

SEGMENTAL AND SUPRASEGMENTAL TRANSCRIPTION RELIABILITY

Phonology Project Technical Report No. 2

Jane L. McSweeney

Lawrence D. Shriberg

Transcriptionists

Joan Kwiatkowski

Jane McSweeney

Carmen Rasmussen

Carol Widder

December, 1995

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

Preparation of this paper was supported by research grants 5 R01 DC 00496-07 (Lawrence D. Shriberg, P.I.) and 5 R01 DC 00528 (Barbara A. Lewis, P.I.) from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health

TABLE OF CONTENTS

INTRODUCTION	3
I. INTERJUDGE RELIABILITY	5
A. Broad and Narrow Transcription Agreement	5
Carmen and Carol	5
Joan and Carmen	8
Joan and Carol	8
Carmen and Jane	9
Carol and Jane	9
B. Diacritic Agreement	10
How to Read the Table	10
Diacritics: Explanation and Analyses	12
Nasality	12
Lip	13
Stop Release	13
Juncture/Stress	14
Tongue Configuration	14
Tongue Position	16
Sound Source	17
Timing/Other	19
Summary	20
Diacritic Agreement Table	23
C. PVSP Agreement	32
How to Read the Tables	32
What Does It All Mean?	36
Phrasing	36
Rate	37
Stress	38
Loudness	39
Pitch	40
Laryngeal Features	40
Resonance	41
Summary	42
Miscellaneous Notes	42
PVSP Comparison Tables	44
II. INTRAJUDGE RELIABILITY	50
A. Carmen	50
Broad and Narrow Phonetic Transcription Agreement	50

Ways to Improve Intrajudge Phonetic Transcription Agreement	50
Prosody-Voice Screening Profile Intrajudge Agreement	51
Phrasing	53
Rate	53
Stress	53
Loudness	53
Pitch	54
Laryngeal Features	54
Resonance	54
Summary	54
PVSP Comparison Table	55
B. Carol	57
Broad and Narrow Phonetic Transcription Agreement	57
Ways to Improve Intrajudge Phonetic Transcription Agreement	57
Prosody-Voice Screening Profile Intrajudge Agreement	58
Phrasing	60
Rate	60
Stress	60
Loudness	61
Pitch	61
Laryngeal Features	61
Resonance	62
Summary	62
PVSP Comparison Table	63

III. TRAINING TAPE (Enclosed)

INTRODUCTION

Background

The Phonology Project processes speech data obtained from several sites in Wisconsin and a number of research centers across the country. Periodic reliability estimates are conducted for all stages of data collection and data reduction, including phonetic transcription and coding prosody-voice status. The specific goals of these estimates are (a) to provide reliability data for the empirical studies, and (b) to continue to identify ways to maximize the validity and reliability of our speech data.

This technical paper reports interjudge and intrajudge reliability for four transcriptionists who have transcribed and coded segmental and suprasegmental data on children from several studies. References to phonetic transcription are based on the system described in Shriberg and Kent (1995), the phonetic and diacritic symbols are produced in WordPerfect using the PEPPER Font (Shriberg, Wilson, & Austin, 1995), and all references to prosody-voice are described in Shriberg, Kwiatkowski, and Rasmussen (1990). Phonetic transcription reliability data were calculated and displayed by means of the PEPAGREE module in PEPPER.

The style of this in-house report is informal, directed specifically at the two goals above. Proper nouns reflecting the short-hand language of the laboratory are used as the most direct way to identify people, studies, and research sites. A narrative style is used to preserve methodological detail, including rationale and data motivating suggestions to improve reliability.

Summary of Findings

Overall, the reliability findings are consistent with prior studies (e.g., Shriberg & Lof, 1991). Broad transcription agreement is well within criteria required for effect sizes anticipated in forthcoming Phonology Project studies (see standard error of measurement project below). However, narrow phonetic transcription, including point-to-point percentages for diacritics, does not meet reliability criteria for certain questions. The same finding obtains for PVSP reliability, with lower levels of description less reliable than summative values.

Transcription reliability needs have been addressed in four ways. First, a training tape based on this report will aid transcriptionists with the most difficult perceptual transcription and coding

tasks. Second, a series of recently validated metrics classify all speech-sound distortions as correct, thus exploiting the demonstrated reliability of broad phonetic transcription (Shriberg, 1993; Shriberg, Austin, Lewis, McSweeny, & Wilson, in press-a). Third, sections of the interjudge reliability data in this report have been used to calculate standard errors of measurement for all metrics used in the Phonology Project (Shriberg et al., in press-a). Finally, a two-year technology project in progress will develop an acoustic-aided procedure for computer-based transcription and prosody-voice coding.

Kudos

Hats off to Jane McSweeny for organizing, completing, and writing up this complex project with her typical intelligence and clear, good-humored style. Also, congratulations, Jane, for successfully entering the prestigious inner circle of research transcriptionists! Thanks to Rachel Phillips for a superb job with the many tables in this report.

Thanks to Joan Kwiatkowski for her single-handed transcription of incredibly l-o-n-g samples from 25 speech-delayed children tested as many as five times over a two-year period.

Very deep bows of appreciation and admiration to Carmen Rasmussen and Carol Widder, long-term colleagues whose skills and insights continue to be reflected in each and every Phonology Project study. May the next box of speech samples be a piece of cake.

LDS

INTERJUDGE RELIABILITY

Broad and Narrow Transcription Agreement: All Comparison Groups

Below are the summary percentage agreement figures for these comparison groups/pairs: Carmen and Carol (IOWA1, Lewis, PRED2, and Gregg's dissertation samples), Joan and Carmen (PRED2), Joan and Carol (PRED2), Carmen and Jane (Gregg's dissertation samples), and Carol and Jane (Gregg's dissertation samples). The "Ø" represents the underbar diacritic symbol (i.e., deletions of phonemes in the z-line), and agreement percentages are provided with and without the deletions included in the calculations. At the end of the Carmen-Carol transcriptionist pair section are percentage agreements for all of the studies combined. (They are the only pair in which transcripts from more than one study were compared.)

CARMEN - CAROL	STUDY: IOWA1	# of Transcripts: 10
<u>Consonants</u>		<u>Range of Percentages</u>
Narrow Agreement: 82.5%		74.9 - 89.0%
Broad Agreement w/Ø: 90.9%		85.2 - 94.8%
Broad Agreement w/o Ø: 94.6%		90.2 - 97.4%
<u>Vowels</u>		<u>Range of Percentages</u>
Narrow Agreement: 80.5%		66.0 - 86.5%
Broad Agreement w/Ø: 87.5%		83.0 - 93.4%
Broad Agreement w/o Ø: 87.7%		83.3 - 93.4%
<u>Diacritics</u>		<u>Range of Percentages</u>
Overall Agreement: 38.7% *		12.0 - 64.2%

* Only 9 of the 10 transcripts included here--could not generate PEPAGREE diacritic printout for transcript ESHOO-C1.

CARMEN - CAROL

STUDY: Lewis

of Transcripts: 12

Consonants

Narrow Agreement: 85.3%

Broad Agreement w/Ø: 90.9%

Broad Agreement w/o Ø: 95.1%

Vowels

Narrow Agreement: 80.5%

Broad Agreement w/Ø: 85.4%

Broad Agreement w/o Ø: 85.7%

Diacritics

Overall Agreement: 28.5%

Range of Percentages

65.3 - 96.8%

77.3 - 97.2%

86.4 - 98.2%

Range of Percentages

65.8 - 86.3%

77.2 - 89.1%

77.8 - 89.6%

Range of Percentages

0.0 - 50.0%

CARMEN - CAROL

STUDY: PRED2

of Transcripts: 7

Consonants

Narrow Agreement: 73.9%

Broad Agreement w/Ø: 85.1%

Broad Agreement w/o Ø: 90.0%

Vowels

Narrow Agreement: 71.2%

Broad Agreement w/Ø: 85.0%

Broad Agreement w/o Ø: 86.0%

Diacritics

N/A

Range of Percentages

69.1 - 77.3%

80.0 - 89.7%

86.5 - 93.7%

Range of Percentages

61.4 - 80.0%

77.6 - 90.4%

79.1 - 91.5%

Range of Percentages

CARMEN - CAROL

STUDY: Gregg's

of Transcripts: 3

Consonants

Narrow Agreement: 83.0%

Broad Agreement w/Ø: 89.6%

Broad Agreement w/o Ø: 93.7%

Vowels

Narrow Agreement: 93.2%

Broad Agreement w/Ø: 96.9%

Broad Agreement w/o Ø: 97.3%

Diacritics

Overall Agreement: 52.0%

Range of Percentages

80.3 - 85.5%

86.5 - 93.0%

92.3 - 94.7%

Range of Percentages

88.0 - 96.1%

94.3 - 99.0%

95.5 - 99.0%

Range of Percentages

49.4 - 53.4%

CARMEN - CAROL

STUDY: All

of Transcripts: 32

Consonants

Narrow Agreement: 82.2%

Broad Agreement w/Ø: 89.8%

Broad Agreement w/o Ø: 94.0%

Vowels

Narrow Agreement: 80.2%

Broad Agreement w/Ø: 87.1%

Broad Agreement w/o Ø: 87.5%

Diacritics*

Overall Agreement: 36.0%

Range of Percentages

73.9 - 85.3%

85.1 - 90.9%

90.0 - 95.1%

Range of Percentages

71.2 - 93.2%

85.0 - 96.9%

85.7 - 97.3%

Range of Percentages

28.5 - 52.0%

* Diacritic information available for 24 transcripts only.

JOAN - CARMEN

STUDY: PRED2

of Transcripts: 7

Consonants

Narrow Agreement: 77.7%

Broad Agreement w/Ø: 84.1%

Broad Agreement w/o Ø: 89.3%

Vowels

Narrow Agreement: 74.1%

Broad Agreement w/Ø: 81.1%

Broad Agreement w/o Ø: 81.8%

Diacritics

Overall Agreement: 24.8%

Range of Percentages

67.9 - 82.5%

75.2 - 90.0%

86.0 - 93.9%

Range of Percentages

69.0 - 82.0%

72.4 - 90.4%

73.6 - 91.5%

Range of Percentages

9.5 - 53.3%

JOAN - CAROL

STUDY: PRED2

of Transcripts: 7

Consonants

Narrow Agreement: 71.1%

Broad Agreement w/Ø: 82.0%

Broad Agreement w/o Ø: 88.6%

Vowels

Narrow Agreement: 70.3%

Broad Agreement w/Ø: 80.5%

Broad Agreement w/o Ø: 81.3%

Diacritics

Overall Agreement: 12.2%

Range of Percentages

64.5 - 77.4%

77.0 - 86.2%

85.5 - 91.5%

Range of Percentages

62.4 - 75.4%

69.9 - 89.2%

71.0 - 90.3%

Range of Percentages

7.2 - 16.9%

CARMEN - JANE

STUDY: Gregg's

of Transcripts: 5

Consonants

Narrow Agreement: 74.7%

Broad Agreement w/Ø: 82.5%

Broad Agreement w/o Ø: 87.3%

Vowels

Narrow Agreement: 78.7%

Broad Agreement w/Ø: 85.9%

Broad Agreement w/o Ø: 86.5%

Diacritics

Overall Agreement: 28.8%

Range of Percentages

70.1 - 81.0%

74.0 - 89.8%

80.9 - 92.6%

Range of Percentages

66.0 - 84.2%

79.5 - 89.9%

80.9 - 89.9%

Range of Percentages

19.8 - 42.0%

CAROL - JANE

STUDY: Gregg's

of Transcripts: 3

Consonants

Narrow Agreement: 74.6%

Broad Agreement w/Ø: 84.2%

Broad Agreement w/o Ø: 89.6%

Vowels

Narrow Agreement: 79.5%

Broad Agreement w/Ø: 87.0%

Broad Agreement w/o Ø: 87.7%

Diacritics

Overall Agreement: 25.5%

Range of Percentages

66.9 - 80.7%

78.8 - 89.3%

86.7 - 92.3%

Range of Percentages

75.6 - 81.6%

84.5 - 88.6%

84.9 - 89.7%

Range of Percentages

18.6 - 29.7%

Diacritic Agreement

How to Read the Table

The information on diacritic usage for Carmen (CR), Carol (CW), and Jane (JM) are organized in the table in three groups, based on who is being compared to whom. For instance, the column headings "CR," "CW," and "Agree." represent one group wherein Carmen and Carol are being compared; the other two groups are comparisons of diacritic usage for Carmen and Jane (CR - JM), and Carol and Jane (CW - JM).

The first column in the table indicates the diacritic symbol being analyzed. The diacritic symbols are grouped into the following categories: Nasality, Lip, Stop Release, Juncture/Stress, Tongue Configuration, Tongue Position, Sound Source, and Timing/Other. Diacritic symbols not used by any of the transcriptionists were not included in the table (most of the omitted symbols are in the Juncture/Stress category).

The second column in the table indicates a particular study or transcript group. The numbers correspond to the various study/transcript groups as follows:

- 1 = Lewis
- 2 = Iowa
- 3 = Gregg's dissertation samples
- 4 = Gregg's dissertation samples
- 5 = Miscellaneous training samples
- 6 = Miscellaneous training samples

Gregg's samples were divided into two groups because Carol only transcribed three samples for reliability purposes, whereas there were five of Gregg's samples (two in addition to the three that Carol transcribed) transcribed by Carmen and Jane. The miscellaneous training samples were pulled from the Lewis and Iowa studies to give Jane some preliminary transcription practice before tackling Gregg's samples. The training samples were divided into two groups because some were originally transcribed by Carmen only, and some by Carol only.

The column "#T" represents how many transcripts in each study/group are being compared (i.e., were pulled out for reliability purposes). The column "#U" is the number of utterances contained in those transcripts. For instance, in the Lewis study (Study 1), there were ten transcripts compared, and in those ten transcripts there were a total of 364 utterances.

Under each transcriptionist's initials are three subheadings: *n* represents the number of times the diacritic in question was used by the transcriptionist in all of the samples compared; %T is not actually a percentage, but a calculation of the number of times the diacritic was used *per transcript* ($n/\#T$); and %U is, again, not a percentage, but a calculation of the number of times the diacritic was used *per utterance* ($n/\#U$). The per transcript and per utterance calculations provide a means of comparing diacritic usage across studies that have different numbers of transcripts and utterances, and they make it easier to spot usage trends among the three transcriptionists.

NOTE: *n* for each transcriptionist can include several situations. It includes any instances of agreement with the other transcriptionist as well as the number of disagreements. Disagreements include cases where one transcriptionist used the symbol when the other transcriptionist used nothing, or one transcriptionist used the symbol while the other transcriptionist used a different symbol. Most cases of disagreement are simply use/non-use differences, but as we go through this table in greater detail, instances of significant disagreement due to the use of *different* symbols on a given phoneme(s) will be discussed.

The "Agree." heading, which represents agreement for the diacritic, has two subheadings: *n* and %. The *n* represents the number of tokens of agreement. For instance, if CR and CW had two instances of agreement for \sim (which they did), *n* would be 4 (i.e., two uses of the diacritic for each of two transcriptionists, or $2 \times 2 = 4$). The % subheading in this case is actually a percentage calculated as the number of tokens of agreement divided by the total number of times the diacritic was used by each transcriptionist $\times 100$. For instance, in the first row of the table, in the Lewis study transcripts, Carmen used the \sim symbol 6 times, and Carol used it 5 times. Of those 11 tokens of \sim usage, 4 were agreements (2 tokens for each transcriptionist). Therefore, the percentage of agreement is $4/11 \times 100 = 36\%$.

At the end of each diacritic section and diacritic group section are total calculations for all of the studies combined. The totals for each diacritic provide more detailed information than the totals for the diacritic group as a whole (i.e., nasality, lip, stop release, etc.), and therefore for our purposes are more useful.

The -'s in the table indicate that information is not available. For example, since Jane did not transcribe the ten Lewis transcripts (Study 1) or the nine Iowa transcripts (Study 2), no comparisons can be made between CR and JM or CW and JM for those studies. The -'s under each subheading in those two transcriptionist comparison groups for the Lewis and Iowa studies represent this.

Diacritics: Explanation and Analyses

Nasality. The totals for $\tilde{}$ (nasalized) usage indicate that Carmen uses this symbol slightly more than Carol and Jane. Agreement between Carmen and Jane is the lowest (0%); agreement between Carmen and Carol is 28%, and agreement between Carol and Jane is 30%. Disagreements were mostly use/non-use and did not seem to be attributable to any particular transcript or study/transcript group.

