CLINICAL LINGUISTICS & PHONETICS, 2001, VOL. 15, NO. 7, 505–528

Clinical research with the prosody-voice screening profile

JANE L. MCSWEENY and LAWRENCE D. SHRIBERG

Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison, USA

(*Received 20 November 2000; accepted 28 April 2001*)

Abstract

The Prosody-Voice Screening Profile (PVSP) is a clinical research instrument that quantifies a speaker's conversational speech status in seven suprasegmental domains: phrasing, rate, stress, loudness, pitch, laryngeal quality and resonance. The PVSP has been used to assess the prosody-voice characteristics of children and adult speakers with typical speech-language development and with a variety of speech-language disorders of known and unknown origin. PVSP coding requires a trained examiner to determine, on an utterance-by-utterance basis within each of the seven domains, whether a speaker's prosody-voice characteristics can be defined as appropriate based on a set of auditory-perceptual criteria. This report provides brief overviews of the development, administration, and psychometric features of this screening tool, and summarizes prosody-voice findings to date for several typically speaking and clinical populations.

Keywords: Assessment, measurement, prosody, speech disorders, suprasegmental.

Introduction

The goal of this report is to provide overviews of the development, administration, and psychometric characteristics of the Prosody-Voice Screening Profile (PVSP), and brief summaries of clinical findings to date. Space limitations preclude a literature review or a comparative analysis of the PVSP with assessment approaches discussed elsewhere in this volume. Rationale leading to a conceptual framework for the PVSP and detailed psychometric data are available in two sources: Shriberg, Kwiatkowski and Rasmussen (1990) and Shriberg, Kwiatkowski, Rasmussen, Lof and Miller (1992). Essentially, the goal was to construct an assessment tool whose coding categories were driven by empirical samples and were orthogonal, rather than to adapt an existing prosody framework for the range of inappropriate prosody observed in individuals with communicative disorders. For example, a traditional

Address correspondence to: Lawrence D. Shriberg, Ph.D., The Phonology Project, Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, WI 53705, USA.

Clinical Linguistics & Phonetics
ISSN 0269-9206 print/ISSN 1464-5076 online © 2001 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/02699200110078159

prosody term such as intonation was found to embrace a complex of interdependent prosodic and vocal behaviours, rather than a single prosodic or vocal behaviour that could be operationally defined and reliably coded for quantitative analysis. For this reason, many of the classificatory terms used in the PVSP, as well as the codes they subsume, may appear somewhat unconventional relative to prosody traditions in descriptive linguistics.

For the interested reader, copies of a technical report that includes PVSP reference data can be downloaded at http://www.waisman.wisc.edu/phonology/. Copies of the PVSP training manual and training tapes are also freely available to clinical investigators on request at this site.

Sample PVSP Output

It is useful to begin with an illustration of PVSP data. Figure 1 is a sample of the four-panel Prosody-Voice Screening Profile produced by the PVSP software (Shriberg, Allen, McSweeny and Wilson, 2000). The data shown in figure 1 were obtained from conversational speech samples from four groups of children whose speech histories were classified as typical speech-sound acquisition (group A, circles), speech delay of unknown origin (group B, squares), speech delay suspected to be associated with developmental psychosocial involvement (group C, triangles), and speech delay associated with suspected apraxia of speech (group D, diamonds). As shown in the upper left panel (panel A) in figure 1, the PVSP procedure provides

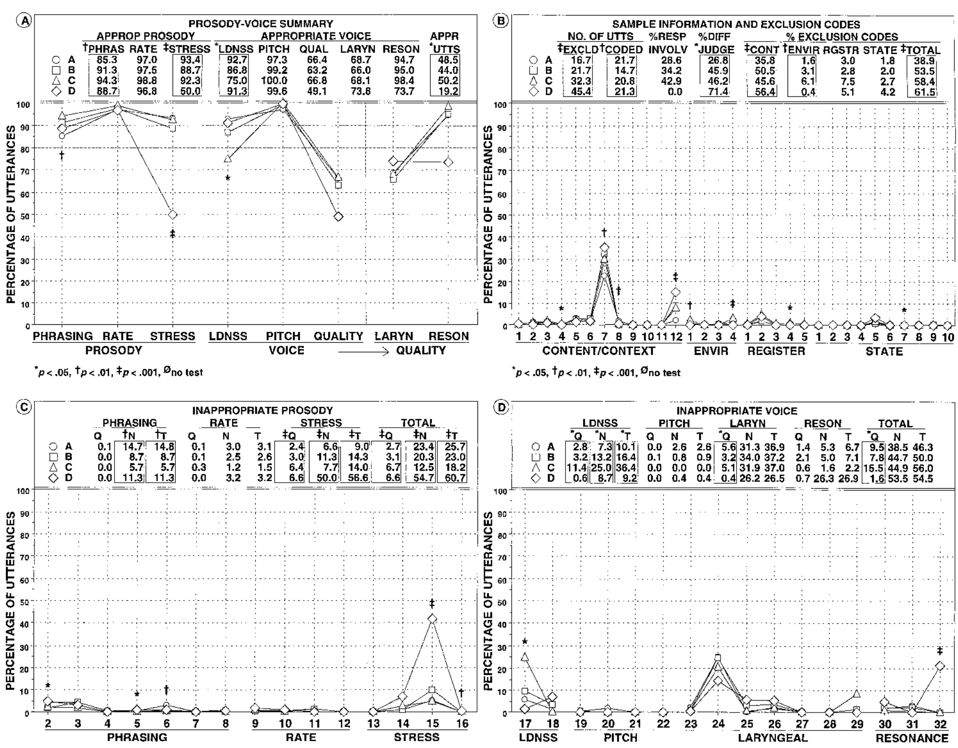

Figure 1. *Sample Prosody-Voice Screening Profile from children in four clinical populations.*

quantitative information on the percentage of utterances judged to meet specified perceptual criteria for the construct of appropriate prosody in three domains (phrasing, rate, stress) and appropriate voice in four domains (loudness, pitch, laryngeal quality, resonance quality). For some applications, data on a combined category termed quality (laryngeal and resonance) have been useful, but these two domains are generally treated separately. Figure 1 also provides descriptive information on 31 exclusion codes that profile a speaker's paralinguistic behaviours and other information relevant to prosody-voice coding (panel B), and on 31 inappropriate prosody-voice codes within the seven prosody-voice domains (panels C and D). The key for the 31 exclusion codes and the 31 prosody-voice codes is shown in figure 2. Additional output termed the 'detailed report' (not shown) provides, for each of the variables in figure 1, descriptive and inferential statistical information that aid in interpretation of the data for clinical and research questions.

As evident in the sample numeric and graphic data in figure 1, the prosody-voice characteristics of children in these four groups differ significantly on at least one prosody-voice variable in each of the panels. In this illustration the inferential statistic selected from the program options to test for between-group differences was the nonparametric Kruskal-Wallis statistic (Siegel and Castellan, 1988). Beginning with panel A, notice that there are several statistically significant between-group findings (unadjusted for multiple comparisons). Two notable group differences, for example, are the lowered average appropriate stress percentages for children in group D (suspected apraxia of speech) and the lowered average appropriate loudness percentages for children in group C (suspected psychosocial involvement), compared to the average percentages of appropriate utterances in these two domains for children in each of the other groups. The code-level sources of these two differences are indicated in panel C and panel D, respectively. The lowered average stress percentages for children in group D appear to be primarily due to one particular inappropriate stress code, PV15: Excessive/Equal/Misplaced Stress (see figure 2). The lowered average loudness percentages for children in group C appear to be due exclusively to one of the two inappropriate loudness codes: PV17: Too Soft. We will later consider clinical-research implications of findings such as these.

