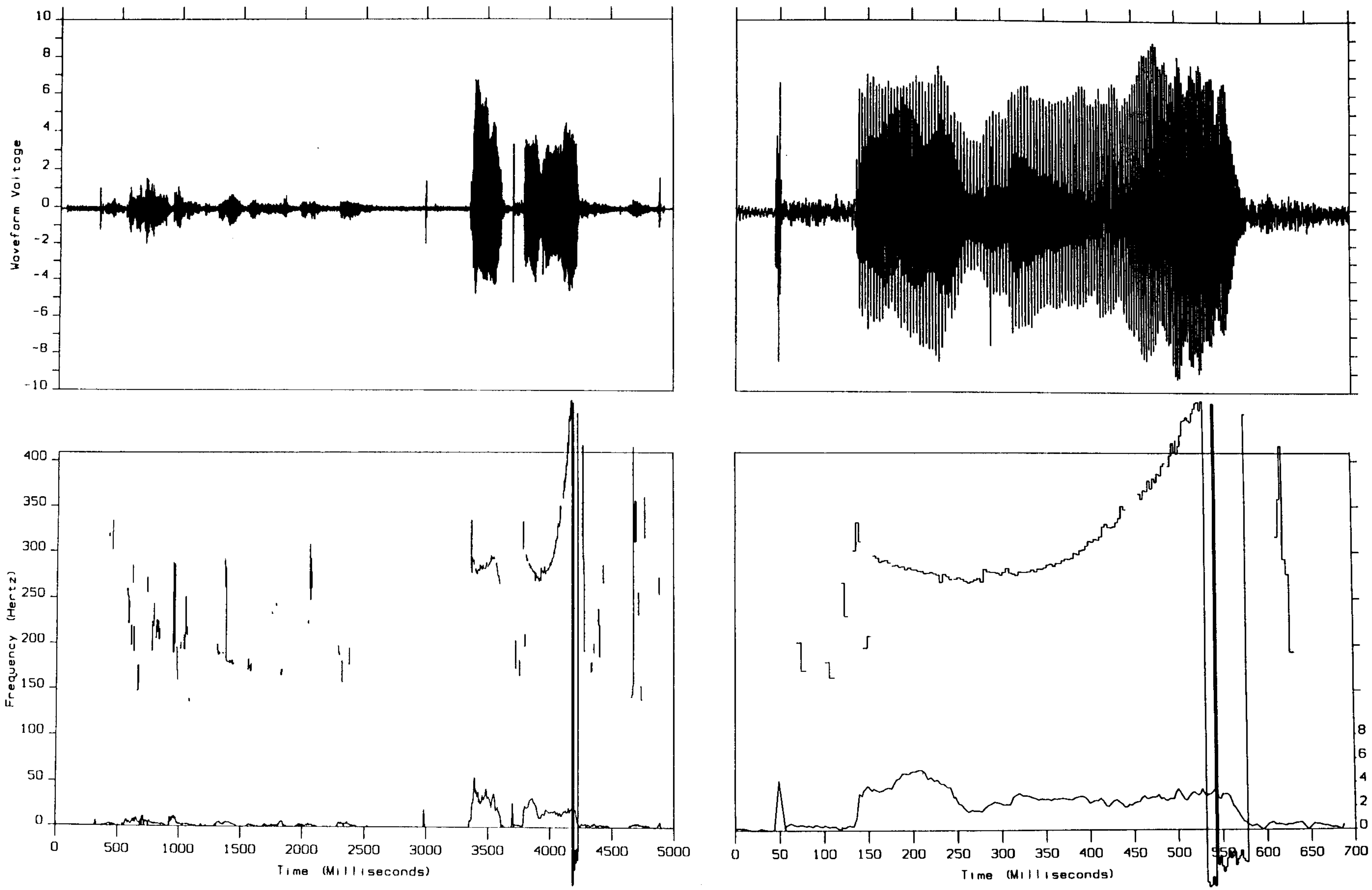

Waveform Voltage
10
8
6
4
2
0
-2
-4
-6
-8
-10
Frequency (Hertz)
400
350
300
250
200
150
100
50
0
0 500 1000 1500 2000 2500 3000 3500 4000 4500 5000
Time (Milliseconds)
0 50 100 150 200 250 300 350 400 450 500 550 600 650 700
Time (Milliseconds)
Peak Volts
8
6
4
2
0