Carmen and Jane use about the same amount of $\tilde{}$ (nasal emission) symbols, and Carol uses considerably fewer. This is most likely due to the word "mhm," which Carmen and Jane usually transcribe as $/m\tilde{h}m/$ and Carol as $/mhm/$. Carmen and Jane's agreement is 86%, Carmen and Carol's agreement is 8%, and Carol and Jane's agreement is 0%. For this symbol, disagreement was all use/non-use. Carol could very easily boost her agreement with Carmen and Jane by changing her transcription of "mhm" by adding $\tilde{}$ to the /h/. After discussing this with Larry, it was decided to make $/m\tilde{h}m/$ a standardized transcription, since there is bound to be some nasal emission on the /h/ when producing this word.

Carmen used more $\tilde{}$ (denasalized) symbols than either Carol or Jane, and Carol used slightly more $\tilde{}$ symbols than Jane as well. The elevated number of $\tilde{}$ used in the Lewis study can be attributed to the file SPINK-C1, where a pervasive denasal resonance was noted. Carmen and Carol's agreement was 46%, Carmen and Jane's agreement was 0%, and Carol and Jane's agreement was 30%. Most disagreements were use/non-use.

Overall agreement for the nasality symbols for the three transcriptionist comparison groups was fairly similar: CR - CW agreement was 33%, CR - JM agreement was 32%, and CW - JM agreement was 29%. However, as indicated previously, this similarity of percentage agreement at the nasality group level does not illuminate the differences noted at the individual diacritic level.

Lip. Carol is the only transcriptionist who used the ^{ɔ̹} (rounded vowel) symbol; Carmen and Jane did not use it at all. Therefore, agreement was 0% for CR - CW and CW - JM comparisons, and the CR - JM comparison agreement could not be calculated since ^{ɔ̹} was not used by either transcriptionist. Nearly all disagreements were use/non-use.

Usage and percentage agreements for ^w (labialized consonant) were similar across all three transcriptionist comparison groups. Carol used slightly more ^w symbols than Carmen; their agreement was 38%. Jane used a few more ^w symbols than Carmen; their agreement was 40%. Jane used one more ^w symbol than Carol; their agreement was 44%. Almost all disagreements were use/non-use.

Carmen and Carol used the same number of ^m (nonlabialized consonant) symbols in the transcripts compared, and Jane used more than Carmen and Carol. Agreement was 0% for all three transcriptionist comparison groups. All disagreements were use/non-use.

Overall agreement for the lip symbols was low: CR - CW agreement was 18%, CR - JM agreement was 25%, and CW - JM agreement was 21%. Carol used more lip symbols than Carmen and Jane. Use of lip symbols is quite low compared to the other diacritic groups (except Juncture/Stress). Low agreement and usage of lip symbols is most likely due to the absence of a visual component in our transcription work (a discussion of this issue can be found in Chapter 8 of *Clinical Phonetics*). Probably a good rule of thumb for making a decision to use a lip symbol is, "When in doubt, leave it out."

Stop Release. In the transcripts compared, Carmen used more ^h (aspirated) symbols than Carol, but Carmen used the same number as Jane. Carol and Jane used the same number of ^h symbols as well. Both usage and percentage figures were low (0% agreement for all three transcriptionist comparison groups). Most disagreements were due to use/non-use.

Use of the [̄] (unaspirated) symbol was distributed fairly equally in the CR - CW and CR - JM comparison groups. However, Jane used almost twice as many [̄] symbols as Carol in the CW - JM group. Of the three stop release symbols, this symbol was used the most, and it also had the highest agreement percentages. CR - CW agreement was 31%, CR - JM agreement was 58%, and CW - JM agreement was 55%. Most disagreements were use/non-use.

Agreement on [̇] (unreleased) was 0% for all three transcriptionist comparison groups. Carmen used slightly more [̇] symbols than Carol and Jane, and Carol used a few more symbols than Jane (Jane did not use any [̇] symbols in the CW - JM comparison group). All disagreements were use/non-use.

Overall agreement for the stop release symbols was lowest (9%) for CR - CW, 42% for CR - JM, and 40% for CW - JM. The 0% agreements for ^h and [̇] bring down the overall percentage agreements. It is probably best to use stop release symbols only in situations where the release (or unrelease) of a stop is atypical for the context in which it occurs, or in situations in which the release (or unrelease) is perceived as being exaggerated.

Juncture/Stress. These symbols are not used often by any of the transcriptionists; the + (open juncture) symbol is the only one included in the table. In the CR - CW comparison group, agreement was 40%. In the CR - JM comparison group, agreement was 0% (Jane did not use the symbol at all). Due to low usage, this diacritic category has very little impact on the overall narrow agreement figures.

Tongue Configuration. The _̣ (dentalized) symbol is one of the most frequently used diacritic symbols. Carmen and Carol used about the same number of _̣ symbols; their agreement was 53%. Jane used nearly twice as many _̣ symbols as Carmen and Carol; agreement was low at 15% and 22%, respectively. It is clear that Jane is using too many _̣ symbols in general in situations where it is not warranted. Jane has spent some time listening to "true" dental /s/'s to "fine-tune" her perception, and is now confident that her reliability in transcribing dentalization will be greater in future comparisons (the training tape includes several speech samples containing dentalization of fricatives).

Carol used many more (58 vs. 11) ɹ̥ (palatalized) symbols than Carmen, and CR - CW agreement was quite low at 14%. Carmen and Jane used ɹ̥ with about the same frequency (not often), and agreement was 0%. Carol used a few more ɹ̥ symbols than Jane, and agreement was 50%. In the Lewis study in particular, Carol used ɹ̥ 17 times when Carmen used the ɹ̥ symbol, all for the transcript TSCHI-C1. Refer to Sample 1 on the training tape for a portion of this particular speech sample, and note in the corresponding key Larry's judgment of the sounds of interest. This hopefully can help us to clarify the perceptual features of palatalized (and rhotacized) fricatives. We also want to make sure that disagreements on ɹ̥ and ɹ̥ are not due to any clerical errors, since it's easy to mix up the two, forgetting which direction the "tail" should go.

Carmen used more ɹ̥ (lateralized) symbols than Carol; CR - CW agreement was 44%. The ɹ̥ symbol was used only once by Carmen in the CR - JM comparison group; agreement was 0%. Carol and Jane each used the ɹ̥ symbol once, and agreement was 100%. All disagreements were use/non-use.

Carmen used about 3 times as many ɹ̥ (rhotacized) symbols as Carol, all in the Lewis group. Agreement was 46%. The ɹ̥ symbol was not used in the CR - JM or the CW - JM comparison groups. Refer to the previous paragraph on the ɹ̥ symbol (palatalized) for additional discussion of this symbol.

Carmen and Carol used the ɹ̥ (velarized) symbol with about the same frequency in comparison to each other, and they both use it more often than Jane. Agreement for all three transcriptionist comparison groups was low: CR - CW agreement was 11%, CR - JM agreement was 0%, and CW - JM agreement was 0%. Most disagreements were use/non-use. It appears that Jane is not using this symbol in enough instances where it should be used. Jane admits that it was perceptually difficult for her to distinguish a velarized /l/ on the *Clinical Phonetics* training examples. All transcriptionists could probably benefit by referring back to "Module 3: Velarized /l/" on *Clinical Phonetics* training tape #3B.

Of all the diacritic symbols, ɹ̥ (derhotacized) was the second most frequently used. Carmen used more ɹ̥ symbols than Carol, Jane used more ɹ̥ symbols than Carmen, and Carol used slightly more ɹ̥ symbols than Jane, making ɹ̥ usage variable and possibly transcript-dependent. Agreements for each transcriptionist comparison group were as follows: CR - CW agreement was 47%, CR - JM

agreement was 34%, and CW - JM agreement was 33%. Almost all agreements were use/non-use. Improving agreement for this symbol is important for several reasons. First, because ʊ has such a high usage rate, it has a higher impact on overall narrow agreement than many of the other symbols. Also, because derhotacization is considered a speech-sound error under PEPPER guidelines, misuse of the symbol can affect, either positively or negatively, a speaker's distortion count and SDCS classification.

Overall agreement for the tongue configuration symbols was not great: CR - CW agreement was 44%, CR - JM agreement was 23%, and CW - JM agreement was 27%. Again, it is more useful to look at the diacritics individually, but because tongue configuration diacritic usage is high and includes some important symbols (ɹ̥, ɹ̥̄), this group as a whole is deserving of attention in efforts to improve reliability/agreement.

Tongue Position. Of the three transcriptionists, Carmen uses the most ɹ̥ (centralized) symbols--considerably more than Carol or Jane. Agreement for this symbol was low: CR - CW agreement was 4%, CR - JM agreement was 13%, and CW - JM agreement was 0%. Most disagreements were use/non-use. It appears that in order to increase agreement on this symbol, Carmen needs to use it less often, and/or Carol and Jane need to use it more often.

The ɹ̥̄ (retracted tongue body) symbol was not used a lot. In the three transcriptionist comparison groups, Carmen used over twice as many ɹ̥̄ symbols as Carol, Carmen used the same number as Jane, and Carol used more than Jane. No clear usage trend was evident. Agreement on this symbol was low: CR - CW and CW - JM agreements were 0%, and CR - JM agreement was 33%. Most disagreements were use/non-use.

The ɹ̥̄̄ (advanced tongue body) symbol was seldom used by the transcriptionists. Agreement percentages were 67% for CR - CW, 0% for CR - JM, and 0% for CW - JM. Because of its low rate of use, this symbol does not greatly affect overall narrow agreement.

In the three transcriptionist comparison groups, Carol used more ɹ̥̄̄̄ (raised tongue body) symbols than Carmen, and Jane used more ɹ̥̄̄̄ symbols than Carol or Carmen. Agreement for all three transcriptionist groups was 0%.

Use of the $\underset{\cdot}{r}$ (lowered tongue body) symbol was about equal for Carol and Jane; Carmen used more $\underset{\cdot}{r}$ symbols in the CR - CW comparison group and no $\underset{\cdot}{r}$ symbols in the CR - JM comparison group. Agreement was 0% for all three transcriptionist pairs. Nearly all disagreements were use/non-use.

Carmen used more $\underset{\cdot}{\epsilon}$ (fronted) symbols than Carol or Jane. Agreement was 0% for the CR - CW and CR - JM groups. The $\underset{\cdot}{\epsilon}$ diacritic was not used in the CW - JM group, so agreement was not calculated. All disagreements were use/non-use.

Carmen used twice as many $\underset{\cdot}{\backslash}$ (backed) symbols as Carol, and both Carmen and Carol used more $\underset{\cdot}{\backslash}$ symbols than Jane, who did not use any in all of the transcripts compared. Agreement for this symbol was low: 22% for CR - CW, and 0% for CR - JM and CW - JM. All disagreements were use/non-use.

Overall agreement for the tongue position symbols was very low: CR - CW agreement was 3%, CR - JM agreement was 11%, and CW - JM agreement was 0%. Carmen used more tongue position symbols than Carol or Jane; overall usage for Carol and Jane was similar. Since most of these diacritics are used to modify vowels, it is not surprising that agreement was low.

Sound Source. The $\underset{\cdot}{v}$ (partially voiced) symbol was seldom used, but used most by Carol in relation to Carmen and Jane. Agreement was 0% for all three transcriptionist comparison groups.

Carol used slightly more $\underset{\cdot}{\circ}$ (partially devoiced) symbols in comparison to Carmen, and Jane used many more $\underset{\cdot}{\circ}$ symbols than Carmen. In the CW - JM comparison group, Carol used it two times, and Jane used it once. Agreement was low for this symbol: CR - CW agreement was 11%, and CR - JM and CW - JM agreement was 0%. Jane may need to shift her perception of "devoiced" slightly; it's possible that she is using the $\underset{\cdot}{\circ}$ symbol for sounds that aren't devoiced to a degree unusual enough to warrant use of the diacritic.

Jane used considerably more $\underset{\cdot}{\text{.}}$ (glottalized) symbols than Carmen and Carol; usage for Carmen and Carol was similar. Agreement for $\underset{\cdot}{\text{.}}$ is higher than for most of the other sound source symbols, probably because it was used more than the others. Agreement between CR - CW was 37%, agreement between CR - JM was 7%, and agreement between CW - JM was 41%. Jane should

probably use this symbol less often in order to improve agreement, similar to the $\underset{\circ}{\text{u}}$ symbol discussed in the previous paragraph.

The $\underset{\cdot}{\text{u}}$ (breathy) symbol is not used a lot, and use is similar among the transcriptionists. Agreement varied among the three transcriptionist comparison groups: CR - CW agreement was 13%, CR - JM agreement was 44%, and CW - JM agreement was 67%. Most disagreements were use/non-use.

Both Carol and Jane used more $\underset{\times}{\text{u}}$ (frictionalized) symbols than Carmen, and Jane used a few more $\underset{\times}{\text{u}}$ symbols than Carol in the five CW - JM transcripts compared. Agreement on this symbol was low: CR - CW agreement was 17%, CR - JM agreement was 12%, and CW - JM agreement was 20%.

The $\underset{\wedge}{\text{u}}$ (whistled) symbol was not used often. Usage for Carmen and Carol was about the same; their agreement was 0%. This symbol was used only in the Lewis transcripts (many on the TSCHI-C1 transcript in particular). The $\underset{\wedge}{\text{u}}$ symbol was not used at all in the CR - JM and CW - JM groups. Listening to the TSCHI-C1 sample and clarifying the perception of $\underset{\wedge}{\text{u}}$ may help to improve agreement, but since $\underset{\wedge}{\text{u}}$ usage is so low, improving agreement on this diacritic alone won't have much impact on overall narrow agreement.

Carmen used slightly more $\underset{\vee}{\text{u}}$ (weak) symbols than Carol, and Jane used more $\underset{\vee}{\text{u}}$ symbols than both Carmen and Carol (who did not use any in the CR - JM and CW - JM comparison groups, respectively). Agreement on this symbol was low: 5% for CR - CW, and 0% for CR - JM and CW - JM. Jane may be using more $\underset{\vee}{\text{u}}$ symbols for sounds that are difficult to hear, but perhaps they are not produced weakly. Clarification of when this symbol should be used might reduce Jane's use of $\underset{\vee}{\text{u}}$, thereby improving agreement.

Overall agreement for sound source symbols is pretty low: 20% for CR - CW, 10% for CR - JM, and 32% for CW - JM. Jane used quite a few more sound source symbols than Carmen and Carol (especially $\underset{\cdot}{\text{u}}$, $\underset{\times}{\text{u}}$, and $\underset{\vee}{\text{u}}$). Ways to improve agreement on the sound source symbols should be explored.

Timing/Other. Carmen used over twice as many : (lengthened) symbols as Carol, and Jane used more : symbols than Carmen and Carol. Agreement could be improved: CR - CW agreement was 33%, CR - JM agreement was 36%, and CW - JM agreement was 17%. Since "lengthened" sounds are not considered errors in PEPPER, low agreement does not make a difference in a speaker's speech profile, but it does have some effect on overall narrow agreement.

The > (shortened) symbol was not used much. Carol and Jane used slightly more > symbols than Carmen. Agreement was low: CR - CW agreement was 8%, CR - JM agreement was 0% (Carmen did not use the symbol at all), and CW - JM agreement was 18%. Again, although a shortened sound is not considered a speech-sound error, it would be nice to have better agreement on this diacritic.

The ̣ (syllabic consonant) symbol is used more often than any of the other timing/other symbols, and it has the highest agreement. All three transcriptionists used it with about the same frequency. Agreement for CR - CW was 90%, for CR - JM it was 96%, and for CW - JM it was 86%. This particular diacritic can be quite a boost to overall narrow agreement figures.

The ̂ (synchronic tie) symbol has low usage and agreement, probably because an on- or offglide symbol is more often the diacritic of choice. Carol used slightly more ̂ symbols than Carmen and Jane, and Jane used more than Carmen. CR - CW agreement was 14%, CR - JM agreement was 22%, and CW - JM agreement was 22%.

Carol and Jane used onglides with similar frequency, but they used considerably more onglides than Carmen. Agreement was quite similar among the three transcriptionist comparison groups: 33% for CR - CW, 34% for CR - JM, and 40% for CW - JM. One way to improve agreement might be for Carmen to listen more closely for onglides.