The sample output in figure 1 provides data on many other prosody-voice variables that have been used to describe individual and grouped speakers. For example, the last column in the numeric (top) section of panel A indicates the percentage of utterances in the sample in which all seven prosody-voice variables were judged appropriate. Notice that, among the four groups of speakers in this comparison, children in group D (suspected apraxia of speech) had significantly fewer utterances coded as completely appropriate (19.2%) compared to the percentage of utterances coded as completely appropriate for the typically speaking children in group A (48.5%), the children with speech delay of unknown origin in group B (44.0%), or the children with speech delay and suspected developmental psychosocial involvement in group C (50.2%). Notice too that even children with typical speech development (group A) may have conversational speech samples that meet criteria for inappropriate phrasing and/or inappropriate laryngeal quality (to be defined later).

Development of the PVSP

In the 1980s a number of clinical-research procedures were developed to quantify phonetic and phonological aspects of a speaker's segmental development and performance (cf. Bernthal and Bankson, 1998). The development of the PVSP during this active period in child phonology arose from the need for parallel information on speakers' prosody and voice characteristics. One early finding supporting this need was that for children with speech-sound disorders of unknown origin, suprasegmentals were rated the fifth most important factor (after intelligibility, age, articulation, and language) contributing to perceived severity of involvement (Shriberg and Kwiatkowski, 1982a). Associated analyses indicated that although Percentage of Consonants Correct (PCC) values were significantly associated with a speaker's severity ratings by trained speech-language pathologists, inappropriate suprasegmentals played a major role in accounting for certain speakers rated more severely impaired than predicted by their PCC (i.e., segmental) scores.

Our earliest attempts to quantify the suprasegmental status of children with speech-sound disorders of unknown origin yielded ordinal estimates based on the percentages of utterances judged inappropriate in prosody or voice features (e.g., Shriberg and Kwiatkowski, 1982a, 1982b; Shriberg, Kwiatkowski, Best, Hengst and Terselic-Weber, 1986). Rather than describe or rate the severity of a speaker's prosody using any of the descriptive-linguistic frameworks available at that time, the goal in these early studies was to code the utterance-by-utterance frequency of occurrence of a speaker's inappropriate prosody-voice behaviours relative to normative age and gender expectations. For each of the seven prosody-voice domains shown in figure 1, a conversational speech sample was classified into the following ordinal polychotomy: 0 if prosody-voice was considered typical (no inappropriate utterances), 1 if questionable (10–15% utterances) and 2 if involved (more than 15% inappropriate utterances). Scoring guidelines for the construct of inappropriate were fairly general, relying primarily on an examiner's clinical experience with children of the ages studied. Not surprisingly, coding agreement for this system was not high: mean interjudge and intrajudge agreement percentages across the seven variables were 67% and 77%, respectively (Shriberg and Kwiatkowski, 1982b).

A second version of the PVSP was developed to meet the needs of a study to describe the prosody-voice characteristics of adult speakers with mental retardation (Shriberg and Widder, 1990). Because inappropriate suprasegmentals and paralinguistic behaviours were frequently noted as barriers to social-vocational adjustment for speakers with mental disability, we were particularly interested in quantifying the utterance-by-utterance frequency of occurrence of such behaviours in these individuals. Descriptors were developed to cover the range of inappropriate prosody-voice and paralinguistic behaviours heard in samples of typical and atypical speakers, procedures for speech sampling were standardized, and guidelines were developed for each coding decision.

The third version of the PVSP was developed over an approximately 3-year period from 1987 to 1989, with the project proceeding in five phases. First, several hundred audio-taped speech samples of children and adult speakers representing ten clinical populations were collected and reviewed. Next, the prosody and voice involvements observed in this database were used to elaborate the form and content of the prior prosody-voice screening procedure. Psychometric studies (some of which are summarized in Appendix B) were then completed. During the fourth phase training materials were constructed, including a detailed manual, audiocassette training tapes, and a set of scoring forms. Finally, the procedure was assessed and field-tested in a number of validity, reliability and efficiency studies.

The fourth and current version of the PVSP, which includes software for data entry and analysis, evolved from computer programming completed during the last two decades, with enhancements continuing to the present. Mainframe, minicomputer, and currently microcomputer platforms for the PVSP have been integrated within a continually evolving suite of programs for research in child speech disorders (Shriberg, 1986; Shriberg, Allen *et al.*, 2000; Shriberg and Nockerts, 1982). The output shown in figure 1 illustrates the type of computer-assisted prosody-voice data currently available to the user, including options for a variety of single-subject and group comparisons, parametric and nonparametric inferential statistical tests and publication-ready graphic formats. The detailed report noted previously provides additional statistical information, such as distributional data for each variable.

For the interested reader, Appendix A and Appendix B provide, respectively, information on the administration and scoring of the PVSP and brief summaries of findings from psychometric studies. The information in Appendix B illustrates the challenge of developing valid, reliable, and time-efficient instruments to assess prosody-voice in a clinical-research setting. As with all other procedures that rely on auditory-perceptual methods and skills (see the useful review of issues and findings in Kent, 1996), the reliability of some levels of PVSP data may not be adequate when effect sizes are small. The next step in the evolution of this assessment task is to develop procedures for acoustics-assisted prosody-voice coding.

Some clinical research with the PVSP

Tables 1 and 2 include summaries of findings from studies to date that have used the PVSP to describe and compare conversational speech samples from a number of speaker populations. Inclusion in one or both tables is based on how the data were reported in each of the citations. Both table 1 and table 2 include PVSP findings from speakers with typical speech development and Speech Delay (SD) of unknown origin. Table 1 also includes data sampled from speakers in several other clinical populations. The data entries for each of the seven prosody-voice domains in table 1 are the percentages (mean, standard deviation) of appropriate utterances for speakers in each of the 13 study samples. The data entries in table 2 are the percentage of speakers in each of the five study samples whose scores on each PVSP domain were classified as pass (90–100%), questionable pass (80–89.9%), or fail (<80%). In both tables the study samples within each population are arranged by increasing age.

The following discussions highlight some comparative aspects of these data for each of the seven prosody-voice domains. All of the brief definitions and overviews of PVSP coding criteria are excerpted from the training manual (Shriberg *et al.*, 1990), which henceforth is not cited. It is important to underscore that the training manual includes several pages of instruction for each inappropriate prosody-voice code, including detailed guidelines, contrastive audio-taped exemplars, and graded (i.e., increasingly challenging) practice items.

Phrasing

Definition and coding criteria

The PVSP defines an utterance with appropriate phrasing as ‘a flow of word and phrase groups that is appropriate for the speaker’s age, emotional state, and the intended propositional content.’ Phrasing behaviours considered to be appropriate include repetitions of entire phrases (more than one word), revisions of entire words,

Exclusion Codes

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

Prosody-Voice Codes

Prosody

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

Voice

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

Figure 2. *Key for the codes used in the Prosody-Voice Screening Profile (PVSP).*

occurrences of filler words ('um', 'uh', etc.), atypical breath groups, and nonphonemic sound productions. The six inappropriate phrasing codes include four codes that are used when the following repetitious behaviours occur (see figure 2): sound/syllable repetitions (PV2), whole-word repetitions (PV3), one sound/syllable and one whole-word repetition in a single utterance (PV4), and multiple-word repetitions within an utterance (PV5). The remaining two inappropriate phrasing codes cover single and multiple occurrences (PV6 and PV7, respectively) of part-word revisions within an utterance.