Figure A1

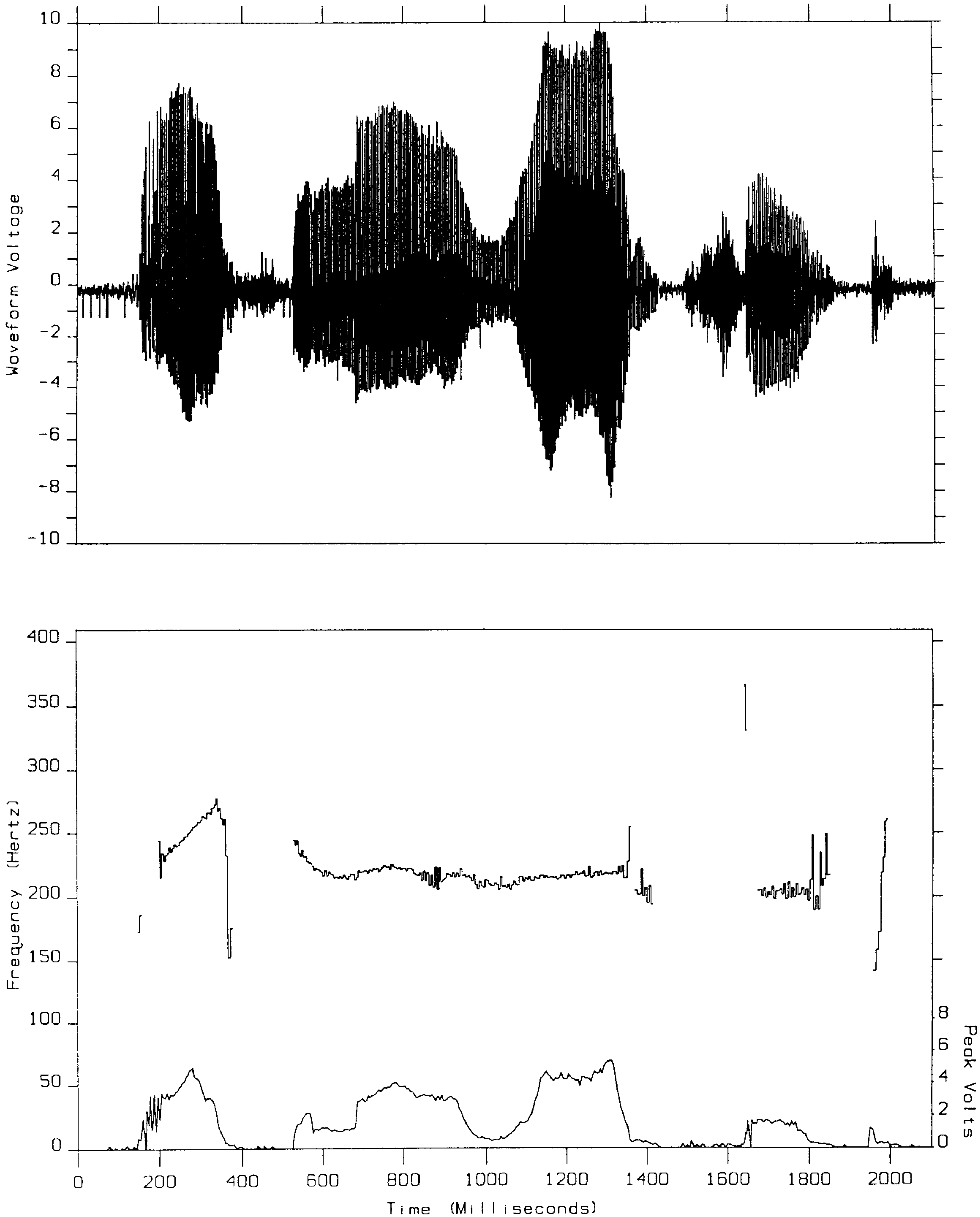

10
8
6
4
2
0
-2
-4
-6
-8
-10
Waveform Voltage
400
350
300
250
200
150
100
50
0
Frequency (Hertz)
8
6
4
2
0
Peak Volts
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Time (Milliseconds)