Carol used many more offglides than Carmen and Jane. Agreement on offglides was quite low: 22% for CR - CW, 11% for CR - JM, and 10% for CW - JM. Decreasing Carol's use of offglides could improve agreement, and it could also affect a speaker's speech profile in instances where Carol codes an offglide as a speech error.

The overall agreement for the timing/other group is the highest of all the diacritic groups. CR - CW agreement is 50%, CR - JM agreement is 55%, and CW - JM agreement is 31%. These agreement figures are elevated by the high agreement and use of the syllabic consonant diacritic, so the overall agreement percentages are a bit misleading. The reliability for the other diacritic symbols in this group needs improvement.

Summary

There are several ways to look at the diacritic reliability data when determining the diacritics most in need of improvement (in terms of agreement). One way to prioritize these diacritics is first to concentrate on those symbols with the highest usage rates and the lowest percentage agreements. This would (hopefully) result in the most drastic improvement in overall narrow agreement figures. Or, we could begin by first working on improving agreement for the diacritics that would be considered speech-sound distortion errors (i.e., in PEPPER, transcribed on the Z-line only). This would not only improve overall narrow agreement figures (though perhaps more modestly) but would also give us more confidence in the speech profiles generated in PEPPER as a result of our transcriptions.

Each transcriptionist needs to honestly evaluate her strengths and weaknesses in broad and narrow transcription, and then review or modify her transcription as appropriate. This analysis and discussion have hopefully taken us further down the road to improved transcription and greater interjudge reliability.

	Carmen	Carol	Jane
Uses more than the other two transcriptionists	[[~]] [_˘] [_˙] [_˚]	[[˚]] [_˙] [offglides]	[^m] [[˚]] [_˘] [_˙] [] [:]
Uses fewer than the other two transcriptionists	[[˘]] [onglides]	[[˚]]	[[˚]] [+] [_˚]

By using the first method of prioritization (usage rates), the following list, in order of most important to least important, would result:

- | | | | | | |
|-----|-------------------|--------------------|-----|------------------|-----------------------|
| 1. | [_̣] | dentalized | 19. | [̃] | nasal emission |
| 2. | [_̤] | derhotacized | 20. | [_̥] | raised tongue body |
| 3. | | onglide | 21. | [[̦]] | unreleased |
| 4. | [: _̄] | lengthened | 22. | [_̨] | breathy |
| 5. | [_̤] | glottalized | 23. | [_̣] | lateralized |
| 6. | | offglide | 24. | [^m] | nonlabialized |
| 7. | [̃] | denasalized | 25. | [_̣] | retracted tongue body |
| 8. | [̃] | nasalized | 26. | [^h] | aspirated |
| 9. | [[̥]] | unaspirated | 27. | [_̣] | retroflexed |
| 10. | [̂] | synchronic tie | 28. | [_̣] | lowered tongue body |
| 11. | [_x] | frictionalized | 29. | [^ɔ] | rounded |
| 12. | [_j] | palatalized | 30. | [_̟] | backed |
| 13. | [_̣] | centralized | 31. | [_̣] | advanced tongue body |
| 14. | [_̤] | weak | 32. | [₊] | open juncture |
| 15. | [^w] | labialized | 33. | [_̣] | whistled |
| 16. | [_̟] | shortened | 34. | [_̣] | fronted |
| 17. | [_̤] | partially devoiced | 35. | [_̤] | partially voiced |
| 18. | [_̣] | velarized | 36. | [_̣] | syllabic consonant |

By using the second method of prioritization (speech errors vs. non-errors), the following list would be generated (again, from most important to least important):

- | | | | | | |
|-----|------------------|-----------------------|-----|-------------------|----------------------|
| 1. | [_n] | dentalized | 19. | [^h] | aspirated |
| 2. | [_u] | derhotacized | 20. | [_r] | lowered tongue body |
| 3. | [[^]] | synchronic tie | 21. | [_ɔ] | backed |
| 4. | [_l] | lateralized | 22. | [_ɹ] | advanced tongue body |
| 5. | [^m] | nonlabialized | 23. | [_ɛ] | fronted |
| 6. | [_ɹ] | retroflexed | 24. | [:] | lengthened |
| 7. | [^o] | rounded | 25. | [_·] | glottalized |
| 8. | onglide | | 26. | [[~]] | denasalized |
| 9. | offglide | | 27. | [[~]] | nasalized |
| 10. | [⁼] | unaspirated | 28. | [_x] | frictionalized |
| 11. | [_j] | palatalized | 29. | [_v] | weak |
| 12. | [_ɹ] | centralized | 30. | [>] | shortened |
| 13. | [^w] | labialized | 31. | [^ʔ] | unreleased |
| 14. | [_o] | devoiced | 32. | [_{..}] | breathy |
| 15. | [[~]] | velarized | 33. | [+] | open juncture |
| 16. | [[~]] | nasal emission | 34. | [_ʌ] | whistled |
| 17. | [_ɹ] | raised tongue body | 35. | [_v] | partially voiced |
| 18. | [_ɹ] | retracted tongue body | 36. | [_ɹ] | syllabic consonant |

Diacritic Agreement

Symbol	Study	#T	#U	CR			CW			Agree.		CR			JM			Agree.		CW			JM			Agree.	
				n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%
Nasality																											
~	1	10	364	6	.6	.01	5	.5	.01	4	36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	22	2.4	.03	16	1.8	.05	6	16	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	4	1.3	.004	5	1.7	.03	6	67	--	--	--	--	--	--	--	5	1.67	.03	3	1.0	.02	0	0.0	
	4	5	336	--	--	--	--	--	--	--	--	6	1.2	.018	4	0.8	.012	0	0.0	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	3	.60	.009	0	0.0	.000	0	0.0	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	5	2.5	.081	7	3.5	.12	6	50	
	Totals				32	1.5	.04	26	1.2	.03	16	28	9	1.13	.021	4	.50	.009	0	0.0	10	2.0	.04	10	2.0	.04	6
~	1	10	364	8	.8	.02	0	0.0	.00	0	0.0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	14	1.6	.04	1	.1	.00	0	0.0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	1	.33	.01	1	.33	.01	2	100	--	--	--	--	--	--	--	1	.33	.01	0	0.0	.00	0	0.0	
	4	5	336	--	--	--	--	--	--	--	--	3	.6	.01	2	.4	.01	4	80	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	4	1.3	.04	5	1.7	.05	8	89	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	0.0	
	Totals				23	1.1	.03	2	.1	.00	2	8.0	7	.88	.02	7	.88	.02	12	86	1	.2	.00	0	0.0	.00	0
~	1	10	364	42	4.2	.12	20	2.0	.05	28	45	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	2	.22	.01	4	.44	.01	2	33	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	1	.33	.01	1	.33	.01	2	100	--	--	--	--	--	--	--	1	.33	.01	3	1.0	.02	0	0.0	
	4	5	336	--	--	--	--	--	--	--	--	6	1.2	.02	4	.8	.01	0	0.0	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	0	.0	.00	1	.33	.01	0	0.0	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	14	7	.24	9	4.5	.15	8	35	
	Totals				45	2.0	.05	25	1.1	.03	32	46	6	.8	.01	5	.6	.01	0	0.0	15	3	.07	12	2.4	.05	8
<u>Nasality Totals</u>				100	4.5	.11	53	2.4	.06	50	33	22	2.8	.05	16	2	.04	12	32	26	5.2	.11	22	4.4	.10	14	29
Lip																											
o	1	10	364	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	0	0.0	.00	1	.11	.00	0	0.0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	0	0.0	.00	4	1.3	.02	0	0.0	--	--	--	--	--	--	--	4	1.3	.02	0	0.0	.00	0	0.0	
	4	5	336	--	--	--	--	--	--	--	--	0	.0	.00	0	.0	.00	0	--	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	0	.0	.00	0	.0	.00	0	--	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	2	1.0	.03	0	0.0	.00	0	0.0	

Symbol	Study	#T	#U	CR			CW			Agree.		CR			JM			Agree.		CW			JM			Agree.	
				n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%
e	Totals			0	0.0	.00	5	.23	.01	0	0.0	0	.0	.00	0	.0	.00	0	n/a	6	1.2	.03	0	0.0	.00	0	0.0
	1	10	364	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	2	9	343	3	.33	.01	5	.56	.01	0	0.0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	3	3	168	4	1.3	.02	4	1.3	.02	6	75	--	--	--	--	--	--	--	--	4	1.3	.02	5	1.7	.03	4	44
	4	5	336	--	--	--	--	--	--	--	--	4	.8	.01	6	1.2	.02	4	40	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	0	.0	.00	0	.0	.00	0	--	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--
	Totals			7	.32	.01	9	.41	.01	6	38	4	.5	.01	6	.8	.01	4	40	4	.8	.02	5	1.0	.02	4	44
m	1	10	364	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	2	9	343	5	.56	.01	5	.56	.01	0	0.0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	3	3	168	1	.33	.01	1	.33	.01	0	0.0	--	--	--	--	--	--	--	--	1	.33	.01	3	1.0	.02	0	0.0
	4	5	336	--	--	--	--	--	--	--	--	1	.2	.00	4	.8	.01	0	0.0	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	0	.0	.00	1	.33	.01	0	0.0	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--
	Totals			6	.27	.01	6	.27	.01	0	0.0	1	.13	.00	5	.63	.01	0	0.0	1	.2	.00	3	.6	.01	0	0.0
	<u>Lip Totals</u>			13	.59	.01	20	.91	.02	6	18	5	.63	.01	11	1.4	.03	4	25	11	2.2	.05	8	1.6	.04	4	21
Stop Release																											
h	1	10	364	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	2	9	343	5	.56	.01	1	.11	.00	0	0.0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	3	3	168	2	.67	.01	0	0.0	.00	0	0.0	--	--	--	--	--	--	--	--	0	0.0	.00	1	.33	.01	0	0.0
	4	5	336	--	--	--	--	--	--	--	--	2	.4	.01	4	.8	.01	0	0	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	2	.67	.02	0	0.0	.00	0	0	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	1	.5	.02	0	0.0	.00	0	0.0
	Totals			7	.3	.01	1	.05	.00	0	0.0	4	.5	.01	4	.5	.01	0	0	1	.2	.00	1	.2	.00	0	0
	"	1	10	364	1	.1	.00	3	.3	.01	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
2		9	343	3	.33	.01	2	.2	.01	2	40	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
3		3	168	2	.67	.01	2	.67	.01	2	50	--	--	--	--	--	--	--	--	2	.67	.01	5	1.7	.03	2	29
4		5	336	--	--	--	--	--	--	--	--	23	4.6	.07	22	4.4	.07	26	58	--	--	--	--	--	--	--	--
5		3	93	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	
6		2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	2	1.0	.03	2	1.0	.03	4	100

Symbol	Study	#T	#U	CR			CW			Agree.		CR			JM			Agree.		CW			JM			Agree.	
				n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%
r	Totals			6	.27	.01	7	.32	.01	4	31	23	2.9	.05	22	2.8	.05	26	58	4	.8	.02	7	1.4	.03	6	55
	1	10	364	2	.2	.01	7	.7	.02	0	0.0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	12	1.3	.03	2	.22	.01	0	0.0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	0	0.0	.00	2	.67	.01	0	0.0	--	--	--	--	--	--	--	--	2	.67	.01	0	0.0	.00	0	0
	4	5	336	--	--	--	--	--	--	--	--	3	.6	.01	2	.4	.01	0	0	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	2	.67	.02	2	.67	.02	0	0	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0	0	0	0	0	0	0	--
	Totals			14	.64	.02	11	.5	.01	0	0	5	.63	.01	4	.5	.01	0	0	2	.4	.01	0	0.0	.00	0	0
<u>Stop Release Totals</u>				27	1.2	.03	19	.86	.02	4	8.7	32	4	.07	30	3.8	.07	26	42	7	1.4	.03	8	1.6	.04	6	40
Juncture/Stress																											
+	1	10	364	3	.3	.01	1	.1	.00	2	50	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	0	0.0	.00	1	.33	.01	0	0	--	--	--	--	--	--	--	--	1	.33	.01	0	0.0	.00	0	0
	4	5	336	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	1	.5	.02	0	0.0	.00	0	0
<u>Juncture/Stress Totals</u>				3	.14	.00	2	.09	.00	2	40	0	0	0	0	0	0	0	n/a	2	.40	.01	0	0.0	.00	0	0
Tongue Configuration																											
n	1	10	364	5	.5	.01	8	.8	.02	6	46	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	127	14.1	.37	124	13.8	.36	136	54	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	12	4.0	.07	15	5	.09	12	44	--	--	--	--	--	--	--	--	15	5	.09	33	11	.20	8	17
	4	5	336	--	--	--	--	--	--	--	--	13	2.6	.04	35	7	.10	6	13	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	10	3.3	.11	9	3	.10	4	21	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	8	4	.14	17	8.5	.29	8	32
Totals				144	6.5	.16	147	6.7	.17	154	53	23	2.9	.05	44	5.5	.10	10	15	23	4.6	.10	50	10	.22	16	22
j	1	10	364	6	.6	.02	30	3.0	.08	8	22	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	3	.33	.01	23	2.6	.07	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	2	.67	.01	5	1.7	.03	2	29	--	--	--	--	--	--	--	--	3	1.0	.02	2	.67	.01	2	40
	4	5	336	--	--	--	--	--	--	--	--	1	.20	.00	2	.40	.01	0	0.0	--	--	--	--	--	--	--	
	5	3	93	--	--	--	--	--	--	--	--	1	.33	.01	0	.00	.00	0	0.0	--	--	--	--	--	--	--	

Symbol	Study	#T	#U	CR			CW			Agree.		CR			JM			Agree.		CW			JM			Agree.	
				n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%
α	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	2	1.0	.03	1	.50	.02	2	67		
	Totals			11	.5	.01	59	2.6	.07	10	14	2	.25	.00	2	.25	.00	0	0.0	5	1.0	.02	3	.60	.01	4	50
	1	10	364	2	.2	.01	1	.1	.00	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	29	3.2	.08	17	1.9	.05	16	35	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	1	.33	.01	1	.33	.01	2	100	--	--	--	--	--	--	--	1	.33	.01	1	.33	.01	2	100	
	4	5	336	--	--	--	--	--	--	--	--	1	.2	.00	0	.0	.00	0	0	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	0	0	0	0	0	0	0	--	--	--	--	--	--	--	--	
Totals				32	1.5	.04	19	.9	.02	18	44	1	.13	.00	0	0	0	0	0	1	.2	.00	1	.2	.00	2	100
β	1	10	364	29	2.9	.08	10	1.0	.03	18	46	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	
	4	5	336	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	
	Totals				29	1.3	.03	10	.45	.01	18	46	0	0	0	0	0	0	0	n/a	0	0	0	0	0	0	0
γ	1	10	364	1	.1	.00	1	.1	.00	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	6	.67	.02	3	.33	.01	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	2	.67	.01	6	2.0	.04	2	25	--	--	--	--	--	--	--	6	2.0	.04	0	0	0	0	0	
	4	5	336	--	--	--	--	--	--	--	--	3	.6	.01	1	.2	.00	0	0	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	0	0	0	0	0	0	0	--	--	--	--	--	--	--	--	
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	3	1.5	.05	0	0	0	0	0	
	Totals				9	.41	.01	10	.45	.01	2	11	3	.38	.01	1	.13	.00	0	0	9	1.8	.04	0	0	0	0
δ	1	10	364	56	5.6	.15	24	2.4	.07	28	35	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	53	5.9	.15	26	2.8	.08	42	53	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	19	6.3	.11	19	6.3	.11	22	58	--	--	--	--	--	--	--	20	6.7	.12	17	5.7	.10	12	32	
	4	5	336	--	--	--	--	--	--	--	--	20	4.0	.06	35	7.0	.10	14	25	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	8	2.7	.09	7	2.3	.08	10	67	--	--	--	--	--	--	--	
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	9	4.5	.15	9	4.5	.15	6	33	
	Totals				128	5.8	.15	69	3.1	.08	92	47	28	3.5	.07	42	5.3	.10	24	34	29	5.8	.13	26	5.2	.11	18