Typical speakers

As found in reference studies of constructs similar to phrasing, such as for the construct of mazing (Miller and Chapman, 1992), typical speakers have a wide range of repetitions and revisions in their conversational speech. Indeed, increases in mean length of utterance and the complexity of language constructions can be accompanied by higher frequencies of repetitions and revisions. As a screening tool, the PVSP simply tabulates such behaviours, with greater than 20% occurrence on an utterance-by-utterance summative basis meeting criteria for fail. As with all screening tools, a fail indicates the need for additional examination of the data and/or additional testing, as might a questionable pass in some situations.

There is good agreement among the mean phrasing scores of the three groups of typical speakers in table 1, with child and adult speakers averaging in the questionable range (mid 80%) and having similar standard deviations of approximately 10%. The two samples providing per-speaker phrasing data in table 2 are also in good agreement, with approximately 18% of speakers meeting criteria for fail (<80%) in both studies. Thus, children and adults with typical speech have repetitions and revisions in approximately 15%, or 1–2 of every 10, utterances, with nearly 20% of typical speakers having them on more than 1 of every 5 utterances. Once again, the task of the PVSP is to tabulate the utterance-by-utterance frequency of occurrence of behaviours such as repetitions and revisions, rather than to evaluate the clinical or social significance of such behaviours, which might vary greatly depending on the propositional importance and complexity of discourse.

Speech delay of unknown origin

The means and standard deviations for phrasing scores for children with SD in table 1 are consistent among one another and with the data for typical speakers summarized above. As aggregated by child in table 2, however, the three estimates of the percentages of children meeting PVSP criteria for pass, questionable pass, and fail are notably different, possibly related to differences in sample size and composition, changing scoring criteria for the version of the PVSP available in 1986 versus the finalized version, and/or the reliability of the coders. From the last two perspectives, the phrasing data obtained from the largest sample of children using the most recent version of the PVSP (table 2; Shriberg, 1993; Shriberg, Austin, Lewis, McSweeny and Wilson, 1997; Shriberg and Kwiatkowski, 1994) are considered the best estimates of the percentages of children with inappropriate phrasing on more than 20% of their utterances. These data, reflecting the means and variance data reported for this same group of 64 children reported in table 1, indicate that approximately 90% of children with speech delay of unknown origin have appropriate or questionably appropriate phrasing. An interesting question for further examination in these data is the percentage of variance in these speech delayed children's phrasing scores that is accounted for by their speech involvement versus the percentage of variance associated with their language involvement.

Clinical populations

PVSP findings have been reported for conversational samples obtained from several clinical populations, including a sample of 40 adults with mental retardation (MR),

Table 1. *PVSP research findings reported as the percentage of appropriate utterances (Mean, Standard Deviation) for each study sample*

				Prosody						Voice							
	Age (yrs;mos)			Phrasing		Rate		Stress		Loudness		Pitch		Laryngeal Quality		Resonance Quality	
Group	*M*[b]	Range	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Typical Speech Acquisition																	
Shriberg and Kwiatkowski, 1994	–	3–5	71	87.7	9.4	98.7	3.4	98.8	2.5	98.6	3.9	99.8	1.1	87.2	25.7	94.0	15.4
Shriberg *et al*., 1992	–	3–19	115	88.2	10.2	97.3	9.1	98.1	3.8	96.3	15.1	99.5	2.9	89.2	21.7	94.3	15.7
Shriberg, Paul *et al*., 2000	26;5	10–30	53	86.3	10.1	95.3	7.6	95.2	7.1	97.0	8.6	94.6	13.6	76.1	31.8	98.1	4.8
Speech Delay of Unknown Origin																	
Shriberg, 1993; Shriberg and Kwiatkowski, 1994; Shriberg *et al*., 1997b	4;3	3;0–6;1	62	91.1	9.8	99.2	2.2	94.2	11.6	91.2	11.8	99.5	2.6	71.1	36.4	95.9	10.0
Shriberg *et al*., 1992	–	3–19	137	88.9	11.5	96.5	10.0	87.3	21.5	89.7	19.7	99.2	3.9	75.3	31.4	83.4	32.0
Clinical Populations																	
Apraxia																	
Shriberg *et al*., 1997b	5;3	4;7–6;6	11	93.9	7.3	95.6	5.7	61.6	22.7	93.2	8.5	100.0	0.0	76.1	31.4	94.3	7.7
Shriberg *et al*., 1997a	5;9	4;10–6;3	6	81.3	17.6	85.0	33.9	64.9	43.4	85.1	23.7	95.2	12.6	73.0	26.6	65.4	45.5
Shriberg *et al*., 1997a	10;2	7;1–14;11	7	84.3	9.7	93.2	10.1	60.4	25.1	98.8	3.1	99.3	1.9	94.4	8.1	65.8	46.2
Shriberg *et al*., 1997b	10;9	7;11–14;4	10	89.6	9.2	97.9	4.5	86.1	11.4	96.2	8.0	100.0	0.0	69.4	37.8	66.5	43.5
Odell and Shriberg, 2000	–	50–81	14	55.3	25.8	42.0	26.6	82.8	16.1	97.3	10.0	99.4	1.5	82.8	22.5	97.6	6.7

Table 1. *(Continued)*

				Prosody						Voice							
	Age (yrs;mos)			Phrasing		Rate		Stress		Loudness		Pitch		Laryngeal Quality		Resonance Quality	
Group	*M*[b]	Range	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Mental Retardation																	
Shriberg and Widder, 1990[a]	–	20–55	40	78.8	–	68.3	–	74.3	–	90.0	–	87.4	–	–	–	–	–
Pervasive Developmental Disorders																	
HFA-Shriberg, Paul *et al.*, 2000	21;7	10–49	15	76.6	16.7	91.9	8.4	77.3	19.2	91.4	16.0	90.2	26.9	84.0	21.9	76.9	32.8
AS-Shriberg, Paul *et al.*, 2000	20;8	10–49	15	75.6	14.1	95.6	6.6	86.5	8.7	96.4	5.6	98.9	2.5	77.2	28.4	80.6	34.4

[a]Used a protoversion of the PVSP; see text.
[b]Mean ages are based on the total number of subjects reported, which in some cases are slightly more than the number of subjects for whom prosody-voice data were obtained.

Table 2. *PVSP research findings reported as the percentage of speakers in each study sample meeting criteria for* pass *(P),* questionable pass *(Q), and* fail *(F)*[a]

				Prosody									Voice											
	Age (yrs;mos)			Phrasing			Rate			Stress			Loudness			Pitch			Laryngeal Quality			Resonance Quality		
Group	*M*	Range	*n*	P	Q	F	P	Q	F	P	Q	F	P	Q	F	P	Q	F	P	Q	F	P	Q	F
Typical Speech Acquisition																								
Shriberg and Kwiatkowski, 1994	–	3–5	71	52.1	29.6	18.3	97.2	1.4	1.4	98.6	1.4	0.0	97.2	0.0	2.8	100.0	0.0	0.0	76.1	7.0	16.9	85.9	4.2	9.9
Lewis and Shriberg, 1994[b]	–	9;0–49+	111	49.5	31.8	18.7	82.2	11.2	6.5	90.6	7.5	1.9	93.5	2.8	3.7	87.7	4.7	7.6	58.1	9.5	32.4	91.6	2.8	5.6
Speech Delay of Unknown Origin																								
Shriberg *et al.*, 1986	4;1	1;11–6;8	14	50.0	14.3	35.7	57.1	21.4	21.4	78.6	7.1	14.3	85.7	0.0	14.3	50.0	21.4	28.6	–	–	–	–	–	–
Shriberg, 1993; Shriberg and Kwiatkowski, 1994; Shriberg *et al.*, 1997	4;3	3;0–6;1	62	71.0	17.7	11.3	100.0	0.0	0.0	82.3	6.5	11.3	69.4	19.4	11.3	98.4	1.6	0.0	51.6	12.9	35.5	88.7	4.8	6.5
Shriberg *et al.*, 1986	4;11	2;10–9;7	38	97.4	2.6	0.0	86.8	2.6	10.5	76.3	7.9	15.8	82.4	5.3	10.5	79.0	2.6	18.4	–	–	–	–	–	–

[a]For findings reported in 1986, *pass* is equivalent to a rating of 0, *questionable pass* to a rating of 1, and *fail* to a rating of 2; see text for description of rating system. For all other studies, *pass* indicates 90% or more appropriate utterances, *questionable pass* indicates 80.0–89.9% appropriate utterances, and *fail* indicates less than 80% appropriate utterances.
[b]This group consists of speakers with normal speech acquisition or residual errors.