Figure A2

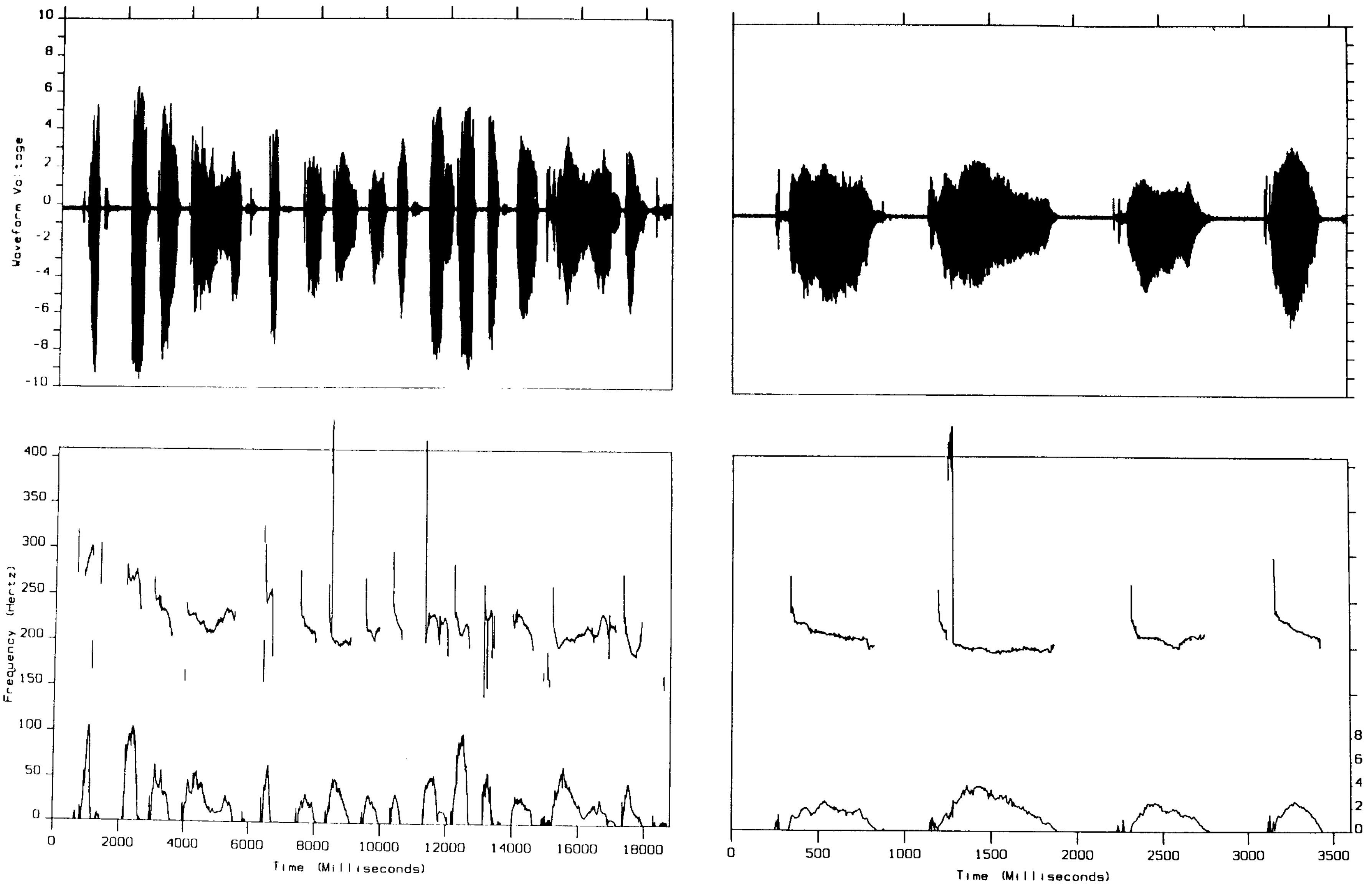

Waveform Voltage
Frequency (Hertz)
Peak Volts
Time (Milliseconds)
Time (Milliseconds)