Symbol	Study	#T	#U	CR			CW			Agree.		CR			JM			Agree.		CW			JM			Agree.	
				n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%
<u>Tongue Configuration</u>				353	16.0	.40	313	14.2	.36	294	44	57	7.1	.13	89	11.1	.21	34	23	67	13.4	.30	80	16	.35	40	27
<u>Totals</u>																											
Tongue Position																											
-	1	10	364	21	2.1	.06	3	.3	.01	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--		
	2	9	343	16	1.8	.05	7	.8	.02	2	9	--	--	--	--	--	--	--	--	--	--	--	--	--	--		
	3	3	168	1	.33	.01	0	0	0	0	0	--	--	--	--	--	--	0	0	0	0	0	0	0	0	--	
	4	5	336	--	--	--	--	--	--	--	--	11	2.2	.03	4	.8	.01	2	13	--	--	--	--	--	--	--	
	5	3	93	--	--	--	--	--	--	--	--	1	.33	.01	0	0	0	0	0	--	--	--	--	--	--	--	
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	2	1.0	.03	0	0	0	0	0	
	Totals			38	1.7	.04	10	.45	.01	2	4	12	1.5	.03	4	.5	.01	2	13	2	.4	.01	0	0	0	0	0
-	1	10	364	2	.2	.01	0	0	0	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--		
	2	9	343	6	.67	.02	0	0	0	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--		
	3	3	168	0	0	0	3	1.0	.02	0	0	--	--	--	--	--	--	3	1.0	.02	1	.33	.01	0	0		
	4	5	336	--	--	--	--	--	--	--	--	3	.6	.01	3	.6	.01	2	33	--	--	--	--	--	--	--	
	5	3	93	--	--	--	--	--	--	--	--	0	0	0	0	0	0	0	--	--	--	--	--	--	--		
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0	0	0	0	0	0	0	--	
	Totals			8	.36	.01	3	.14	.00	0	0	3	.38	.01	3	.38	.01	2	33	3	.6	.01	1	.2	.00	0	0
-	1	10	364	0	0	0	0	0	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--		
	2	9	343	0	0	0	0	0	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--		
	3	3	168	1	.33	.01	2	.67	.01	2	67	--	--	--	--	--	--	2	.67	.01	1	.33	.01	0	0		
	4	5	336	--	--	--	--	--	--	--	--	1	.2	.00	1	.2	.00	0	0	--	--	--	--	--	--	--	
	5	3	93	--	--	--	--	--	--	--	--	0	0	0	0	0	0	0	--	--	--	--	--	--	--		
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0	0	0	0	0	0	0	--	
	Totals			1	.05	.00	2	.09	.00	2	67	1	.13	.00	1	.13	.00	0	0	2	.4	.01	1	.2	.00	0	0
1	1	10	364	7	.7	.02	9	.9	.02	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--		
	2	9	343	2	.22	.01	6	.67	.02	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--		
	3	3	168	0	0	0	0	0	0	0	--	--	--	--	--	--	--	0	0	0	3	1.0	.02	0	0		
	4	5	336	--	--	--	--	--	--	--	--	0	0	0	3	.6	.01	0	0	--	--	--	--	--	--	--	
	5	3	93	--	--	--	--	--	--	--	--	0	0	0	3	1.0	.03	0	0	--	--	--	--	--	--	--	
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0	0	0	2	1.0	.03	0	0	
	Totals			9	.41	.01	15	.68	.02	0	0	0	0	0	6	.75	.01	0	0	0	0	0	5	1.0	.02	0	0

Symbol	Study	#T	#U	CR			CW			Agree.		CR			JM			Agree.		CW			JM			Agree.	
				n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%
r	1	10	364	5	.5	.01	1	.1	.00	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	9	1.0	.03	3	.33	.01	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	0	0.0	.00	1	.33	.01	0	0	--	--	--	--	--	--	1	.33	.01	0	0.0	.00	0	0	0	0
	4	5	336	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	0	0.0	.00	2	.67	.02	0	0	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0	0.0	.00	1	.5	.02	0	0	0	0
	Totals			14	.64	.02	5	.23	.01	0	0	0	0.0	.00	2	.25	.00	0	0	1	.2	.00	1	.2	.00	0	0
<	1	10	364	1	.1	.00	0	0.0	.00	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	5	.56	.01	1	.11	.00	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	1	.33	.01	0	0.0	.00	0	0	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	
	4	5	336	--	--	--	--	--	--	--	--	1	.2	.00	0	0.0	.00	0	0	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	
	Totals			7	.32	.01	1	.05	.00	0	0	1	.13	.00	0	0.0	.00	0	0	0	0	0	0	0	0	0	--
>	1	10	364	1	.1	.00	0	0.0	.00	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	3	.33	.01	2	.22	.01	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	2	.67	.01	1	.33	.01	2	67	--	--	--	--	--	--	1	.33	.01	0	0.0	.00	0	0	0	
	4	5	336	--	--	--	--	--	--	--	--	2	.4	.01	0	0.0	.00	0	0	--	--	--	--	--	--	--	--
	5	3	93	--	--	--	--	--	--	--	--	1	.33	.01	0	0.0	.00	0	0	--	--	--	--	--	--	--	--
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	
	Totals			6	.27	.01	3	.14	.00	2	22	3	.38	.01	0	0.0	.00	0	0	1	.2	.00	0	0.0	.00	0	0
<u>Tongue Position Totals</u>				83	3.8	.09	39	1.8	.04	4	3	20	2.5	.05	16	2.0	.04	4	11	9	1.8	.04	8	1.6	.04	0	0
Sound Source																											
v	1	10	364	1	.1	.00	1	.0	.00	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	0	0.0	.00	3	.33	.01	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	
	4	5	336	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	
	5	3	93	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	1	.5	.02	0	0.0	.00	0	0	0	
	Totals			1	.05	.00	4	.18	.00	0	0	0	0.0	.00	0	0.0	.00	0	--	1	.2	.00	0	0.0	.00	0	0

Symbol	Study	#T	#U	CR			CW			Agree.		CR			JM			Agree.		CW			JM			Agree.	
				n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%	n	%T	%U	n	%T	%U	n	%
)	3	3	168	5	1.7	.03	5	1.7	.03	10	100	--	--	--	--	--	--	5	1.7	.03	5	1.7	.03	10	100		
	4	5	336	--	--	--	--	--	--	--	--	18	3.6	.05	18	3.6	.05	36	100	--	--	--	--	--	--	--	
	5	3	93	--	--	--	--	--	--	--	--	21	7.0	.23	24	8.0	.26	42	93	--	--	--	--	--	--	--	
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	2	1.0	.03	2	1.0	.03	2	50		
	Totals			99	4.5	.11	101	4.6	.12	180	90	39	4.9	.09	42	5.3	.10	78	96	7	1.4	.03	7	1.4	.03	12	86
	1	10	364	1	.1	.00	2	.2	.01	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	1	.11	.00	2	.22	.01	0	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	2	.67	.01	6	2.0	.04	2	25	--	--	--	--	--	--	--	6	2.0	.04	7	2.3	.04	2	15	
	4	5	336	--	--	--	--	--	--	--	--	2	.4	.01	7	1.4	.02	2	22	--	--	--	--	--	--	--	
	5	3	93	--	--	--	--	--	--	--	--	0	0.0	.00	0	0.0	.00	0	--	--	--	--	--	--	--	--	
6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	4	2.0	.07	1	.5	.02	2	40		
Totals			4	.18	.00	10	.45	.01	2	14	2	.25	.00	7	.88	.02	2	22	10	2.0	.04	8	1.6	.04	4	22	
onglide	1	10	364	5	.5	.01	13	1.3	.04	4	22	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	8	.89	.02	33	3.7	.10	6	15	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	14	4.7	.08	29	9.7	.17	24	56	--	--	--	--	--	--	--	24	8.0	.14	20	6.7	.12	20	45	
	4	5	336	--	--	--	--	--	--	--	--	14	2.8	.04	31	6.2	.09	16	36	--	--	--	--	--	--	--	
	5	3	93	--	--	--	--	--	--	--	--	0	0.0	.00	2	.67	.02	0	0	--	--	--	--	--	--	--	
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	1	.5	.02	5	2.5	.08	0	0	
	Totals			27	1.2	.03	75	3.4	.09	34	33	14	1.8	.03	33	4.1	.08	16	34	25	5.0	.11	25	5.0	.11	20	40
offglide	1	10	364	4	.4	.01	26	2.6	.07	4	13	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	2	9	343	14	1.6	.04	30	3.3	.09	6	14	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	3	3	168	6	2.0	.04	13	4.3	.08	10	53	--	--	--	--	--	--	--	11	3.7	.07	2	.67	.01	2	15	
	4	5	336	--	--	--	--	--	--	--	--	10	2.0	.03	5	1.0	.01	2	13	--	--	--	--	--	--	--	
	5	3	93	--	--	--	--	--	--	--	--	2	.67	.02	2	.67	.02	0	0	--	--	--	--	--	--	--	
	6	2	59	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	7	3.5	.12	1	.5	.02	0	0	
	Totals			24	1.1	.03	69	3.1	.08	20	22	12	1.5	.03	7	.88	.02	2	11	18	3.6	.08	3	.6	.01	2	10
<u>Timing/Other Totals</u>				240	10.9	.27	304	13.8	.35	274	50.4	86	10.8	.20	127	15.9	.30	118	55	80	16	.35	69	13.8	.30	46	31

PVSP Agreement

How to Read the Tables

There are three tables containing PVSP information, one for each of the transcriptionist groups CR - CW (Carmen and Carol), CR - JM (Carmen and Jane), and CW - JM (Carol and Jane). In each group, the transcriptionist/coder listed first in the pair is considered the "standard" against which the other coder's decisions are compared.

Each table is broken down into the categories of Phrasing, Rate, Stress, Loudness, Pitch, Laryngeal Features, and Resonance. For each category, the appropriate (always "1") and inappropriate codes are listed right below the category heading from left to right. For instance, under the "RATE" category heading is listed a "1" for Appropriate, "9" for Slow Articulation/Pause Time, "10" for Slow/Pause Time, "11" for Fast, and "12" for Fast/Acceleration.

The Study column indicates which studies the transcripts/PVSPs being compared came from. In all cases (CR - CW, CR - JM, and CW - JM) transcripts/PVSPs were taken from the Lewis, Iowa1, and Pred2 studies. The "#T" column represents the number of transcripts/PVSPs used for comparison from each study. [NOTE: In the CR - JM and CW - JM comparison groups there are only two transcripts/PVSPs each from the Lewis and Iowa1 studies, so totals and percentages derived from sample numbers that small need to be interpreted with caution. We can't really make generalizations or draw conclusions based on the small individual study sample numbers, so it's best to talk about overall totals and percentages from the three study groups combined.]

Under each appropriate and inappropriate code in the PVSP category sections are listed four columns labeled A, B, C, and D. "A" represents the number of exact agreements, i.e., the number of times the two coders used that same code (whether appropriate or inappropriate) on the same utterance. Included in the exact agreements on appropriate codes (code "1" subheading "A") are instances where both coders used "1" to indicate appropriate and instances where one coder used a "1" to indicate appropriate, and the other coder used an inappropriate, circled code, since arbitrarily we've decided to consider a circled code "appropriate" for purposes of comparison. However, instances where both coders used the same inappropriate code and either one or both of them circled the code would also be considered agreements, but under the "A" column of the inappropriate code

used by the coders. This is because we think that both coders heard something they considered to be inappropriate, even though it may have been a borderline case. So, the circled codes can be recorded in the totals under an appropriate or an inappropriate column, depending on the code used by the other coder in the comparison group for the particular utterance in question.

Column "B" represents within-code agreements on inappropriate codes. For instance, if Carmen and Carol both considered an utterance to be inappropriate in the Phrasing category, but Carmen coded it "4" and Carol coded it "3", it would be counted as an within-code agreement but not an exact agreement. The number of instances of within-code agreement are found under the "B" column of the inappropriate code used by the first, "standard" coder. Therefore, in the example above, the "4" vs. "3" instance would be recorded under the Code 4 heading, column B since Carmen used the "4" inappropriate code, and she is considered the "standard" in the CR - CW comparison group.

Column "C" represents the number of times the "standard" coder used an inappropriate code when the other coder judged an utterance to be appropriate. For instance, in the CR - JM group, if Carmen coded a particular utterance as "24" (rough), and Jane coded it "1" (appropriate) for Laryngeal Features, this instance of disagreement would be recorded under heading "24", column C in the Laryngeal section of the CR - JM table.

Column "D" on the other hand represents disagreements where the standard coder judged the utterance to be appropriate and the second coder used an inappropriate code. For instance, in the CW - JM comparison group, if Carol coded an utterance as "1" (appropriate) and Jane coded it as "12" (fast/acceleration) in the rate category, this instance would be coded as a disagreement under Heading "12" column D, in the Rate section of the CW - JM table. Therefore, since column C represents the disagreements when the "standard" coder considers the utterance to be inappropriate (when the second coder calls it appropriate), and column D represents the disagreements when the second coder considers the utterance inappropriate when the standard coder calls it appropriate, comparing the totals under the C and D columns gives us some information on who might be using an inappropriate code too much or too little, thereby bringing down the percentage agreement for a particular category (especially where inappropriate codes are concerned).

Under some inappropriate code headings there are " - "'s under the A, B, C, and D columns. This means that the code in question was not used at all by either coder.

For each PVSP section (Phrasing, Rate, Stress, etc.), the A, B, C, and D columns are totaled, and these totals are recorded in the "All" row labeled at the bottom of the "Study" column heading. Percentage agreements are then provided for each individual study, as well as for the studies combined ("All").

Remember that the characteristics (such as subjects' ages and SDCS classifications) of each study group may influence the difficulty or ease with which the PVSP is coded and the number of appropriate and inappropriate codes likely to be used. For instance, since some adults were included in the Lewis study, we would expect the "Exact Agreement With Appropriate" agreements to be higher than for other studies where the subjects were younger and had "lower" SDCS classifications. This is because we would expect adult samples to contain mostly appropriate utterances, and fewer instances of coder use (and therefore disagreement) on inappropriate codes.

The "Exact Agreement with Appropriate" line at the end of each section displays the percent agreement of all appropriate and inappropriate codes used for each group (Iowa1, Lewis, Pred2, All). This includes exact agreements only (refer back to paragraph 4 of this section for explanation on how circled codes are handled). For each group, the percent agreement is calculated by adding up the numbers in the A columns for the particular study of interest and dividing that number by the sum of the numbers in the A, B, C, and D columns for that same study, multiplied by 100. For instance, if we look at the CR - JM table under the Rate category, we see that for the Lewis study there were 43 agreements on the "1" code (column A), and 2 agreements on the "11" code (column A under heading "11"). This gives us a total of 45 agreements for rate. This is our numerator for the equation. The denominator includes exact agreements (A columns), within-code agreements (B columns), and disagreements (C and D columns). Adding up all numbers in the Lewis study row gives us our denominator (the denominator is equal to the total number of utterances coded and compared in the study). So by plugging the CR - JM rate numbers (excluding 0's to make the equation less cumbersome) from the Lewis study into our equation, we get

$$\frac{43+2}{43+2+3} = \frac{45}{48} \times 100 = 93.8\%$$

This same method/calculation was used to calculate the "Exact Agreement with Appropriate" percentages for all categories, study groups, and transcriptionist/coder comparison groups.

The "Exact Agreement on Inappropriate Only" percentages do not include any instances of agreement on the "1" (appropriate) code. These percentages are unfortunately low in many cases. This percentage is calculated by adding up the A columns for the inappropriate codes only and dividing that by the sum of the A, B, C, and D columns (again, for the inappropriate codes only). For example, if we again look at the CR - JM table under the Rate section, we see with the Pred2 study that there were two instances of exact agreement on utterances coded inappropriate by both coders (one in column A under code "9" and one in column A under code "10"). Our numerator, therefore, is 2. Codes 11 and 12 were not used at all in the transcripts/PVSPs compared for this study, so to get our denominator we add up the A, B, C, and D columns under codes 9 and 10 (remember, appropriate agreements are not included in this calculation). So, if we plug the numbers (again, taking out the 0's) into our "Exact Agreement on Inappropriate Only" equation, we get

$$\frac{1+1}{1+2+1+1+1} = \frac{2}{6} \times 100 = 33.3\%$$

This percentage is the exact agreement in those instances where one or both coders considered one or more utterances to be inappropriate in a prosody-voice domain of interest.