14 adults with apraxia of speech (AOS), several samples of children with suspected apraxia of speech (AOS), and adolescents and adults with two forms of pervasive developmental disabilities—15 individuals with high-functioning autism (HFA) and 15 individuals with Asperger syndrome (AS). The numbers of speakers in each of these groups have been relatively small, reflecting the difficulty in assembling sociodemographically well-balanced groups of individuals from special populations.

As shown in table 1, the mean phrasing scores for the samples of children with suspected AOS are within the range of the scores reported above for typical speakers and speakers with speech delay of unknown origin. Mean phrasing scores for speakers in the MR, HFA and AS groups are approximately 0.5–1 standard deviation below those of typical speakers in the same age ranges, and mean scores for the adult speakers with AOS are several standard deviations below the mean for adults with typical speech. Thus, as considered in each of the references for these studies, the high frequency of occurrence of repetitions and revisions for individuals in these clinical populations is associated with semantic, syntactic and pragmatic challenges. Interestingly, in the study of adults with mental retardation, appropriate phrasing was not found to be significantly associated with gender or level of intellectual functioning, but was estimated to be one of the prosodic variables associated with a higher probability of an individual transitioning to successful independent living.

Rate

Definition and coding criteria

The PVSP codes a speaker's rate in an utterance as appropriate if it is 'within the normal limits for a speaker's age, dialect, and emotional state'. Based on findings reported in normative studies of speech rate, appropriate rate is defined as 2–4 syllables per second for young children and 4–6 syllables per second for older children and adults. For one study that included senior speakers (Odell and Shriberg, 2001), alternative rate criteria were used based on reference data from a study of elderly speakers. The four inappropriate rate codes include two codes that correspond to inappropriately slow rates of speech, due to either slow articulation and pause time (PV9) or pause time alone (PV10), and two codes that correspond to excessively fast rates of speech (PV11: Fast and PV12: Fast/Acceleration). PV12 is coded when fast rate is accompanied by the percept of a breakdown in articulatory control.

Typical speakers

The means for syllable rates reported in table 1 for the three samples of children and adults with typical speech acquisition are well within the appropriate range. As indicated by the standard deviations, however, typical speakers may also have speech rates in the questionable range. As indicated more directly in table 2, particularly for the relatively large sample of 111 speakers from 9 to over 50 years of age (Lewis and Shriberg, 1994), nearly 20% of typical speakers have rates that are either inappropriately fast or inappropriately slow on over 10% of their utterances.

Speech delay of unknown origin

A longstanding question in the literature on children with speech delay of unknown origin is whether the source of the delay is associated with clinical or subclinical impairment in speech motor control processes. As a widely studied correlate of deficits in speech motor control, speech rate is one prosodic variable that might provide a relatively objective metric for studies in this clinical population (cf. Flipsen, 1999). The PVSP data to date, however, do not support this hypothesis. Both of the relatively large samples of children with speech delay in table 1 have means and standard deviations well within the pass or questionable pass range on this suprasegmental. The three samples providing per-speaker data in table 2 suggest a similar conclusion, although the first study found that only approximately 79% of these relatively few (14) children had scores meeting criteria for pass or questionable pass.

Clinical populations

Among the studies of speakers with AOS, MR, HFA and AS, the rate findings in table 1 indicate possible rate involvements for individuals in only two of the clinical populations: adults with AOS and adults with MR. Adults with AOS averaged 42% utterances with appropriate rate, and adults with MR averaged approximately 68% utterances with appropriate rate. The PVSP subcode data for each study group were quite different, however; all speakers with AOS had slowed rates, whereas rates for many speakers with MR were too fast.

In addition to stretches of fast speech in the 40 adults with MR (including eight persons with Down syndrome), it is useful to note that information from the PVSP exclusion codes (see figure 2) was particularly revealing for the individuals in this study. Unlike PVSP findings for speakers in each of the other groups in tables 1 and 2, the conversational speech findings for many of the speakers with MR included frequent and distracting non-speech noises, including yawns, sighs, throat clears, grunts, burps, lip smacks, snorts, sniffs, and tongue clicks. Such behaviours occurred in at least one utterance for 68% of the adult speakers with MR. As with the effects of high rates of inappropriate phrasing and fast speaking rates for these adult speakers, the frequency of occurrence of these paralinguistic behaviours was negatively associated with the probability of independent living.

Stress

Definition and coding criteria

The PVSP defines appropriate stress in conversational speech as variations in the pitch, loudness, and duration of sounds that are consistent with lexical, morphologic, syntactic, and pragmatic rules. Inappropriate stress does not include the many variations in emphasis and style that define a speaker's unique prosody-voice pattern. Inappropriate lexical stress is coded when it meets criteria for PV13: Multisyllabic Word Stress. Inappropriate phrasal stress may alternatively meet coding criteria for PV14: Reduced/Equal Stress (an overall reduction in the forcefulness of typically stressed syllables and words and/or a weak monostress pattern) or PV15: Excessive/Equal/Misplaced Stress (excessively forceful, punctuated monostress, misplaced word stress and/or sound blocks or prolongations). PV16: Multiple Stress Features is used to code utterances that meet criteria for more than one of the other three inappropriate stress codes.

Typical speakers

As indicated in tables 1 and 2, few children and adults with typical speech acquisition have inappropriate stress as coded by PVSP criteria. Well over 90% of the utterances of typical speakers have appropriate lexical and phrasal stress (table 1), with fewer than 10% of the 182 typical speakers assessed in two studies having stress scores in the questionable pass or fail range.

Speech delay of unknown origin

There is good intersample agreement on the stress status of children with speech delay of unknown origin, as indicated by the entries for the four studies in table 1. Mean appropriate stress scores averaged in the low pass to high questionable pass range, with the relatively large standard deviation in the Shriberg *et al.* (1992) sample indicating that some children's stress scores were below the 80% criterion for questionable pass. The estimates from the three samples in table 2 confirm that approximately 11% to 16% of children with speech delay of unknown origin have stress scores below 80%, indicating inappropriate stress based on one or more of the codes in this prosody domain. As noted earlier and described in more detail below, it is at the detailed level of the codes for lexical and phrasal stress—as well as a further level of subtypes within codes—where interesting clinical-research questions arise in these data. For example, studies in progress are examining the frequency of occurrence of PV14: Reduced/Equal Stress in children whose speech delay is suspected to be associated with certain types of mild neurogenic speech impairments.