Figure A3

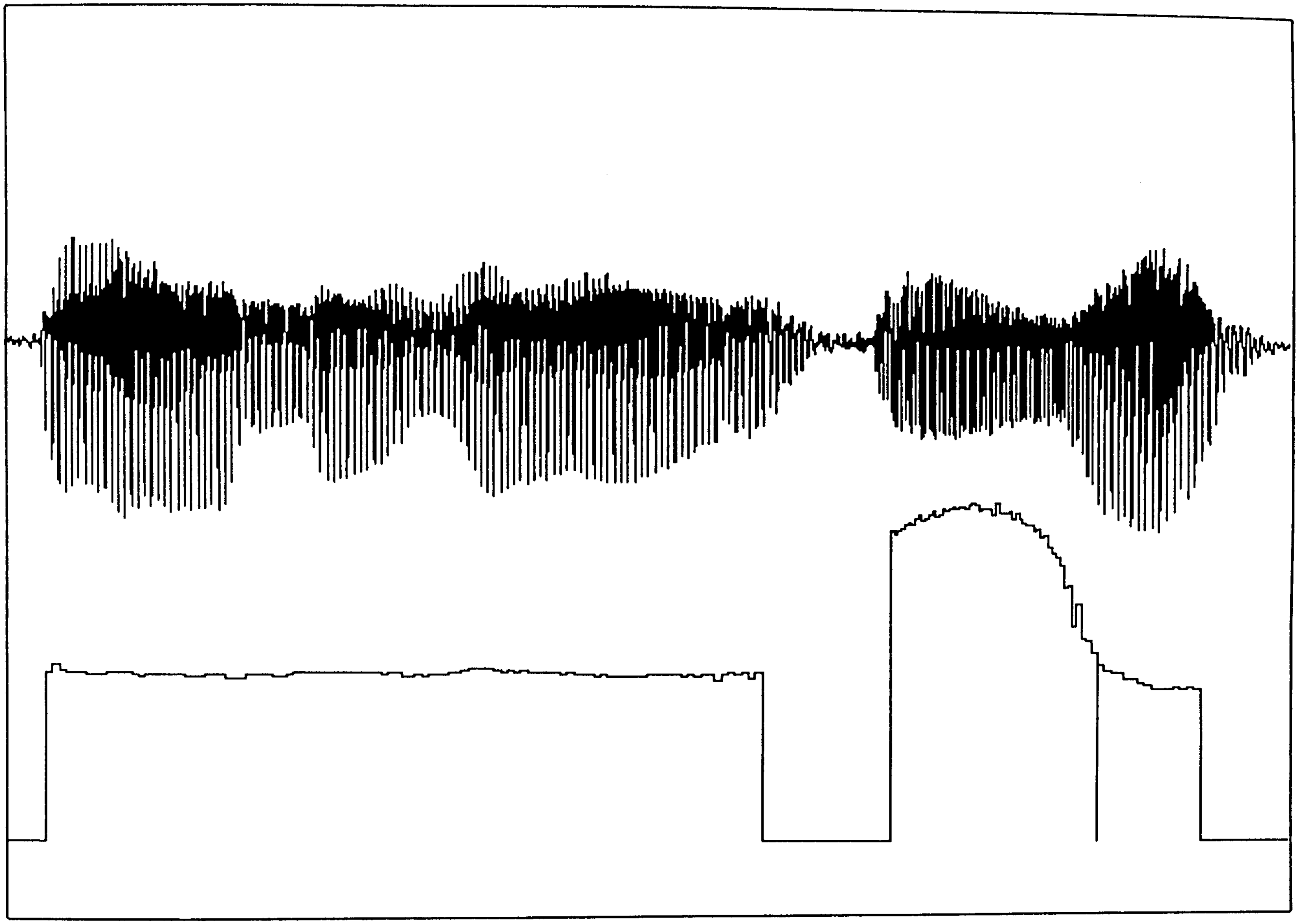