Finally, within-code agreement is a calculation of the percentage of coder agreement on the use of an inappropriate code, whether or not it was an exact agreement. Basically, it tells us the coder agreement for use of an appropriate vs. an inappropriate code (no matter what inappropriate code was used by each coder). Again, we do not include the exact agreements on appropriate codes (Code 1, column A) in this calculation. The "Within-Code Agreement" percentage will always be the same or better than the "Exact Agreement on Inappropriate Only" percentage because the numerator will be the sum of the A and B columns (for the inappropriate codes only) rather than just the A columns. The denominator, again, is the sum of the A, B, C, and D columns for the inappropriate codes only. For instance, if we again look at the CR - JM table, this time under Stress in the Pred2 study, we find 10 instances of exact agreement on code 15 (column A) and 1 within-code agreement on code 16 (column B). The sum of these agreements makes up our numerator. The denominator is the sum of the A, B, C, and D columns for all of the inappropriate codes (13, 14, 15, 16). Plugging the numbers (0's excluded) into the equation, we get

$$\frac{10+1}{2+1+10+8+3+1} = \frac{11}{25} \times 100 = 44.0\%$$

Note that in this case the within-code agreement is slightly higher than the exact agreement for the inappropriate codes only (because of the one within-code agreement on code 16).

What Does It All Mean?

Phrasing. Of all the PVSP categories, phrasing has one of the highest percentage agreements for all three coder comparison groups. The "Exact Agreement With Appropriate" (hereafter referred to as EAWA) calculations/percentages appear consistently high and stable across study groups. There is, however, more variation in percentage agreements for the "Exact Agreement on Inappropriate Only" (EAIO) and "Within-Code Agreement" (WCA) calculations.

It may be that taking the "appropriate" agreements out of the equation reduces the numerators and denominators to such a degree that a small difference in the numerator or denominator can change the percentage quite drastically (i.e., $1/3 = 33.3\%$, but $2/3 = 66.6\%$). Also, inappropriate phrasing can often be a characteristic of the speech of certain subjects (with dysfluencies, cluttering problems, word retrieval problems, etc.) rather than a characteristic present to some extent in all subjects. So, again, it is important to remember that the subject characteristics of the various studies may at least partially account for the differences in the percent agreement among the three study groups.

One trend important to note in the Phrasing section is the consistent increase in percentage agreement from the EAIO to the WCA. This means that, while the coders usually agree that phrasing is inappropriate in an utterance, in some cases they are using different inappropriate phrasing codes to describe the inappropriate behavior. Fine-tuning (or just plain reviewing!) the definitions for the inappropriate phrasing codes could help us improve our exact agreement. A quick scan of the within-code phrasing agreements indicates a possible difficulty in judging a repetition (codes 2, 3) vs. a revision (codes 6, 7, 8). Single (codes 2, 3) vs. multiple repetitions (codes 4, 5, 7, 8) of sound/syllables and/or words is also disagreed upon fairly often.

Several points regarding judgments in the Phrasing category came up as Larry, Carmen, and Jane were going through the training tape, and it seems most appropriate to mention those points here. One involves capturing all glossed part words with a phrasing code. If a part word comes at the end of an utterance without completion of that word, code this event as "6" (one word revision). If the word is completed in the next utterance, code the part word event as a "2" (sound/syllable repetition), since the completed word does not need to immediately follow the part word, nor does it even need to occur in the same utterance as the part word. There are some examples of these things in Sample 6 on the training tape.

Rate. Considering that rate coding is fairly straightforward in the sense that an utterance is timed in syllables per second and then coded as appropriate or inappropriate based on pre-established criteria, we're not doing very well agreement-wise in this category. The good news is that, given the objective nature of the decision-making process for judging rate, we should easily be able to boost our EAIO and WCA percentages!

Again, the EAWA agreement percentages are quite high, because the majority of utterances for all speakers in general are judged to be appropriate. With the exception of a few (five) disagreements on using code 9 vs. code 10, most of the disagreements for all three coder comparisons groups concern whether or not rate can be considered appropriate.

As with phrasing, all it may take to improve agreement in the Rate category is a review of the PVSP guidelines/criteria for rate described in the manual. Remember that the "cut-off" rates for speech are less than 2 syllables/second for slow rate and greater than 4 syllables/second for fast rate in children (with additional judgments concerning pause time or acceleration). In adult (12 years and older) speech, the cut-off rates are less than 2 syllables/second for slow rate and greater than 6 syllables/second to qualify as fast rate. Probably the best rule of thumb is, "If you think it's inappropriate, use a stopwatch to time the utterance (preferably several times to be sure of your results)." Also note that in order to use code "12" (Fast/Acceleration), the whole utterance needs to be fast (>4 syl/sec for children or >6 syl/sec for adults), not just the portion of the utterance considered to be accelerated. Another way to say it is that although an utterance may be accelerated, if the utterance as a whole does not exceed the fast rate criteria indicated above, it cannot be coded "12".

Stress. As you can see from the PVSP tables, the percentage agreement for stress is all over the board (from 0.0% agreement to 100.0% agreement from one study and coder comparison group to another). Again, the EAWA percentages look pretty good (with the exception of the CW - JM comparison group, where overall EAWA was 76.1%), but the EAIO and WCA percentage agreements are quite low.

Agreement in the judgment of stress may be so low because judging stress is a highly perceptual task, and making decisions about the appropriateness of stress in conversational speech is extremely difficult. The inappropriate stress code used most often is 15, and while it accounts for the highest number of agreements for all of the inappropriate stress codes used, it also accounts for the highest number of disagreements. Nearly all of these disagreements are over the decision of appropriate ("1") vs. inappropriate (mostly "15," with a smattering of the other inappropriate stress codes), as opposed to disagreement over which inappropriate code to use.

In comparing the C and D columns for the three coder comparison groups, we see the following: In the CR - CW comparison group, Carol used about 1.5 times as many "15" codes overall as Carmen; in the CW - JM group, Carol used over twice as many "15" codes as Jane; and in the CR - JM group, Carmen used almost twice as many "15" codes as Jane. We infer from this information that Carmen's judgments may represent a nice "middle ground," with Carol being too strict at one end and Jane being too lenient at the other.

Improving agreement in the Stress category may be a pipe dream, but let's give it a shot. If Carol and Jane make an effort to listen for inappropriate stress less and more, respectively, and attempt to retrain themselves in their perception of appropriate vs. inappropriate stress using past training tapes and the 1995 training tape, we should be able to reach Carmen's hallowed "middle ground" and thus achieve respectable percentage agreements in the Stress category.

Several points relevant to stress follow. First, remember that revisions and repetitions are accounted for under the Phrasing category, not the Stress category. Secondly, a prolongation is appropriate if judged to be a "thought prolongation" (i.e., it seems that the speaker is "buying time" or trying to keep a conversational turn), especially when the speaker is describing something that involves a list or series (i.e., "This summer I went to the Dells, went to camp, took swimming lessons," etc.). These situations can be coded "1" or a circled "15", with a note ("Thought PRO")

in the Comments section. Third, it is important to distinguish among the various types of inappropriate stress when using Code 15 (Excessive/Equal/Misplaced Stress). Whenever the 15 code is used, be sure to specify why the utterance has been coded "15" in the Comments section (MP, EE, PRO, BLO). As a guideline, it has been found that most utterances coded "15" can be attributed to "EE" (excessive/equal) stress in speakers, while "MP" (misplaced) stress shows up infrequently, occurring most often as a unique event possibly due to a dialectal variation in speech.

Occasionally a child's speech (especially stress) is influenced by a "play register"--a particular mode of speech the child uses while in a play context. These utterances cannot be excluded unless the child is clearly using a character register. If the child's stress can be considered normal within the play context and register, code it as appropriate but make a note in the Comments section ("play register").

Loudness. Inappropriate loudness codes are not used often, so EAWA agreements are high, with the agreement on appropriate utterance judgments being the most positive contributing factor as usual. The EAIO and WCA agreements tend to be lower and more variable. I think part of the reason for this variability is the idea that once a coder determines that an utterance is too soft or loud, she in essence makes a decision that the speaker has a soft or a loud voice, and that decision, on some level, may influence perceptual judgments on subsequent utterances in the sample. If the other coder does not classify that same speaker as loud or soft to begin with, she will have tendency to code the utterances as appropriate for loudness. This sets up a situation for low agreement (especially EAIO and WCA).

In terms of usage, Carol appears to use the most inappropriate codes (when another coder uses an appropriate code), and Carmen uses more inappropriate codes than Jane (remember the "middle ground" under the Stress section?). The strategy for improving agreement on inappropriate loudness codes may be similar to that described previously in the Stress section.

As with stress, making judgments about a speaker's loudness level is difficult because it is so perceptual. It can also be confounded by the quality of the recording and mic. distance (for both the clinician and the speaker of interest). It seems that the best we can do to improve agreement is to review the guidelines in the PVSP manual, be sure that we're judging on an utterance-by-utterance basis rather than on a decision we've made about the speaker's overall loudness level, be sensitive

to other signals and noises on the recording that can give us clues concerning the loudness of the speaker's voice relative to those other sounds, and, when there is not a clear reference (from the examiner for instance) be conservative when judging loudness. As with stress coding, strive for that middle ground!

Pitch. As with loudness codes, inappropriate pitch codes are seldom used relative to the number of appropriate codes used. Consequently, the EAWA percentages are high for all three coder comparison groups, and the EAIO and WCA are lower. Codes 21 and 22 were not used at all by any of the coders in any of the samples compared.

All disagreements were appropriate vs. inappropriate (no within-code agreements). Code 19 appears to be used most often, although in the CR - CW comparison group, Carol used quite a few 20 codes (see Column D under code 20 in the Pitch section) in the Iowa1 and Lewis samples compared. All "20" codes in the Iowa1 study were used on sample AKLEI-C1, a six-year-old female. Most of the "20" codes used in the Lewis study were on BHIZA-C1 (an adolescent male) and TSCHI-C1 (an adult male).

Probably the best strategy for improving agreement in the Pitch category, in addition to recalibrating our pitch perceptions with training tapes, is to use the rule of thumb stated in the PVSP manual: "...for appropriate pitch...the pitch level should not 'call attention to itself' in a social situation." Also, in situations where the speaker's voice contains elements of both low pitch and roughness, capture the quality under the inappropriate laryngeal code rather than the inappropriate pitch code.

Laryngeal Features. Judgment of laryngeal features is problematic, and it's not surprising that the agreement in this PVSP category is low. Of all the PVSP categories, Laryngeal Features has the highest inappropriate code usage, which lowers the EAWA percentages (we don't have all those nice "1" agreements to boost the numerator!). The EAIO and WCA percentages are low too, but not much lower overall in comparison to the EAIO's and WCA's in the other PVSP categories.

In terms of usage, codes 27 and 28 were not used at all in the samples compared. Of the inappropriate codes used, "29" was used least often, and code 24 was used most often. Carol used the most inappropriate codes -- in our comparison it was about 1.5 times as many as Carmen and two

times as many as Jane. Carmen used about 1.5 times as many inappropriate codes as Jane, with Jane using the least number of inappropriate codes.

In nearly all cases, the WCA's are higher than the EAIO's, meaning that there are agreements on the inappropriateness of utterances, but not on the inappropriate codes used. Carol appears to use more 23 codes when another coder uses 24. Jane appears to use a greater variety of codes (25, 26, 29) when another coder uses 23 or 24. The code 23 vs. code 24 judgment seems to pose the greatest difficulty, so fine-tuning our perception of breathy (23) vs. rough (24) vs. hoarse (rough & breathy - again, code 24) would be a great way to boost our exact agreement in the Laryngeal Features category (easier said than done, right?).

In order to improve our agreement in the Laryngeal Features category, there are several things we can do. First, I think it is important for Carol to become a bit more lenient in her judgment of what qualifies as inappropriate, and Jane needs to become more strict. On our training tape, there are many examples of voices that caused disagreements among coders. Listening to the tape and seeing how the other coders scored the utterances should help narrow our perceptions of 1) appropriate vs. inappropriate laryngeal quality, and 2) the type of laryngeal quality exhibited if it is judged to be inappropriate. Reviewing the guidelines in the PVSP manual, reviewing the training tapes periodically to prevent "drift," and always having Table 2 on page 41 of the PVSP manual close by for easy reference should also help to fine-tune our perceptions and improve our reliability. [NOTE: Always keep in mind the 50%/4+ criterion in operation for inappropriate laryngeal codes 23, 24, 25, and 29, and the fact that filler words **are not** included in the total word count.] Any other suggestions?

Resonance. Coding of Resonance, like Laryngeal Features, requires difficult perceptual judgments concerning whether a speaker's vocal quality is appropriate (i.e., "within the normal range") or inappropriate, with further classification of the inappropriate quality. The EAWA percentages in this category are pretty good; the majority of judgments were "appropriate," with "inappropriate" judgments confined to a few of the subjects in each study group. As usual, the EAIO and WCA percentages were pretty low.

Unlike other PVSP categories, Carol was the most lenient with speakers when judging resonance as appropriate or inappropriate. Carol used fewer inappropriate codes than Carmen or Jane, and Carmen and Jane used inappropriate resonance codes with similar frequency.

In all three coder comparison groups, there is an increase in overall percentage agreement from the EAIO's to the WCA's, indicating a fair number of instances of disagreement over which inappropriate code to use. All but one within-code agreement was on the use of code 31 (denasal) vs. code 32 (nasopharyngeal). In many of these cases, Carmen used code 32 when Carol and Jane used code 31. This may be a very important distinction to consider in our strategy for improving agreement for Resonance coding.

I hope that the training tape will help us fine-tune the perceptual distinctions among the three types of inappropriate resonance. This is a difficult feature to code reliably.

Summary

As you can see, we need improvement in our PVSP coding reliability, especially on the exact agreement of inappropriate codes. The Laryngeal Features category is in need of the most attention, based on our three comparison groups. We also need to improve our reliability in the coding of Resonance and Stress. While the reliability/agreement information presented may seem overwhelming and discouraging, it's important to get a sense of where we are so that we can determine what we need to do to get where we want to be (not perfect, just "reliable"). Hopefully the training tape will aid in improving agreement on the problem areas in transcription and PVSP coding.

As you listen to the training tape and review the information on reliability, you may have additional ideas and suggestions to help improve agreement. Share these ideas! That goes for transcription as well as PVSP coding.

Miscellaneous Notes

During meetings to make definitive judgments on the training samples, some general conventions for glossing, segmentation, and transcription came up that, for lack of a better spot, will be

mentioned here. There appear to be some disagreements among the transcriptionists on how an utterance should be segmented. In the PVSP manual, Segmentation Rule 4 on page 10 is the most relevant in these cases. Please refer to this rule (as well as the others) to help clear up these disagreements.

Dates (such as "May twentythird") are considered as two words and should not be excluded as one-word utterances in cases where a date is the entire utterance. The same goes for an utterance such as "eighth grade."

Words such as "yeah" and "yep" that occur very frequently in speech samples can be transcribed "loosely." Don't get too hung up on narrow transcription and realize the wide array of acceptable forms for these words.