Clinical populations

Inappropriate lexical and phrasal stress appears to be a frequent finding in conversational speech studies of individuals in several clinical populations, including persons with speech motor, intellectual, and affective challenges. As shown in table 1, three of four subsets of children with suspected AOS had stress scores averaging in the low 60% range, with large standard deviations indicating that some individuals' scores may be considerably lower. These studies indicated that over 50% of such children have stress scores in this range. As discussed in the references cited in table 1 and consistent with a metrical analysis of one sample of children with suspected AOS (Velleman and Shriberg, 1999), inappropriate stress might serve as a clinical marker to identify at least one subtype of childhood AOS. Crucially, however, the type of stress observed in these studies of children differs from the stress deficit described in a PVSP study of adults with AOS (Odell and Shriberg, in press). In children with suspected AOS, the stress deficit is an excessive/equal type of syllable and word stress coded within PV15: Excessive/Equal/Misplaced Stress. In adults with acquired AOS, the primary type of stress deficit is blocks and prolongations of sounds, associated with reductions in speech rate (see above). Collaborative molecular genetic studies in progress will test whether inappropriate excessive/equal stress in children and their siblings is the phenotype marker for a genetically transmitted form of AOS.

As indicated in the three studies in table 1, stress findings for individuals with mental retardation and high-functioning autism averaged below the 80% criteria for pass and questionable pass, with average stress scores for a sample of individuals with Asperger syndrome within the questionable pass range. The most frequent inappropriate stress codes obtained in the studies of persons with MR and HFA differed from the codes obtained in the studies of children with AOS (cf. Shriberg, Paul, McSweeny, Klin, Volkmar and Cohen, in press; Shriberg and Widder, 1990), which were interpreted as support for deficits in speech processing. As discussed in the studies cited, inappropriate stress in the adolescent and adult speakers in these samples appeared to be associated with constraints in pragmatic and affective aspects of discourse.

Loudness, pitch, laryngeal quality and resonance

It is efficient to discuss collectively definitions and findings for the four PVSP voice domains of loudness, pitch, laryngeal quality, and resonance. Consistent with findings in the extensive literature on the auditory-perceptual identification and rating of voice disorders, interjudge and intrajudge percentage of agreement data for the four voice domains are often below a nominally acceptable 80%. PVSP data in clinical-research studies are particularly advisory, requiring instrumental analysis to validate and quantify types and levels of vocal and resonance involvement. These constraints notwithstanding, we have found it useful to retain auditory-perceptual voice codes in this screening procedure, allowing the vocal or signal aspects of nonsegmental data to be coded independently of prosodic features. As discussed below, there have been only a few findings to date in which differences in one or more voice domains differentiate among children with speech delay or among speakers in different clinical populations.

Definition and coding criteria for loudness

The PVSP criterion for appropriate loudness of an utterance is met when 'at least 50% of the words [are] appropriately loud for the sampling content/context and the speaker's age, gender, dialect, and current emotional state'. For utterances longer than six words, no more than three of the words can exceed normal loudness levels. The two inappropriate loudness codes are used when an utterance is unacceptably too soft (PV17) or too loud (PV18). The PVSP guidelines for loudness coding includes information on recording and microphone placement techniques that are used to assess relative loudness.

Definition and coding criteria for pitch

The PVSP manual defines appropriate pitch as 'within normal limits for the speaker's age and gender ... the pitch level should not "call attention to itself" in a social situation'. The two inappropriate codes used for utterances in which pitch is too low include PV19: Low Pitch/Glottal Fry (low pitch accompanied by glottal fry register distributed across an utterance) and PV20: Low Pitch. The two codes used for inappropriately high pitch include PV21: High Pitch/Falsetto (high pitch accompanied by a voice quality change termed falsetto register) and PV22: High Pitch (high pitch not perceived to be in a falsetto register).

Definition and coding criteria for laryngeal quality

The PVSP general definition for appropriate laryngeal quality is 'vocal characteristics that are within the normal range for the speaker's age, gender, and dialect'. PV23: Breathy (audible airflow at least equal to the accompanying laryngeal tone), PV24: Rough (an aperiodic, 'gravelly' sounding voice), and PV25: Strained (a strident vocal tone) codes are used when at least 50% of the words within an utterance (or at least three words in utterances that are 6+ words long) are judged as inappropriately breathy, rough, or strained. Just one inappropriate vocal occurrence is sufficient for the use of PV26: Break/Shift/Tremulous (voice break, pitch shift, tremulous voice), PV27: Register Break (a break from modal to falsetto register within a syllable or between two syllables) and PV28: Diplophonia (the simultaneous production of two vocal pitches). Finally, PV29: Multiple Laryngeal Features is used for combinations of vocal quality or behaviours not already built into the codes themselves.

Definition and coding criteria for resonance

The general definition for appropriate resonance is similar to definitions for the other PVSP voice domains: '... oral and nasal features that are within the normal range for a speaker's age, gender, and dialect'. Inappropriate resonance often results in a less intense vocal signal that may be inappropriately nasal (PV30) in vowel or diphthong contexts, denasal (PV31) in contexts where nasality is expected or appropriate, or nasopharyngeal (PV32), with the last code being used to capture a 'backed', 'muffled' form of inappropriate resonance that is neither nasal nor denasal.

Typical speakers

Most conversational utterances from typical speakers have appropriate loudness and pitch (table 1), with the adult sample in table 2 indicating that approximately 92% of typical adult speakers have appropriate pitch on at least 80% of their utterances. The data in table 1 on laryngeal quality in these speakers are somewhat lower, averaging in the upper 80% range of appropriate utterances for the two samples of typically speaking children, but in the mid 70% range for typically speaking adults. The per-individual percentages in table 2 illustrate the differences in the two ways of summarizing the data, with approximately 17% of typically speaking children and 32%, or nearly one of every three typically speaking adults, scoring below 80% appropriate laryngeal quality. Such findings are not inconsistent with survey data indicating a high prevalence of voice differences and disorders in children and adults (see reviews in Ramig and Verdolini, 1998). The resonance data in tables 1 and 2 indicate that relatively few speakers with typical speech acquisition have frequently occurring inappropriate resonance in conversational speech samples.

Speech delay

The sample estimates in tables 1 and 2 indicate that most children with speech delay have appropriate loudness and pitch in conversational speech, with approximately 86–89% of children having appropriate loudness and/or pitch on at least 80% of their utterances. Similar to the findings for typically speaking adults, appropriate laryngeal quality was coded for only approximately 71–75% of utterances, with more than one of every three children with speech delay having fewer than 80% utterances coded as having appropriate laryngeal quality. Resonance scores averaged in the pass or questionable pass range for children with speech delay. The one available resonance estimate in table 2 indicates that only approximately 7% of children have fewer than 80% of utterances with appropriate resonance.

Clinical populations

Children and adults in the AOS, MR, HFA and AS groups also had appropriate loudness and pitch, with scores in all but one of the studies in table 1 above 90% appropriate utterances. Group-averaged scores for laryngeal quality across eight of the nine studies in table 1 were once again considerably lower than the scores for loudness and pitch, ranging from approximately 69% to 94% appropriate utterances. Finally, considerably lowered appropriate resonance scores, averaging in the mid

60% range, were obtained for four of the five samples of children with AOS, whereas the one sample of adults with AOS averaged nearly 98% appropriate resonance. Resonance scores for speakers in the HFA and AS groups were also lower, averaging approximately 77% and 81%, respectively.

Conclusion

This overview of prosody-voice findings in speech samples from several typically speaking and clinical populations illustrates the range of potentially useful information on suprasegmentals—information that, to date, has remained relatively unstudied in communicative disorders. The following observations are offered in the spirit of promoting further research and development in clinical aspects of prosody.