Figure A4

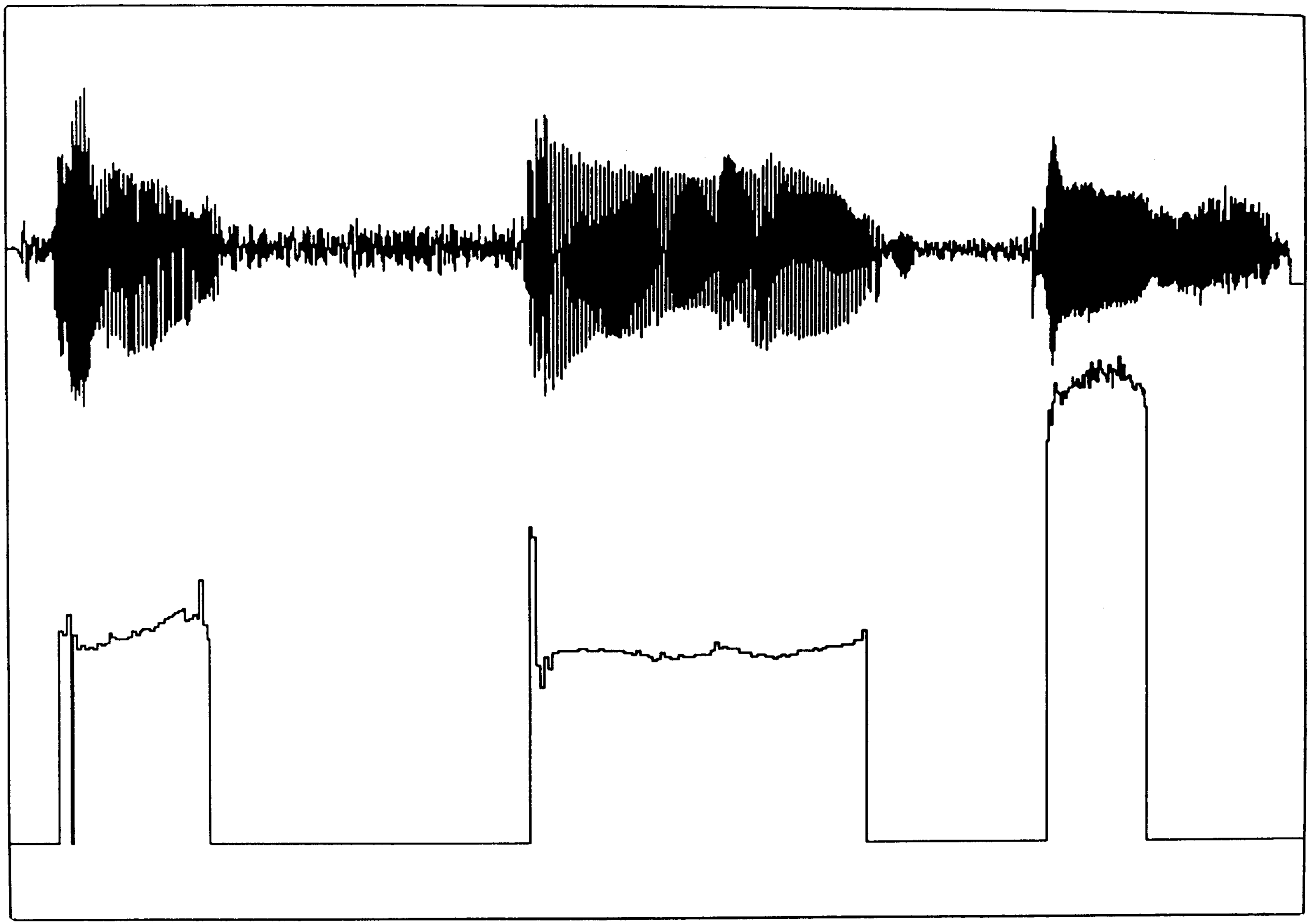


Figure A5