PVSP Comparison: CR and CW

PHRASING

		1				2				3				4				5				6				7				8			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	198	n\a	n\a	n\a	2	2	2	0	6	0	0	0	1	0	0	0	--	--	--	--	9	1	4	3	--	--	--	--	1	3	0	0
Lewis	12	246	n\a	n\a	n\a	5	2	1	0	10	0	4	3	1	2	0	0	1	1	0	0	2	0	4	4	--	--	--	--	1	1	0	0
Pred2	6	94	n\a	n\a	n\a	4	1	0	0	11	5	1	1	0	1	0	0	1	1	0	0	0	0	0	1	--	--	--	--	1	1	0	0
All	28	538	--	--	--	11	5	3	0	27	5	5	4	2	3	0	0	2	2	0	0	11	1	8	8	--	--	--	--	3	5	0	0

PHRASING: Exact Agreement with Appropriate: Iowa1= 93.5% Lewis= 92.4% Pred2= 90.2% All= 92.4%
 Exact Agreement on Inappropriate Only: Iowa1= 55.9% Lewis= 47.6% Pred2= 58.6% All= 53.3%
 Within Code Agreement: Iowa1= 73.5% Lewis= 61.9% Pred2= 89.7% All= 73.3%

RATE

		1				9				10				11				12			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	217	n\a	n\a	n\a	0	0	0	1	1	0	0	4	1	0	5	3	--	--	--	--
Lewis	12	273	n\a	n\a	n\a	--	--	--	--	--	--	--	--	1	0	9	3	0	0	0	2
Pred2	6	114	n\a	n\a	n\a	0	0	1	1	0	1	1	0	0	0	2	1	0	0	1	1
All	28	604	--	--	--	0	0	1	2	1	1	1	4	2	0	16	7	0	0	1	3

RATE: Exact Agreement with Appropriate: Iowa1= 94.4% Lewis=95.1% Pred2= 92.7% All= 94.4%
 Exact Agreement on Inappropriate Only: Iowa1=13.3% Lewis= 6.7% Pred2= 0.0% All= 7.7%
 Within- Code Agreement: Iowa1= 13.3% Lewis= 6.7% Pred2= 11.1% All= 10.3%

STRESS

		1				13				14				15				16			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	209	n\a	n\a	n\a	--	--	--	--	0	0	1	1	4	0	5	12	--	--	--	--
Lewis	12	261	n\a	n\a	n\a	--	--	--	--	--	--	--	--	8	0	8	11	--	--	--	--
Pred2	6	84	n\a	n\a	n\a	0	1	0	1	0	1	0	0	13	1	5	16	0	0	1	0
All	28	554	--	--	--	0	1	0	1	0	1	1	0	25	1	18	39	0	0	1	0

STRESS: Exact Agreement with Appropriate: Iowa1= 91.8% Lewis= 93.4% Pred2= 78.9% All= 90.0%
 Exact Agreement on Inappropriate Only: Iowa1= 17.4% Lewis= 29.6% Pred2= 33.3% All= 28.1%
 Within-Code Agreement: Iowa1= 17.4% Lewis= 29.6% Pred2= 41.0% All= 31.5%

LOUDNESS

		1				17				18			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	190	n\ a	n\ a	n\ a	4	0	0	27	4	0	1	6
Lewis	12	244	n\ a	n\ a	n\ a	0	0	0	34	1	0	3	6
Pred2	6	108	n\ a	n\ a	n\ a	6	0	1	3	0	0	5	0
All	28	542	--	--	--	10	0	1	64	5	0	9	12

LOUDNESS: Exact Agreement With Appropriate: Iowa1= 85.3% Lewis= 85.1% Pred2= 92.7% All= 86.6
 Exact Agreement on Inappropriate Only: Iowa1= 19.0% Lewis= 2.3% Pred2= 40.0% All= 14.9%
 Within-Code Agreement: Iowa1= 19.0% Lewis= 2.3% Pred2= 40.0% All= 14.9%

PITCH

		1				19				20				21				22			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	208	n\ a	n\ a	n\ a	2	0	1	2	0	0	0	19	--	--	--	--	--	--	--	--
Lewis	12	269	n\ a	n\ a	n\ a	0	0	2	5	0	0	0	12	--	--	--	--	--	--	--	--
Pred2	6	116	n\ a	n\ a	n\ a	3	0	2	2	--	--	--	--	--	--	--	--	--	--	--	--
All	28	593	--	--	--	5	0	5	9	0	0	0	31	--	--	--	--	--	--	--	--

PITCH: Exact Agreement with Appropriate: Iowa1= 90.5% Lewis= 93.4% Pred2= 96.7% All= 93.0%
 Exact Agreement on Inappropriate Only: Iowa1= 8.3% Lewis= 0.0% Pred2= 42.9% All= 10.0%
 Within-Code Agreement: Iowa1= 8.3% Lewis= 0.0% Pred2= 42.9% All= 10.0%

LARYNGEAL

		1				23				24				25				26				27				28				29			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D				
Iowa1	10	105	n\ a	n\ a	n\ a	0	0	0	22	30	5	5	34	0	0	0	16	5	2	1	5	--	--	--	--	--	--	--	--	0	0	0	2
Lewis	12	155	n\ a	n\ a	n\ a	1	0	0	14	56	24	17	13	0	1	0	1	1	0	1	2	--	--	--	--	--	--	--	--	1	0	0	1
Pred2	6	43	n\ a	n\ a	n\ a	2	1	0	9	28	3	7	18	0	0	0	1	3	3	0	2	--	--	--	--	--	--	--	0	2	1	0	
All	28	303	--	--	--	3	1	0	45	114	32	29	65	0	1	0	18	9	5	2	9	--	--	--	--	--	--	--	1	2	1	3	

LARYNGEAL: Exact Agreement with Appropriate: Iowa1= 60.3% Lewis= 74.3% Pred2= 61.8% All= 66.9%
 Exact Agreement on Inappropriate Only: Iowa1= 27.6% Lewis= 44.4% Pred2= 41.3% All= 37.4%
 Within-Code Agreement: Iowa1= 33.1% Lewis= 63.2% Pred2= 52.5% All= 49.4%

RESONANCE

		1				30				31				32			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	199	n\ a	n\ a	n\ a	4	0	5	10	3	0	9	2	--	--	--	--
Lewis	12	232	n\ a	n\ a	n\ a	0	0	2	3	21	0	3	3	0	7	17	0
Pred2	6	105	n\ a	n\ a	n\ a	0	0	10	8	--	--	--	--	--	--	--	--
All	28	536	--	--	--	4	0	17	21	24	0	12	5	0	7	17	0

RESONANCE: Exact Agreement with Appropriate: Iowa1= 88.8% Lewis= 87.8% Pred2=85.4% All= 87.7%
 Exact Agreement on Inappropriate Only: Iowa1= 21.2% Lewis= 37.5% Pred2= 0.0% All= 26.2%
 Within-Code Agreement: Iowa1= 21.2% Lewis= 50.0% Pred2= 0.0% All= 32.7%

PVSP Comparison: CW and JM

PHRASING

		1				2				3				4				5				6				7				8			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	2	37	n/a	n/a	n/a	--	--	--	--	1	0	0	0	--	--	--	--	--	--	--	--	1	0	1	0	--	--	--	--	--	--	--	--
Lewis	2	38	n/a	n/a	n/a	--	--	--	--	6	0	0	0	1	1	0	0	--	--	--	--	1	0	0	0	--	--	--	--	1	0	0	0
Pred2	6	93	n/a	n/a	n/a	4	0	0	0	11	0	1	1	0	1	0	0	2	3	0	0	1	1	1	1	1	1	0	0	1	2	0	0
All	10	168	--	--	--	4	0	0	0	18	0	1	1	1	2	0	0	2	3	0	0	3	1	2	1	1	1	0	0	2	2	0	0

PHRASING: Exact Agreement with Appropriate: Iowa1= 97.5% Lewis= 97.9% Pred2=90.4% All= 93.4%
 Exact Agreement on Inappropriate Only: Iowa1= 66.7% Lewis= 90.0% Pred2= 62.5% All= 68.9%
 Within-Code Agreement: Iowa1= 66.7% Lewis= 100.0% Pred2= 87.5% All= 88.9%

RATE

		1				9				10				11				12			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	2	35	n/a	n/a	n/a	0	0	1	0	0	2	1	0	--	--	--	--	0	0	0	1
Lewis	2	43	n/a	n/a	n/a	--	--	--	--	--	--	--	--	0	0	2	2	0	1	0	0
Pred2	6	114	n/a	n/a	n/a	0	1	1	2	0	1	0	1	0	0	0	3	0	0	0	2
All	10	192	--	--	--	0	1	2	2	0	3	1	1	0	0	2	5	0	1	0	3

RATE: Exact Agreement with Appropriate: Iowa1= 87.5% Lewis= 89.6% Pred2=91.2% All= 90.1%
 Exact Agreement on Inappropriate Only: Iowa1= 0.0% Lewis= 0.0% Pred2= 0.0% All= 0.0%
 Within-Code Agreement: Iowa1= 40.0% Lewis= 20.0% Pred2= 18.2% All= 23.8%

STRESS

		1				13				14				15				16			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	2	33	n/a	n/a	n/a	--	--	--	--	--	--	--	--	1	0	4	2	--	--	--	--
Lewis	2	35	n/a	n/a	n/a	--	--	--	--	--	--	--	--	0	0	12	1	--	--	--	--
Pred2	6	84	n/a	n/a	n/a	0	0	1	2	--	--	--	--	9	0	23	6	--	--	--	--
All	10	152	--	--	--	0	0	1	2	--	--	--	--	10	0	39	9	--	--	--	--

STRESS: Exact Agreement with Appropriate: Iowa1= 85.0% Lewis= 72.9% Pred2=74.4% All= 76.1%
 Exact Agreement on Inappropriate Only: Iowa1= 14.3% Lewis= 0.0% Pred2= 21.9% All= 16.4%
 Within-Code Agreement: Iowa1= 14.3% Lewis= 0.0% Pred2= 21.9% All= 16.4%

LOUDNESS

		1				17				18			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	2	39	n/a	n/a	n/a	1	0	0	0	--	--	--	--
Lewis	2	42	n/a	n/a	n/a	0	0	0	1	0	0	5	0
Pred2	6	114	n/a	n/a	n/a	3	0	6	2	--	--	--	--
All	10	195	--	--	--	4	0	6	3	0	0	5	0

LOUDNESS: Exact Agreement with Appropriate: Iowa1= 100% Lewis= 87.5% Pred2=93.6% All= 93.4%
 Exact Agreement on Inappropriate Only: Iowa1= 100% Lewis= 0.0% Pred2= 27.3% All= 22.2%
 Within-Code Agreement: Iowa1= 100% Lewis= 0.0% Pred2= 27.3% All= 22.2%

PITCH

		1				19				20				21				22			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	2	40	n/a	n/a	n/a	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Lewis	2	48	n/a	n/a	n/a	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Pred2	6	116	n/a	n/a	n/a	2	0	3	4	--	--	--	--	--	--	--	--	--	--	--	--
All	10	204	--	--	--	2	0	3	4	--	--	--	--	--	--	--	--	--	--	--	--

PITCH: Exact Agreement with Appropriate: Iowa1= 100.0% Lewis= 100.0% Pred2=94.4% All= 96.7%
 Exact Agreement on Inappropriate Only: Iowa1= n/a Lewis= n/a Pred2= 22.2% All= 22.2%
 Within-Code Agreement: Iowa1= n/a Lewis= n/a Pred2= 22.2% All= 22.2%

LARYNGEAL

		1				23				24				25				26				27				28				29			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D				
Iowa1	2	4	n/a	n/a	n/a	4	1	0	0	4	0	19	0	0	0	7	0	--	--	--	--	--	--	--	--	--	--	--	0	1	0	0	
Lewis	2	39	n/a	n/a	n/a	0	0	1	0	0	0	6	0	--	--	--	--	1	0	1	0	--	--	--	--	--	--	--	--	--	--		
Pred2	6	50	n/a	n/a	n/a	3	4	7	0	13	9	26	4	0	0	1	0	4	0	2	1	--	--	--	--	--	--	--	--	0	0	0	1
All	10	93	--	--	--	7	5	8	0	17	9	41	4	0	0	8	0	5	0	3	1	--	--	--	--	--	--	--	0	1	0	1	

LARYNGEAL: Exact Agreement with Appropriate: Iowa1= 30.0% Lewis= 83.3% Pred2= 56.0% All= 57.3%
 Exact Agreement on Inappropriate Only: Iowa1= 22.2% Lewis= 11.1% Pred2= 26.7% All= 24.2%
 Within-Code Agreement: Iowa1= 27.8% Lewis= 11.1% Pred2= 44.0% All= 36.7%

RESONANCE

		1				30				31				32			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	2	38	n/a	n/a	n/a	--	--	--	--	0	0	0	2	--	--	--	--
Lewis	2	29	n/a	n/a	n/a	0	0	1	0	5	2	0	6	0	0	0	5
Pred2	6	106	n/a	n/a	n/a	0	1	7	0	0	0	0	11	--	--	--	--
All	10	173	--	--	--	0	1	8	0	5	2	0	19	0	0	0	5

RESONANCE: Exact Agreement with Appropriate: Iowa1= 95.0% Lewis= 70.8% Pred2= 84.8% All= 83.6%
 Exact Agreement on Inappropriate Only: Iowa1= 0.0% Lewis= 26.3% Pred2= 0.0% All= 12.5%
 Within-Code Agreement: Iowa1= 0.0% Lewis= 36.8% Pred2= 5.3% All= 20.0%

PVSP Comparison: CR and JM

PHRASING

		1				2				3				4				5				6				7				8			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	2	36	n/a	n/a	n/a	0	1	0	0	1	0	0	0	--	--	--	--	--	--	--	--	1	0	1	0	--	--	--	--	--	--	--	--
Lewis	2	38	n/a	n/a	n/a	--	--	--	--	5	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0	--	--	--	--	1	0	0	0
Pred2	6	94	n/a	n/a	n/a	5	0	0	0	15	1	1	1	1	0	0	0	1	1	0	0	0	0	0	1	--	--	--	--	1	1	0	0
All	10	168	--	--	--	5	1	0	0	21	1	1	2	2	0	0	0	1	2	0	0	2	0	1	1	--	--	--	--	2	1	0	0

PHRASING: Exact Agreement with Appropriate: Iowa1= 95.0% Lewis= 95.8% Pred2= 95.1% All= 95.3%
 Exact Agreement on Inappropriate Only: Iowa1= 50.0% Lewis= 80.0% Pred2= 79.3% All= 76.7%
 Within-Code Agreement: Iowa1= 75.0% Lewis= 90.0% Pred2= 89.7% All= 88.4%

RATE

		1				9				10				11				12			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	2	37	n/a	n/a	n/a	0	0	0	2	--	--	--	--	--	--	--	--	0	0	0	1
Lewis	2	43	n/a	n/a	n/a	--	--	--	--	--	--	--	--	2	0	3	0	--	--	--	--
Pred2	6	117	n/a	n/a	n/a	1	0	0	2	1	0	1	1	--	--	--	--	--	--	--	--
All	10	197	--	--	--	1	0	0	4	1	0	1	1	2	0	3	0	0	0	0	1

RATE: Exact Agreement with Appropriate: Iowa1= 92.5% Lewis= 93.8% Pred2= 96.7% All= 95.3%
 Exact Agreement on Inappropriate Only: Iowa1= 0.0% Lewis= 40.0% Pred2= 33.3% All= 28.6%
 Within-Code Agreement: Iowa1= 0.0% Lewis= 40.0% Pred2= 33.3% All= 28.6%

STRESS

		1				13				14				15				16			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	2	37	n/a	n/a	n/a	--	--	--	--	--	--	--	--	3	0	0	0	--	--	--	--
Lewis	2	41	n/a	n/a	n/a	--	--	--	--	--	--	--	--	1	0	6	0	--	--	--	--
Pred2	6	98	n/a	n/a	n/a	0	0	0	2	0	0	1	0	10	0	8	3	0	1	0	0
All	10	176	--	--	--	0	0	0	2	0	0	1	0	14	0	14	3	0	1	0	0

STRESS: Exact Agreement with Appropriate: Iowa1= 100.0% Lewis= 87.5% Pred2= 87.8% All= 90.0%
 Exact Agreement on Inappropriate Only: Iowa1= 100.0% Lewis= 14.3% Pred2= 40.0% All= 40.0%
 Within-Code Agreement: Iowa1= 100.0% Lewis= 14.3% Pred2= 44.0% All= 42.9%

LOUDNESS

		1				17				18			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	2	39	n/a	n/a	n/a	1	0	0	0	--	--	--	--
Lewis	2	43	n/a	n/a	n/a	0	0	0	1	0	0	4	0
Pred2	6	109	n/a	n/a	n/a	4	0	4	1	0	0	5	0
All	10	191	--	--	--	5	0	4	2	0	0	9	0

LOUDNESS: Exact Agreement with Appropriate: Iowa1= 100% Lewis= 89.6% Pred2= 91.9% All= 92.9%
 Exact Agreement on Inappropriate Only: Iowa1= 100% Lewis= 0.0% Pred2= 28.6% All= 25.0%
 Within-Code Agreement: Iowa1= 100% Lewis= 0.0% Pred2= 28.6% All= 25.0%

PITCH

		1				19				20				21				22			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	2	40	n/a	n/a	n/a	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Lewis	2	48	n/a	n/a	n/a	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Pred2	6	115	n/a	n/a	n/a	5	0	1	2	--	--	--	--	--	--	--	--	--	--	--	--
All	10	203	--	--	--	5	0	1	2	--	--	--	--	--	--	--	--	--	--	--	--