For most of the seven PVSP summative domains and their respective codes there are few comparable data available to assess the validity of the descriptive findings summarized in table 2. In children with speech delay, for example, most available epidemiologic data on the proportion of children with suprasegmental involvements are qualitative and anecdotal (cf. Shriberg, Tomblin and McSweeny, 1999). Thus, the concurrent and criterion validity of these data remain untested. Cross-sample consistencies in the data obtained on different samples in tables 1 and 2 lend some confidence to claims for the validity and reliability of these data and PVSP descriptions of several clinical populations have begun to appear (e.g., Campbell *et al.*, 2000; Campbell and Dollaghan, 1995; Del Duca, Petrille, Bourne, Leavy, McMillin, Wieting, Dollaghan, Davis, Liddell and Metzler, 1998; Pittelko, 2000; Porteous, 1995). Cross-laboratory work has also begun on two acoustic-based protocols that may have the sensitivity, reliability and efficiency required for fine-grained analysis of lexical and phrasal stress.

Assessment procedures such as those described in this report and elsewhere in this volume are constrained by efficiency concerns, particularly the investment of time needed to acquire the requisite skills and knowledge to code a speech sample. Skills in the assessment of prosody are not routinely taught in academic training programmes in communicative disorders, compared, for example, to the training provided in phonetic transcription and analysis of language samples. Obligatory training in prosody assessment will likely be motivated only if the correct diagnosis and funding for treatment of an individual with a communicative disorder is crucially dependent on quantitative data on his or her prosody status. In this regard, we are particularly encouraged by the preliminary studies indicating that such a role for prosody might be found for children with suspected apraxia of speech, mental retardation, or pervasive developmental disability.

Finally, from a discipline perspective, it might be noted that investigators are generally wary of assessment data obtained using noninstrumental auditory-perceptual procedures. For the present concerns, the complexity of methods used to code inappropriate prosody make it especially difficult to study theoretically and clinically relevant aspects of prosody in natural discourse and to clearly report findings. Until acoustic-based procedures are widely available to assess prosody-voice, possibly aided by preprocessing with speech recognition software, procedurally disparate auditory-perceptual approaches such as those described and cited in this volume remain the only way to address this important assessment task. From this perspective, we are especially pleased to be a part of this volume and heartily endorse the editor's goal of encouraging increased research in clinical aspects of prosody.

Acknowledgements

Our sincere thanks to: David Wilson for his significant contributions to the development of the PVSP software; Chad Allen for his editorial assistance with this manuscript; Joan Kwiatkowski, Carmen Rasmussen and Carol Widder, who, along with the authors, were the PVSP coders for the data described in this report; and Sheryl Hall and Jennine Sprangers, PVSP coders to whom the torch has been passed. This paper was supported by the National Institutes of Deafness and Other Communication Disorders, National Institutes of Health, DC00496.

References

Bernthal, J. E. and Bankson, N. W., 1998, *Articulation and phonological disorders* (Boston, MA: Allyn and Bacon).

Campbell, T. F., Adelson, D., Dollaghan, C. A., Sabo, D., Janosky, J. and Feldman, H., 2000, Speech change in children after traumatic brain injury. Manuscript in preparation.

Campbell, T. F and Dollaghan, C. A., 1995, Speech production characteristics of children following traumatic brain injury. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, **4**, 476–479.

Del Duca, G., Petrille, L., Bourne, P., Leavy, H., McMillin, A., Wieting, S., Dollaghan, C., Davis, D., Liddell, M and Metzler, R., 1998, Prosodic features of children with language impairments. Poster session presented at the Symposium for Research in Child Language Disorders. University of Wisconsin-Madison, June.

Flipsen, P., Jr., 1999, Articulation rate and speech-sound normalization following speech delay. Unpublished doctoral dissertation, University of Wisconsin-Madison, USA.

Kent, R. D., 1996, Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology*, **5**, 7–23.

Lewis, B. A and Shriberg, L. D., 1994, Life span interrelationships among speech, prosody-voice and nontraditional phonological measures. Miniseminar presented at the *Annual Convention of the American Speech-Language-Hearing Association*, New Orleans, LA, November.

Miller, J. F and Chapman, R. S., 1992, *Systematic Analysis of Language Transcripts (SALT)* (Madison, WI: Language Analysis Laboratory, Waisman Center).

Morrison, J. A and Shriberg, L. D., 1992, Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research*, **35**, 259–273.

Odell, K. H. and Shriberg, L. D., 2001, Prosody-voice characteristics of children and adults with apraxia of speech. *Clinical Linguistics and Phonetics*, **15**, 275–307.

Pittelko, S. L., 2000, Prosody-voice characteristics of children with Williams syndrome. Unpublished master's thesis, Minnesota State University, Mankato, USA.

Porteous, E. A., 1995, Relationship between prosody-voice and speech intelligibility scores in dysarthric children. Unpublished master's thesis, University of Alberta, Edmonton, Alberta, Canada.

Ramig, L. O and Verdolini, K., 1998, Treatment efficacy: Voice disorders. *Journal of Speech, Language, and Hearing Research*, **41**, S101–S116.

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

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

Shriberg, L. D., Allen, C. T., McSweeny, J. L. and Wilson, D. L., 2000, PEPPER: Programs to examine phonetic and phonologic evaluation records [Computer software] (Madison, WI: Waisman Center, University of Wisconsin).

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

Shriberg, L. D., Aram, D. M. and Kwiatkowski, J., 1997b, Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *Journal of Speech, Language, and Hearing Research*, **40**, 313–337.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L. and Wilson, D. L., 1997, The Speech Disorders Classification System (SDCS): Extensions and lifespan reference data. *Journal of Speech, Language, and Hearing Research*, **40**, 723–740.

Shriberg, L. D. and Kwiatkowski, J., 1982a, Phonological disorders I: A diagnostic classification system. *Journal of Speech and Hearing Disorders*, **47**, 226–241.

Shriberg, L. D. and Kwiatkowski, J., 1982b, Phonological disorders II: A conceptual framework for management. *Journal of Speech and Hearing Disorders*, **47**, 242–256.

Shriberg, L. D. and Kwiatkowski, J., 1983, Computer-assisted Natural Process Analysis (NPA): Recent issues and data. In J. Locke (ed.) *Assessing and treating phonological disorders: current approaches. Seminars in Speech and Language*, **4**, 389–406.

Shriberg, L. D. and 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. and Kwiatkowski, J., 1994, Developmental phonological disorders I: A clinical profile. *Journal of Speech and Hearing Research*, **37**, 1100–1126.

Shriberg, L. D., Kwiatkowski, J., Best, S., Hengst, J. and Terselic-Weber, B., 1986, Characteristics of children with phonologic disorders of unknown origin. *Journal of Speech and Hearing Disorders*, **51**, 140–161.

Shriberg, L. D., Kwiatkowski, J. and Rasmussen, C., 1989, 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, USA, November.

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

Shriberg, L. D., Kwiatkowski, J., Rasmussen, C., Lof, G. L. and Miller, J. F., 1992, The Prosody-Voice Screening Profile (PVSP): Psychometric data and reference information for children (Tech. Rep. No. 1). Phonology Project, Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison.

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

Shriberg, L. D. and Nockerts, A., 1982, A speech analysis program for the Harris/800 minicomputer. Waisman Center Research Computing Facility, University of Wisconsin, Madison, USA.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Volkmar, F. R. and Cohen, D. J., in press, Speech and prosody characteristics of adolescents and adults with High Functioning Autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research*.