PITCH: Exact Agreement with Appropriate: Iowa1= 100.0% Lewis= 100.0% Pred2= 97.6% All= 98.6%
 Exact Agreement on Inappropriate Only: Iowa1= n/a Lewis= n/a Pred2= 62.5% All= 62.5%
 Within-Code Agreement: Iowa1= n/a Lewis= n/a Pred2= 62.5% All= 62.5%

LARYNGEAL

		1				23				24				25				26				27				28				29			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D				
Iowa1	2	27	n/a	n/a	n/a	--	--	--	--	4	1	7	0	--	--	--	--	0	0	0	1	--	--	--	--	--	--	--	--	--			
Lewis	2	45	n/a	n/a	n/a	--	--	--	--	0	0	2	0	--	--	--	--	1	0	0	0	--	--	--	--	--	--	--	--	--			
Pred2	6	64	n/a	n/a	n/a	3	0	0	1	13	3	22	6	0	0	0	1	4	2	0	1	--	--	--	--	--	--	--	--	1	1	1	0
All	10	136	--	--	--	3	0	0	1	17	4	31	6	0	0	0	1	5	2	0	2	--	--	--	--	--	--	--	--	1	1	1	0

LARYNGEAL: Exact Agreement with Appropriate: Iowa1= 77.5% Lewis= 95.8% Pred2= 69.1% All= 76.8%
 Exact Agreement on Inappropriate Only: Iowa1= 30.8% Lewis= 33.3% Pred2= 35.6% All= 34.7%
 Within-Code Agreement: Iowa1= 38.5% Lewis= 33.3% Pred2= 45.8% All= 44.0%

RESONANCE

		1				30				31				32			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	2	37	n/a	n/a	n/a	--	--	--	--	1	0	1	1	--	--	--	--
Lewis	2	24	n/a	n/a	n/a	0	12	0	0	--	--	--	--	7	0	5	0
Pred2	6	101	n/a	n/a	n/a	0	0	10	0	0	0	0	12	--	--	--	--
All	10	162	--	--	--	0	12	10	0	1	0	1	13	7	0	5	0

RESONANCE: Exact Agreement with Appropriate: Iowa1= 95.0% Lewis= 64.6% Pred2= 82.1% All= 80.6%
 Exact Agreement on Inappropriate Only: Iowa1= 33.3% Lewis= 29.2% Pred2= 0.0% All= 16.3%
 Within-Code Agreement: Iowa1= 33.3% Lewis= 79.2% Pred2= 0.0% All= 40.8%

INTRAJUDGE RELIABILITY

Carmen

Broad and Narrow Phonetic Transcription Agreement

A total of 22 transcripts were compared, 12 from the Lewis study and 10 from the IOWA1 study. Percentage agreements are given below for each study individually and for the two studies combined.

In the Lewis study, **consonant** percentage agreements were as follows: narrow agreement was 87.5% (range 75.6% - 95.2%) and broad agreement (with the underbar symbol, or "deletions," not included) was 95.9% (range 92.3% - 99.0%). **Vowel** percentage agreements were 85.0% (range 77.2% - 92.1%) for narrow and 89.2% (range 85.1% - 93.4%) for broad. Overall **diacritic** agreement was 41.5% (range 22.2% - 54.5%).

In the IOWA1 study, **consonant** percentage agreements were 88.6% (range 83.1% - 92.9%) for narrow and 97.1% (range 94.9% - 98.7%) for broad. **Vowel** percentages were 87.2% (range 77.1% - 92.6%) for narrow agreement and 92.6% (range 89.7% - 96.9%) for broad agreement. Overall agreement for **diacritics** was 58.0% (range 44.8% - 68.8%).

Agreement percentages were also computed for the two studies combined (i.e., 22 transcripts). Narrow agreement for **consonants** was 88.0% (range 75.6% - 95.2%), and broad agreement for **consonants** was 96.4% (range 92.3% - 99.0%). Narrow agreement for **vowels** was 85.9% (range 77.1% - 92.6%), and broad agreement for **vowels** was 90.7% (range 85.1% - 96.9%). Overall **diacritic** agreement was 50.2% (range 22.2% - 68.8%).

Ways to Improve Intrajudge Phonetic Transcription Agreement

The greatest sources of narrow disagreement for consonants are as follows: use of the nasality symbols on nasal consonants, use of stop release symbols on the stop consonants, use of the dentalized and palatalized symbols on the alveolar fricatives, and the labialized and derhotacized

symbols on the liquids. There are no particular vowels that seem to contribute the most to narrow disagreement; some of the symbols associated with the vowels that may be contributing to lower narrow agreements are the centralized symbol, the lengthened symbol, the glottalized symbol, the lowered symbol, and the nasality symbols.

Prosody-Voice Screening Profile Intrajudge Agreement

Following this discussion is a PVSP Comparison table that is set up similarly to the tables used for the interjudge PVSP comparison. In this table, Carmen's first coding of the sample is considered the "standard" against which the second coding of the sample is compared.

The table is broken down into the categories of Phrasing, Rate, Stress, Loudness, Pitch, Laryngeal Features, and Resonance. For each category, the appropriate (always "1") and inappropriate codes are listed right below the category heading from left to right. For instance, under the "Rate" category heading is listed a "1" for Appropriate, "9" for Slow Articulation/Pause Time, "10" for Slow/Pause Time, "11" for Fast, and "12" for Fast/Acceleration.

The Study column indicates which studies the transcripts/PVSPs being compared came from. In this case, the samples were all taken from the IOWA1 and Lewis studies. The "#T" column represents the number of transcripts/PVSPs used for comparison from each study.

Under each appropriate and inappropriate code in the PVSP category sections are listed four columns labeled A, B, C, and D. "A" represents the number of exact agreements, i.e., the number of times the same code was used by Carmen (whether appropriate or inappropriate) on the same utterance for "time 1" (the first time the utterance was coded) and "time 2" (the second time the same utterance was coded). Included in the exact agreements on appropriate codes (code "1" subheading "A") are instances where Carmen used "1" to indicate appropriate both times, and instances where she used a "1" one time and a circled, inappropriate code the other time for that same utterance. (Remember that a circled code is arbitrarily considered appropriate for purposes of comparison here). However, cases in which Carmen used the same inappropriate code both times to describe an utterance, even if one or both of the codes was/were circled, are also considered exact agreements and would be recorded under column "A" under that particular inappropriate code heading.

Column "B" represents within-code agreements on inappropriate codes. For instance, if Carmen used two different inappropriate Phrasing codes to describe the same utterance, that situation would be recorded in column "B" under the code she used in her first coding (time 1). That is because, as mentioned previously, the first coding of the transcript is considered the "standard" here.

Column "C" represents the number of times Carmen used an inappropriate code to describe an utterance the first coding time, and an appropriate code to describe that same utterance the second time. For instance, if Carmen coded an utterance "24" the first time and "1" the second time, this would be recorded in column "C" under Code "24" in the Laryngeal category.

Column "D" on the other hand represents the disagreements where Carmen coded an utterance as appropriate the first time and inappropriate the second time. For instance, if she coded an utterance "1" the first time and "15" the second time, this event would be recorded in the Stress category in column "D" under code 15.

To summarize and clarify, column "A" represents exact agreements for each code, column "B" represents within-code agreements for each code, column "C" represents a first time inappropriate and second time appropriate judgment, and column "D" represents a first time appropriate and second time inappropriate judgment. Therefore, under columns "A" and "B" are recorded instances of agreement; under columns "C" and "D" are recorded instances of disagreement between times 1 and 2.

Since the "B," "C," and "D" columns are not applicable in the case of appropriate codes, there are "-"s in these areas of the table. All "0"s indicate that the code in question was not used.

The "Exact Agreement With Appropriate" line at the end of each section displays the percent agreement of all appropriate and inappropriate codes used for each group (IOWA1, Lewis, and "All", which is both studies combined). This includes exact agreements only and is calculated by adding up all the "A" columns and dividing by the total of all of the columns (A, B, C, and D). The "Exact Agreement on Inappropriate Only" percentages do not include agreements on the "1" (appropriate) code. This is calculated by adding all the "A" columns on the inappropriate codes only and dividing that by the total of all of the columns (A, B, C, and D) under the inappropriate codes only. The "Within-Code" agreement is the percentage of exact agreements and within-code agreements on

inappropriate codes only; this is calculated by adding all the "A" and "B" columns under the inappropriate codes only and dividing that by the total of all of the columns (A, B, C, and D) under the inappropriate codes only. Therefore, the "Within-Code Agreement" percentages will always be the same or higher than the "Exact Agreement on Inappropriate Only" percentages.

If needed, please refer to the more detailed description of this information under "How to Read the Tables" in your interjudge agreement discussion.

Phrasing. Percentage agreements are consistently high in this category. Three of the five within-code agreements involved using "8" versus another inappropriate code (i.e., "2" or "5"), signaling perhaps a different interpretation of a sound or part-word from one time to the next (repetition vs. revision). However, there are instances of disagreement, meaning that either something was missed or maybe not even heard from one time to the next.

Rate. Inappropriate rate codes were not used very frequently, but unfortunately the times they were used resulted more often than not in a disagreement in appropriate vs. inappropriate (see columns "C" and "D" under the inappropriate codes, especially under "11"). An accurate timing and calculation of rate in all instances where rate is in question should clear up these disagreements.

Stress. A "15" was used to code most of the utterances considered to have inappropriate stress in the PVSPs compared. There were 21 agreements and 20 disagreements on the "15" code. Because the "D" column shows a total of 15, compared to a total of 5 under the "C" column, there appeared to be a tendency to use more inappropriate stress codes on the second coding of these transcripts. It could represent some "drift" in the perception of inappropriate stress over time, or a clarification of the definition of this particular type of inappropriate stress over time. The training tape might serve as a useful reference in this area.

Loudness. There were more disagreements than agreements on the inappropriate loudness codes (20 total disagreements vs. 9 total agreements). This judgment is highly perceptual, so it's not surprising that agreement percentages on the inappropriate codes are low. Please refer to the discussion of loudness in the interjudge agreement text for some suggestions for improvement.

Pitch. In all 22 PVSPs compared there were only 6 times that an inappropriate pitch code was used. This makes the agreement with appropriate percentages high and the inappropriate only agreement percentages low. Code 19 (Low Pitch/Glottal Fry) was the only inappropriate pitch code used. Remember that use of this code will affect the judgment of laryngeal and resonance features ("*" vs. a code), so it is important to use it reliably.

Laryngeal Features. Laryngeal codes are the most frequently used inappropriate codes of all the categories. The "24" code is the most popular; codes 27 and 28 were not used at all. There were more agreements than disagreements in the use of inappropriate laryngeal codes, which is encouraging. The training tape examples should help to clarify the perception of hoarse voice and help to improve agreement in this category.

Resonance. You can see in the Lewis study that there is quite a jump from the "Exact Agreement on Inappropriate Only" percentage (42.9%) to the "Within-Code Agreement" percentage (69.6%). This was due to the use of a "31" code (Denasal) used the first time and a "32" code (Nasopharyngeal) used the second time (in 15 instances). Maybe going back and listening to the PVSP training tapes could aid in clarifying the perception of these two resonance qualities, thereby improving exact agreement.

Summary. The categories that appear to need the greatest attention as far as improving intrajudge agreement are Rate (especially use of code "11"), Stress (especially use of code "15"), Laryngeal (especially code "24"), the use of code "31" (Denasal) versus "32" (Nasopharyngeal) for Resonance, and, to a lesser extent, the Loudness category. Hopefully, the above observations and guidelines, combined with those in the interjudge text, will help to clear up some of the discrepancies in coding from one time to the next and from one transcriptionist to the next.

PVSP Comparison: CR(Time 1) and CR(Time 2)

PHRASING

		1				2				3				4				5				6				7				8			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	198	--	--	--	5	1	0	1	5	0	1	0	0	1	0	0	0	0	0	0	13	0	1	2	0	0	0	0	3	1	0	0
Lewis	12	250	--	--	--	8	0	0	0	12	0	3	2	3	0	0	0	1	0	0	0	5	0	1	0	0	0	0	0	1	2	0	0
Total	22	448	--	--	--	13	1	0	1	17	0	4	2	3	1	0	0	1	0	0	0	18	0	2	2	0	0	0	0	4	3	0	0

PHRASING: Exact Agreement with Appropriate: Iowa1= 96.6% Lewis= 97.2% All= 96.9%
 Exact Agreement on Inappropriate Only: Iowa1= 76.5% Lewis= 78.9% All= 77.8%
 Within-Code Agreement: Iowa1= 85.3% Lewis= 84.2% All= 84.7%

RATE

		1				9				10				11				12			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	220	--	--	--	0	0	0	3	0	0	1	1	1	0	5	1	0	0	0	0
Lewis	12	269	--	--	--	0	0	0	0	0	0	0	0	5	0	4	10	0	0	0	0
Total	22	489	--	--	--	0	0	0	3	0	0	1	1	6	0	9	11	0	0	0	0

RATE: Exact Agreement with Appropriate: Iowa1= 95.3% Lewis= 95.1% All= 95.2%
 Exact Agreement on Inappropriate Only: Iowa1= 8.3% Lewis= 26.3% All= 19.4%
 Within-Code Agreement: Iowa1= 8.3% Lewis= 26.3% All= 19.4%

STRESS

		1				13				14				15				16			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	216	--	--	--	0	0	0	0	1	0	0	0	9	0	2	4	0	0	0	0
Lewis	12	260	--	--	--	0	0	0	0	0	0	0	2	12	0	3	11	0	0	0	0
Total	22	476	--	--	--	0	0	0	0	1	0	0	2	21	0	5	15	0	0	0	0

STRESS: Exact Agreement with Appropriate: Iowa1= 97.4% Lewis= 94.4% All= 95.8%
 Exact Agreement on Inappropriate Only: Iowa1= 62.5% Lewis= 42.9% All= 50.0%
 Within-Code Agreement: Iowa1= 62.5% Lewis= 42.9% All= 50.0%

LOUDNESS

		1				17				18			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	215	--	--	--	3	0	1	2	5	0	0	6
Lewis	12	276	--	--	--	0	0	0	8	1	0	3	0
Total	22	491	--	--	--	3	0	1	10	6	0	3	6

LOUDNESS: Exact Agreement with Appropriate: Iowa1= 96.1% Lewis= 96.2% All= 96.2%
 Exact Agreement on Inappropriate Only: Iowa1= 47.1% Lewis= 8.3% All= 31.0%
 Within-Code Agreement: Iowa1= 47.1% Lewis= 8.3% All= 31.0%

PITCH

		1				19				20				21				22			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	228	--	--	--	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0
Lewis	12	286	--	--	--	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	22	514	--	--	--	1	0	4	1	0	0	0	0	0	0	0	0	0	0	0	0

PITCH: Exact Agreement with Appropriate: Iowa1= 98.3% Lewis= 99.7% All= 99.0%
 Exact Agreement on Inappropriate Only: Iowa1= 0.0% Lewis= 50.0% All= 16.7%
 Within-Code Agreement: Iowa1= 0.0% Lewis= 50.0% All= 16.7%

LARYNGEAL

		1				23				24				25				26				27				28				29			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D				
Iowa1	10	169	--	--	--	0	0	0	0	30	2	7	13	0	0	0	0	6	2	2	1	0	0	0	0	0	0	0	0	0	0	0	
Lewis	12	139	--	--	--	0	1	0	0	77	1	22	41	0	1	0	1	2	0	0	2	0	0	0	0	0	0	0	0	0	0	1	0
Total	22	308	--	--	--	0	1	0	0	107	3	29	54	0	1	0	1	8	2	2	3	0	0	0	0	0	0	0	0	0	1	0	

LARYNGEAL: Exact Agreement with Appropriate: Iowa1= 88.4% Lewis= 75.7% All= 81.3%
 Exact Agreement on Inappropriate Only: Iowa1= 57.1% Lewis= 53.0% All= 54.2%
 Within-Code Agreement: Iowa1= 63.5% Lewis= 55.0% All= 57.5%

RESONANCE

		1				30				31				32			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	210	--	--	--	4	1	4	0	5	0	7	1	0	0	0	0
Lewis	12	232	--	--	--	1	0	1	2	22	15	2	4	1	0	0	8
Total	22	442	--	--	--	5	1	5	2	27	15	9	5	1	0	0	8

RESONANCE: Exact Agreement with Appropriate: Iowa1= 94.4% Lewis= 88.9% All= 91.3%
 Exact Agreement on Inappropriate Only: Iowa1= 40.9% Lewis= 42.9% All= 42.3%
 Within-Code Agreement: Iowa1= 45.5% Lewis= 69.6% All= 62.8%

Carol

Broad and Narrow Phonetic Transcription Agreement

A total of 22 transcripts were compared, 12 from the Lewis study and 10 from the IOWA1 study. Percentage agreements are given below for each study individually and for the two studies combined.