Shriberg, L. D., Tomblin, J. B. and McSweeny, J. L., 1999, Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language, and Hearing Research*, **42**, 1461–1481.

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

Siegel, S. and Castellan, N. J., Jr., 1988, Nonparametric statistics for the behavioral sciences, second edition (New York: McGraw-Hill).

Velleman, S. L. and Shriberg, L. D., 1999, Metrical analysis of the speech of children with suspected developmental apraxia of speech. *Journal of Speech, Language, and Hearing Research*, **42**, 1444–1460.

Appendix A: administration and coding of the PVSP

The purpose of this appendix is to provide a brief overview of procedural aspects of the PVSP. This information should provide a methodological context for the research findings summarized in this report and may be of use to readers interested in developing alternative approaches to prosody-voice assessment. Software for the

PVSP provides considerable efficiencies in data entry and data analyses, but the data collection and data coding procedures are essentially similar for paper-pencil and computer versions. The present review is from the perspective of the paper-pencil version of the PVSP, with the entire process divided into the following four phases. These guidelines underscore the many potentially confounding variables that require controlled procedures when attempting to quantify a speaker's prosody and voice characteristics from a recorded conversational speech sample.

Obtaining a Conversational Speech Sample for PVSP Coding

The first task is to obtain a conversational speech sample of sufficient length to include the required number of utterances for coding on the seven prosody-voice parameters: phrasing, rate, stress, pitch, loudness, laryngeal quality and resonance. The recommended target is 24 utterances that meet eligibility criteria for prosody-voice coding. The examiner who obtains the conversational speech sample must follow a set of procedural guidelines to ensure that the linguistic content and audio or video recording of the sample meet the needs for valid prosody-voice coding, as well as for broad and narrow phonetic transcription if segmental analyses will also be completed. The PVSP manual (Shriberg *et al.*, 1990) includes the following five guidelines for obtaining speech samples with good linguistic content.

1. Establish rapport with the speaker before prompting conversation. The position of the speech sample within the framework of a larger assessment protocol requires careful consideration and should probably take place after the speaker has had an opportunity to become comfortable with the examiner and the test situation. This will help ensure a more natural conversational interchange between the examiner and the speaker.
2. Prepare the speaker physically for conversational speech sampling. The speaker should be physically comfortable at the beginning and throughout the sample. The examiner has the option of informing the speaker that she is obtaining a conversational speech sample, using language appropriate to the age of the speaker. This indicates to the speaker that the conversational sample is part of the protocol, which clarifies the speaker's understanding that, although less structured than other tasks in the protocol, it is an important part of the protocol.
3. Use topics and materials that are within the speaker's cognitive level and interests. An examiner may need age-appropriate materials, such as pictures or books to prompt conversational speech from young or less verbal speakers. Experienced examiners have lists of topics and associated materials available to successfully evoke conversational speech from speakers of different ages and cognitive backgrounds.
4. Monitor examiner prosody-voice throughout the sample. The examiner must be careful not to model inappropriate prosody or voice characteristics (e.g., narrative register, character register) that could influence the speaker's prosody or voice characteristics.
5. Obtain at least 12 codable utterances. Twelve utterances is the minimum number of utterances needed for a reliable prosody-voice screening assessment (see Appendix B) and is likely to be sufficient and efficient for speakers with

typical prosody-voice or evident utterance-to-utterance consistency in prosody-voice involvement. For speakers with notable but variable involvement in one or more of the prosody and voice domains, the clinical-research recommendation is to obtain 24 codable utterances.

In addition to these five sampling guidelines, technical and environmental factors must be monitored to ensure a high-quality speech sample. The recording equipment should be of high quality and well maintained. An external microphone should be used, and a recorder with a VU indicator is desirable so that input levels can be set and monitored throughout the recording session. Environmental noise should be minimized so that there are no sounds that could potentially interfere with the transcription or coding of the speech sample. For this reason, the use of noisy toys during the speech sample is discouraged. The microphone should be placed approximately 6 inches from the speaker's mouth, and this lip-to-microphone distance should be maintained throughout the sample. A head-mounted microphone can also be used and is preferable if acoustic analyses will be performed.

The 31 exclusion codes (figure 2) discussed next require additional procedures to maximize the quality and efficiency of speech sampling. As reviewed in detail in the references cited, additional speech sampling strategies include: minimizing questions that require no more than a single-word response (Exclusion Code C7); glossing utterances during the conversation to minimize the number of utterances that would be excluded due to unintelligibility (Exclusion Code C12); avoiding interruptions or overtalk (Exclusion Code C5); avoiding test situations that could encourage the use of character or narrative registers (such as role-playing or describing books or movies; Exclusion Codes R1 and R2); and monitoring body movements and avoiding topics or situations that encourage the speaker to sing, whisper, use sound effects, or laugh excessively (Environment, Register and States codes). As noted in the review of PVSP findings, the occurrence of high rates of such behaviours, despite examiner precautions, may itself provide information on paralinguistic aspects of a speaker's communicative deficits.

Glossing and Segmenting a Speech Sample for PVSP Coding

The second step in the PVSP procedure requires glossing and segmenting the utterances in the speech sample using a playback device to listen to an audio or video recording of the sample. Due to the variance associated with technical differences in audiocassette technology, all such listening for audiocassette tapes to date has been standardized to one analogue device. Clinical researchers interested in using digital audiotape procedures will need to explore such technical issues when using, for example, the audio output from microcomputer speakers for glossing and coding PVSP data. The speaker on the playback unit should face the listener, and the volume may need to be turned up and/or the playback speed may need to be slowed down to capture faint or fleeting noises (such as certain intermittent behaviours or part-word repetitions). A playback device that allows the listener to easily control loudness, speed, and the number and length of repeats is desirable.

Making an accurate, exact written gloss of each utterance is vital to the validity of the coding procedure. The PVSP procedural manual includes examples and guidelines for the variety of glossing and segmentation decisions that occur in conversational samples from typical and atypical speakers. As with transcriptions for the purposes of language analyses and speech-sound analyses, glossing for the purposes of prosody-voice analyses requires conventions for semantic, syntactic, and morphological decisions that ultimately affect the validity and clinical utility of the analysis.

Excluding Utterances and Coding Utterances for Prosody-Voice

The third step in the PVSP procedure is to go through the sample a second time and exclude those utterances that meet the criteria for one or more of the 31 exclusion codes listed in figure 2 and to code non-excluded utterances for each of the prosody-voice domains. The PVSP manual contains detailed information on the criteria for each of the inappropriate codes and descriptions of exemplars on the accompanying audiocassette training tape. Several of the inappropriate prosody-voice codes provide a way to indicate utterances with multiple occurrences of inappropriate prosody-voice. A record kept in the Notes section of the Prosody-Voice Coding Log provides additional descriptive information. For example, PV15: Excessive/Equal/Misplaced Stress includes four types of behaviour that coders annotate to provide the specific source of the behaviour(s) used to code stress as inappropriate. These four behaviours include excessive/equal stress at the phrase or utterance level, misplaced stress relative to normal phrasal or emphatic stress, sound blocks and sound prolongations. Similarly, PV26: Break/Shift/Tremulous is an inappropriate voice code that includes three types of inappropriate vocal behaviours; the examiner annotates which one or more of these vocal behaviours was the basis for coding the utterance as inappropriate for PV26. Rationale for the use of multiple codes is provided in the references cited, essentially reflecting validity and reliability constraints obtained in the psychometric studies.