In the Lewis study, **consonant** percentage agreements were as follows: narrow agreement was 87.3% (range 70.0% - 97.9%) and broad agreement (with the underbar symbol, or "deletions" not included) was 95.9% (range 89.2% - 98.9%). **Vowel** percentage agreements were 85.3% (range 73.4% - 92.5%) for narrow and 89.7% (range 85.5% - 93.2%) for broad. Overall **diacritic** agreement was 36.2% (range 0.0% - 61.5%).

In the IOWA1 study, **consonant** percentage agreements were 85.3% (range 73.5% - 90.3%) for narrow and 95.5% (range 91.3% - 98.2%) for broad. **Vowel** percentages were 85.6% (range 79.9% - 91.3%) for narrow agreement and 90.7% (range 86.1% - 94.7%) for broad agreement. Overall agreement for **diacritics** was 42.7% (range 32.2% - 56.1%).

Agreement percentages were also computed for the two studies combined (i.e., 22 transcripts). Narrow agreement for **consonants** was 86.4% (range 70.0% - 97.9%), and broad agreement for **consonants** was 95.7% (range 89.2% - 98.9%). Narrow agreement for **vowels** was 85.5% (range 73.4% - 92.5%), and broad agreement for **vowels** was 90.1% (range 85.5% - 94.7%). Overall **diacritic** agreement was 39.7% (range 0.0% - 61.5%).

Ways to Improve Intrajudge Phonetic Transcription Agreement

Not surprisingly, the fricatives and liquids are the greatest sources of narrow disagreement for the consonants. The diacritic symbols generally associated with these consonants are the dentalization and palatalization symbols for the fricatives and the labialization and derhotacization symbols for liquids. Transcribing these sounds as accurately as possible is important. There are no particular vowels that seem to contribute the most to narrow disagreement; some of the symbols

associated with the vowels that may be contributing to lower narrow agreements are the glottal symbol, the lengthened symbol, the centralized symbol, and the raised and lowered symbols.

Prosody-Voice Screening Profile Intrajudge Agreement

Following this discussion is a PVSP Comparison table that is set up similarly to the tables used for the interjudge PVSP comparison. In this table, Carol's first coding of the sample is considered the "standard" against which the second coding of the sample is compared.

The table is broken down into the categories of Phrasing, Rate, Stress, Loudness, Pitch, Laryngeal Features, and Resonance. For each category, the appropriate (always "1") and inappropriate codes are listed right below the category heading from left to right. For instance, under the "Rate" category heading is listed a "1" for Appropriate, "9" for Slow Articulation/Pause Time, "10" for Slow/Pause Time, "11" for Fast, and "12" for Fast/Acceleration.

The Study column indicates which studies the transcripts/PVSPs being compared came from. In this case, the samples were all taken from the IOWA1 and Lewis studies. The "#T" column represents the number of transcripts/PVSPs used for comparison from each study.

Under each appropriate and inappropriate code in the PVSP category sections are listed four columns labeled A, B, C, and D. "A" represents the number of exact agreements, i.e., the number of times the same code was used by Carol (both appropriate and inappropriate) on the same utterance for "time 1" (the first time the utterance was coded) and "time 2" (the second time the same utterance was coded). Included in the exact agreements on appropriate codes (code "1" subheading "A") are instances where Carol used "1" to indicate appropriate both times, and instances where she used a "1" one time and a circled, inappropriate code the other time for that same utterance. (Remember that a circled code is arbitrarily considered appropriate for purposes of comparison here). However, cases in which Carol used the same inappropriate code both times to describe the utterance, even if one or both of the codes was/were circled, are also considered exact agreements and would be recorded under column "A" of the corresponding inappropriate code heading.

Column "B" represents within-code agreements on inappropriate codes. For instance, if Carol used two different inappropriate Phrasing codes to describe the same utterance, that situation

would be recorded in column "B" under the code she used in her first coding (time 1). That is because, as mentioned previously, the first coding of the transcript is considered the "standard" here.

Column "C" represents the number of times Carol used an inappropriate code to describe an utterance the first coding time, and an appropriate code to describe that same utterance the second time. For instance, if Carol coded an utterance "24" the first time and "1" the second time, this would be recorded in column "C" under Code "24" in the Laryngeal category.

Column "D" on the other hand represents the disagreements where Carol coded an utterance as appropriate the first time and inappropriate the second time. For instance, if she coded an utterance "1" the first time and "15" the second time, this event would be recorded in the Stress category in column "D" under code 15.

To summarize and clarify, column "A" represents exact agreements for each code, column "B" represents within-code agreements for each code, column "C" represents a first time inappropriate and second time appropriate judgment, and column "D" represents a first time appropriate and second time inappropriate judgment. Therefore, under columns "A" and "B" are recorded instances of agreement; under columns "C" and "D" are recorded instances of disagreement between times 1 and 2.

Since the "B," "C," and "D" columns are not applicable in the case of appropriate codes (i.e., "1" codes only), there are "-"s in these areas of the table. All "0"s indicate that the code in question was not used.

The "Exact Agreement With Appropriate" line at the end of each section displays the percent agreement of all appropriate and inappropriate codes used for each group (IOWA1, Lewis, and "All", which is both studies combined). This includes exact agreements only and is calculated by adding up all the "A" columns and dividing by the total of all of the columns (A, B, C, and D). The "Exact Agreement on Inappropriate Only" percentages do not include agreements on the "1" (appropriate) code. This is calculated by adding all the "A" columns on the inappropriate codes only and dividing that by the total of all of the columns (A, B, C, and D) under the inappropriate codes only. The "Within-Code" agreements are the percentage of exact agreements and within-code agreements on inappropriate codes only; this is calculated by adding all the "A" and "B" columns under the

inappropriate codes only and dividing that by the total of all of the columns (A, B, C, and D) under the inappropriate codes only. Therefore, the "Within-Code Agreement" percentages will always be the same or higher than the "Exact Agreement on Inappropriate Only" percentages.

If needed, please refer to the more detailed description of this information under "How to Read the Tables" in the interjudge agreement discussion. That discussion contains examples that help to further clarify the tables and the calculations.

Phrasing. The "Exact Agreement with Appropriate" percentages are high in this category (lower 90s), but the percentage agreements for the inappropriate codes only are considerably lower. Also, notice that there is quite a difference between the "Exact Agreement on Inappropriate Only" and the "Within-Code Agreement" figures. Since the Phrasing definitions are fairly straightforward, Carol must be interpreting inappropriate phrasing events differently from one time to the next. Hopefully a review of the Phrasing coding guidelines in the PVSP manual would help to improve future reliability in this category.

Rate. As with Phrasing, "Exact Agreement with Appropriate" percentages were high, but the percentage agreements calculated using the inappropriate only codes were low. Since the inappropriate rate codes are determined based on a timing of the utterance and a determination of the syllables-per-second rate for the utterance, it is unclear why the agreement on the inappropriate codes is so low here. Accurate timing is important, as well as calibrating the speed of the transcription machine using the speed calibration tape recorded by Joan (Carol should have a copy of this tape). Note the high number of disagreements on Code 11 ("C" and "D" columns). Because of the straightforwardness of this category, Carol should be able to greatly reduce the number of disagreements without too much trouble.

Stress. The "Exact Agreement With Appropriate" percentages are lower for this category than for Phrasing and Rate. Again, the agreement on only the inappropriate codes is low. Stress is a problematic category, so this is not surprising. Still, there must be something that can be done to improve reliability here. As you can see from the figures in the "D" columns under codes 14 and 15, Carol used many more inappropriate codes in her second coding of the samples than she did in her first coding. This could represent some "drift" in her perception of inappropriate stress over the past few years. Remember that in the interjudge reliability comparison of Carol's use of inappropriate

stress codes, she used more inappropriate stress codes than the other two transcriptionists. "Easing up" on her use of inappropriate stress codes may improve her intrajudge reliability as well as the reliability with the other transcriptionists.

Loudness. The "Exact Agreement With Appropriate" percentage agreements for Loudness were in the low to mid 70s. There were many disagreements on the use of code 17, and quite a few disagreements on code 18 as well. As with the Stress category, you can see by looking in the "D" columns that Carol used more inappropriate loudness codes in her second coding of these samples. Because she used considerably more inappropriate loudness codes than Carmen did on these same samples, it seems that Carol needs to become generally less strict in her judgment of inappropriate loudness. This should improve her intrajudge and interjudge reliability in the long run.

Pitch. The "Exact Agreement With Appropriate" agreement percentages, especially for the Lewis study, is lower than expected here. By looking at column "D" under Code 20, it is clear that Carol is using more "20" codes in her second coding than she did in her first. Remember that some of the speakers in the Lewis study are adult males, and while their voices may sound low in comparison to other speakers, they are probably not abnormal for their age and gender. Therefore, Carol needs to exercise caution in the use of this code, especially when judging the voices of adult males. It's possible that there has been some drift in her perception of an abnormally low-pitched voice; referring back to the PVSP training tapes may help.

Laryngeal Features. The intrajudge agreement in this category needs improvement. In most cases it is even lower than the interjudge agreement for this category. Most of the agreements and disagreements are on the use of code 24. Based on the "D" columns, again, it appears that Carol used more inappropriate laryngeal codes the second time than she did the first time. Most of the within-code agreements were 23 and 24 or 25 and 24. Hopefully the new training tape (and the PVSP training tapes) will help to clarify the perception of true hoarseness and the components of breathiness and strain. Again, in the interjudge comparison it was found that Carol was using many more inappropriate laryngeal codes than the other transcriptionists. Becoming less strict in her judgment of inappropriate laryngeal quality, thereby using fewer inappropriate laryngeal codes, should help to improve intrajudge agreement as well as interjudge agreement in the future.

Resonance. There was no within-code agreement in this category, meaning that Carol either agreed exactly in each coding, or disagreed. As with inappropriate Rate, Stress, Loudness, Pitch, and Laryngeal codes, Carol used more inappropriate resonance codes in her second coding of the samples (See "D" columns under codes 30 and 31) than she did in her first coding. In the interjudge reliability discussion, it mentions that Carol actually used fewer inappropriate resonance codes than the other two transcriptionists (the transcripts compared were the first-time codings of the samples). In her second coding of the samples, Carol used more inappropriate resonance codes than Carmen did. Resonance is a difficult feature to judge; as usual, it seems that all that can be done to improve the reliability in this area is to listen to the training tape examples and the PVSP training tapes for assistance in making decisions regarding resonance.

Summary. In all categories but Phrasing, Carol used more inappropriate codes in her second coding than she did in her first coding. In general it seems she may be a little strict in her judgment of inappropriateness in these categories. Hopefully the above observations and guidelines, combined with those in the interjudge text, will help to clear up some of the discrepancies in coding from one time to the next.

PVSP Comparison: CW(Time 1) and CW(Time 2)

PHRASING

		1				2				3				4				5				6				7				8			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	191	--	--	--	1	2	0	2	4	2	0	1	0	2	0	0	0	0	0	0	5	4	4	2	0	0	0	0	0	0	1	1
Lewis	12	241	--	--	--	4	2	0	1	10	0	3	1	0	1	0	0	1	0	0	0	3	2	3	4	0	0	0	0	0	1	0	0
Total	22	432	--	--	--	5	4	0	3	14	2	3	2	0	3	0	0	1	0	0	0	8	6	7	6	0	0	0	0	0	1	1	1

PHRASING: Exact Agreement with Appropriate: Iowa1= 90.5% Lewis= 93.5% All= 92.2%
 Exact Agreement on Inappropriate Only: Iowa1= 32.3% Lewis= 50.0% All= 41.8%
 Within-Code Agreement: Iowa1= 64.5% Lewis= 66.7% All= 65.7%

RATE

		1				9				10				11				12			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	202	--	--	--	1	0	0	2	0	2	3	1	2	0	2	6	0	0	0	1
Lewis	12	250	--	--	--	0	0	0	0	0	0	0	3	2	0	2	17	0	1	1	1
Total	22	452	--	--	--	1	0	0	2	0	2	3	4	4	0	4	23	0	1	1	2

RATE: Exact Agreement with Appropriate: Iowa1= 92.3% Lewis= 91.0% All= 91.6%
 Exact Agreement on Inappropriate Only: Iowa1= 15.0% Lewis= 7.4% All= 10.6%
 Within-Code Agreement: Iowa1= 25.0% Lewis= 11.1% All= 17.0%

STRESS

		1				13				14				15				16			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	168	--	--	--	0	0	0	0	1	0	0	18	6	4	5	19	0	0	0	1
Lewis	12	212	--	--	--	0	0	0	0	0	0	0	21	13	2	5	24	0	0	0	0
Total	22	380	--	--	--	0	0	0	0	1	0	0	39	19	6	10	43	0	0	0	1

STRESS: Exact Agreement with Appropriate: Iowa1= 78.8% Lewis= 81.2% All= 80.2%
 Exact Agreement on Inappropriate Only: Iowa1= 12.9% Lewis= 20.0% All= 16.8%
 Within-Code Agreement: Iowa1= 20.4% Lewis= 23.1% All= 21.8%

LOUDNESS

		1				17				18			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	133	--	--	--	28	0	3	40	8	0	2	8
Lewis	12	169	--	--	--	24	0	11	39	7	0	1	26
Total	22	302	--	--	--	52	0	14	79	15	0	3	34

LOUDNESS: Exact Agreement with Appropriate: Iowa1= 76.1% Lewis= 72.2% All= 73.9%
 Exact Agreement on Inappropriate Only: Iowa1= 40.4% Lewis= 28.7% All= 34.0%
 Within-Code Agreement: Iowa1= 40.4% Lewis= 28.7% All= 34.0%

PITCH

		1				19				20				21				22			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	172	--	--	--	3	0	1	3	8	0	11	24	0	0	0	0	0	0	0	0
Lewis	12	169	--	--	--	1	4	0	1	10	1	0	91	0	0	0	0	0	0	0	0
Total	22	341	--	--	--	4	4	1	4	18	1	11	115	0	0	0	0	0	0	0	0

PITCH: Exact Agreement with Appropriate: Iowa1= 82.4% Lewis= 65.0% All= 72.7%

Exact Agreement on Inappropriate Only: Iowa1= 22.0% Lewis= 10.2% All= 13.9%

Within-Code Agreement: Iowa1= 22.0% Lewis= 14.8% All= 17.1%

LARYNGEAL

		1				23				24				25				26				27				28				29				
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D					
Iowa1	10	50	--	--	--	3	20	0	8	56	5	2	33	5	17	0	6	3	9	1	0	0	0	0	0	0	0	0	0	0	0	4	0	0
Lewis	12	75	--	--	--	16	20	2	5	64	2	3	76	1	0	0	2	1	2	1	5	0	0	0	0	0	0	0	0	0	2	0	0	
Total	22	125	--	--	--	19	40	2	13	120	7	5	109	6	17	0	8	4	11	2	5	0	0	0	0	0	0	0	0	6	0	0		

LARYNGEAL: Exact Agreement with Appropriate: Iowa1= 52.7% Lewis= 56.7% All= 54.9%

Exact Agreement on Inappropriate Only: Iowa1= 39.0% Lewis= 40.6% All= 39.8%

Within-Code Agreement: Iowa1= 70.9% Lewis= 53.5% All= 61.5%

RESONANCE

		1				30				31				32			
Study	#T	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Iowa1	10	176	--	--	--	11	0	2	17	5	0	0	11	0	0	0	0
Lewis	12	220	--	--	--	2	0	1	4	29	0	1	20	0	0	0	0
Total	22	396	--	--	--	13	0	3	21	34	0	1	31	0	0	0	0

RESONANCE: Exact Agreement with Appropriate: Iowa1= 86.5% Lewis= 90.6% All= 88.8%

Exact Agreement on Inappropriate Only: Iowa1= 34.8% Lewis= 54.4% All= 45.6%

Within-Code Agreement: Iowa1= 34.8% Lewis= 54.4% All= 45.6%