In paper-pencil PVSP processing, prosody-voice codes are recorded on a four-page PVSP Scoring Form that includes speaker and sample identification information, utterance-by-utterance tallies of exclusion and prosody-voice codes, a section to graph summative scores, and space to annotate comments and recommendations regarding the validity of the sample and any other information that may aid interpretation of the screening results. The computerized version provides for similar information to be stored and retrieved using database formats.

Analysis and Interpretation of PVSP Data

The final steps in the PVSP differ considerably in the computer compared to pencil and paper versions of the procedure, with the software providing considerably more analytic information. Whichever version is used, findings should be viewed as screening information, with the limitations attendant to auditory-perceptual data obtained in a single sample of behavioural data. The psychometric studies described in Appendix B provide validity support for dividing the summative PVSP scores on the seven dimensions into a three-way ordinal classification. For each of the seven suprasegmentals, summative scores of 90% or above appropriate indicate a pass, scores of 80.0% to 89.9% indicate a questionable pass, and scores below 80.0% indicate a fail. In keeping with customary screening conventions, higher rates of false positives (speakers with fails on the PVSP who have appropriate prosody-voice) are more desirable than false negatives (speakers who pass the PVSP and have inappropriate prosody-voice). In the clinical diagnostic context, failing scores indicate areas of prosody-voice that require further review and assessment using other available methods to make clinical decisions and to monitor progress.

Appendix B: psychometric studies of the PVSP

Validity Studies

Face, content and consensual validity for the PVSP were developed and supported from a number of sources and studies, including reviews of the literature on prosody and voice in descriptive linguistics and clinical speech pathology; discussions of terms and concepts with researchers and clinicians experienced in the areas of speech, language, fluency, voice, and hearing disorders; and pilot studies that yielded necessary and sufficient exclusion and prosody-voice codes, based on several hundred conversational speech samples from speakers with typical and disordered speech.

Criterion and concurrent validity for the PVSP was assessed in acoustic studies for five of the seven prosody-voice variables, including instrumental studies of rate, stress, pitch, laryngeal quality and resonance quality (Shriberg *et al.*, 1992). Criterion validity studies for these variables, as well as for phrasing and loudness, included comparisons of PVSP findings with the perceptual decisions of a panel of expert listeners. Concurrent validity was assessed by comparing data from the prior studies that used the first version of the prosody-voice coding procedure to estimates obtained using the present version of the PVSP. Acceptably close results were obtained for all but two of the seven prosody-voice domains. Differences between the two versions on speakers' rate and pitch appeared to be due to the more stringent criteria for inappropriate in the present version of the PVSP.

Reliability Studies

Conversational speech sampling. The stability of conversational speech samples has been assessed in considerable detail, including information on intra- and inter-sample occurrence of consonants and vowel/diphthongs, canonical forms, parts of speech, type-token ratios per minute, intelligible words per minute, speech registers, and the effects of alternative sampling modes (e.g., Morrison and Shriberg, 1992; Shriberg and Kwiatkowski, 1983, 1985; Shriberg *et al.*, 1986; Shriberg and Lof, 1991; Shriberg and Widder, 1990). Essentially, studies indicate that conversational speech samples obtained using standardized evocation and recording conventions are quite robust relative to the linguistic stability and signal fidelity needed for prosody-voice studies.

Internal consistency studies. Two studies support the internal consistency of the PVSP (Shriberg *et al.*, 1992). One study compared the first 12 utterances to the second 12 utterances in 24-utterance samples of 64 children with developmental phonological disorders. Using a 90% pass criterion for each suprasegmental, it was found that 80–100% of the retest decisions were similar for those speakers that passed based on the first 12 utterances. Too few speakers received failing scores on the first 12 utterances to assess their stability. A second study split the transcripts of 40 speakers with normal speech and 40 speakers with speech delay into odd and even utterances; similar scores were obtained for the odd compared to the even utterances. The internal consistency of the pass scores was again found to be more stable than the fail scores, but too few fail scores were available for inferential statistical comparison.

Prosody-voice coding. Estimates of interjudge and intrajudge agreement for each phase of the PVSP procedures summarized in Appendix A were obtained in the developmental studies, and have been reported in nearly all of the clinical studies using the procedure.

For segmenting conversational speech, estimates of average interjudge agreement have ranged from 90–100% (Shriberg, Kwiatkowski and Rasmussen, 1989; Shriberg *et al.*, 1990). The detailed segmentation rules provided in the PVSP manual appear to be sufficient for even inexperienced users to obtain high reliability in segmenting a conversational speech sample for prosody-voice coding.

For exclusion coding, estimates of average agreement on exclusion codes have been essentially similar, with average intrajudge percentages of agreement ranging from 71% to 100% and average interjudge percentages of agreement ranging from 76% to 100% (Shriberg *et al.*, 1989; Shriberg *et al.*, 1992). For prosody-voice coding, estimates of intrajudge and interjudge agreement have been obtained both in the developmental studies and in nearly all of the clinical studies with this procedure. In the developmental studies, average intrajudge agreement ranged from 85% to 99% and average interjudge agreement ranged from 78% to 96%. In the clinical studies, including four separate estimates, interjudge and intrajudge agreement has ranged from 74% to 99%. These agreement estimates reflect only agreement at the summative level averaged across the seven prosody-voice domains. Agreement in the use of the inappropriate codes within each domain are notably lower, depending on the severity of involvement of the speakers in each study. Thus, although the summative data are generally reliable for screening purposes, they are not considered reliable at the level of codes for clinical diagnostic purposes. For such needs, especially for the three voice suprasegmentals, additional assessment is warranted using instrumental measures.

Efficiency Studies

Efficiency data on administering the PVSP are assessed by estimates of the time required to obtain a speech sample long enough to yield the required number of codable utterances. An approximately 10 minute conversational speech sample has been found to be sufficient for the various speech and prosody-voice procedures and analyses used for research purposes, assuming that the speaker produces an average number of utterances, most of which would not be excluded using PVSP guidelines. The same conversational speech sample can be used to assess a speaker's speech and language as well as prosody-voice, an efficiency issue that is not always acknowledged when estimating the efficiency of procedures separately in each domain.

Efficiency data on the time needed to code utterances for PVSP analysis are described in Shriberg *et al.* (1992). Pilot and larger group studies indicated that examiners require approximately 2 minutes per utterance to code the seven suprasegmentals as appropriate or inappropriate. This average varies greatly depending upon the speech and prosody-voice involvement of the speaker, the quality of the sample, the efficiency and skill of the examiner who obtained the conversational speech sample, and, most importantly, on the experience and auditory-perceptual skills of the coder.

A related efficiency issue concerns the number of utterances in a conversational speech sample that must be excluded from prosody-voice coding. As described in the PVSP manual, from 15% to 85% of utterances may need to be excluded from prosody-voice coding because they meet criteria for one or more of the exclusion codes. For a group of young children with speech delay, the average percentage of excluded utterances was 51% with a standard deviation of 17%. Of the excluded utterances, 87% were due to three codes: C7: Only One Word, C12: Too Many Unintelligibles, and R2: Narrative Register. As suggested in Appendix A, the use of speech sampling procedures that minimize the occurrence of these types of utterances greatly increases the efficiency of the prosody-voice coding.

Finally, several studies have provided efficiency data on the time needed to learn the PVSP, including an initial study involving two clinical instructors, an independent learning group study, and a group learners study (Shriberg *et al.*, 1992). For the independent learning group study, in which three students were recruited to learn the PVSP procedure on their own using only the training materials, the average training time was 15 hours, 17 minutes. As with other procedures for prosody assessment cited in this volume, acquisition of the skills required to complete prosody-voice analysis with the PVSP is a challenging task.