DOCUMENT RESUME

ED 088 287 FL 005 219

AUTHOR Lea, Wayne A.
TITLE Perceived Stress as the "Standard" for Judging Acoustical Correlates of Stress.
PUB DATE 1 Nov 73
NOTE 12p.; Paper presented at the Meeting of the Acoustical Society of America (86th, Los Angeles, California, November 1, 1973)

EDRS PRICE MF-$0.75 HC-$1.50
DESCRIPTORS Acoustic Phonetics; Articulation (Speech); Auditory Discrimination; Auditory Perception; Connected Discourse; *Distinctive Features; Graphs; Language Patterns; *Language Research; Oral Expression; Oral Reading; Phonetics; *Phonology; Standard Error of Measurement; *Standards; *Syllables; Tape Recordings; Written Language

ABSTRACT

Acoustical correlates of stress can only be evaluated in comparison with some "standard" specifying which syllables are actually stressed. The Standard should be consistent from time to time, and largely independent of talker and listener idiosyncrasies. Three phonetically-trained subjects listened to repeatedly spoken texts and spontaneous sentences until they could categorize each syllable as either stressed, unstressed, or reduced. This procedure was repeated three times for each speech text and listener. Two listeners differed from each other on only 5% of all syllables as to whether they were perceived as stressed or not. Each showed about 5% confusion in decisions about stressed syllables from one trial to another. Unstressed and reduced levels were confused more frequently. The third listener gave less consistent results. Subjects' judgements of stress when given only the written text were of comparable consistency but did not correspond well with perceptions with speech, if the speech was spontaneous rather than spoken texts. Stress perceptions consequently may be suitable for evaluating acoustical correlates to within a 5% tolerance in overall location scores. Pooling the perceptions from several trials and several listeners may improve the stability of this "standard" for stress assignment. (Author/DD)

PERCEIVED STRESS AS THE "STANDARD" FOR JUDGING ACOUSTICAL CORRELATES OF STRESS

Wayne A. Lea
Sperry Univac DSD
P.O. Box 3525
St. Paul, Minn. 55165

ABSTRACT

Acoustical correlates of stress can only be evaluated in comparison with some "standard" specifying which syllables are actually stressed. The standard should be consistent from time to time, and largely independent of talker and listener idiosyncrasies. Three phonetically-trained subjects listened repeatedly to spoken texts and spontaneous sentences, until they could categorize each syllable as either stressed, unstressed, or reduced. This procedure was repeated three times for each speech text and listener. Two listeners differed from each other on only 5% of all syllables as to whether they were preceived as stressed or not. Each also showed only about 5% confusions in decisions about stressed syllables from one trial to another. Unstressed and reduced levels were much more frequently confused. The third listener gave less consistent results. Subjects' judgments of stress when given only the written text were of comparable consistency, but did not correspond well with perceptions with speech, if the speech was spontaneous rather than spoken texts. Stress perceptions consequently may be suitable for evaluating acoustical correlates to within a 5% tolerance in overall location scores. Pooling the perceptions from several trials and several listeners may improve the stability of this "standard" for stress assignment.

Presented at the 86th Meeting, Acoustical Society of America, Los Angeles, California, November 1, 1973. This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC 15-73-C-0310, ARPA Order No. 2010.

PERCEIVED STRESS AS THE "STANDARD" FOR JUDGING ACOUSTICAL CORRELATES OF STRESS

Wayne A. Lea

Acoustical correlates of stress can only be evaluated in comparison with some "standard" specifying which syllables are actually stressed. For studies of isolated words, such as minimal pairs of noun versus verb, a desk dictionary or a researcher's own intuitions may be sufficient. However, for studies of the stress patterns throughout sentences and discourses, that "standard" for stress assignment is not as readily established. I will report here on some experiments regarding the effectiveness and stability of listener's perceptions of stressed, unstressed, and reduced syllables in continuous speech.

The procedure[1] used in the present study was to have an individual repeatedly hear tape recordings, through earphones, and mark, for each syllable, whether he heard that syllable as stressed, unstressed, or reduced. The listener could listen to portions of the tape as often as necessary, until he could mark each syllable. He was free to back up the tape at his choice, and no time limits or procedural constraints were placed on him. The listeners did endeavor to rewind far enough to always hear an entire clause or sentence, to have a constant context within which to judge relative stress levels.

Slide one illustrates the method for recording a listener's perceptions. To facilitate marking for each syllable, the script of each recording was typed on a sheet of paper, with vertical slashes between syllables. The listener received one such sheet for each recorded text, and a mark (such as S, U, or R) was required for each syllable.

Three phonetically-trained listeners were used in this study. An earlier study showed that two of these listeners gave similar stress perceptions to those of four other listeners used in experiments previously reported on by Li, Hughes and Snow. Each listener repeated the perception test at least three times, to determine listener

---

[1] Li, K.-P., Hughes, G. W., and Snow, T. B. (1973), Segment Clascification in Continuous Speech, IEEE Trans. on Audio and Electroacoustics, Vol. AU-21, No. 1, pp. 50-57.

consistency from one time to another. The listeners were also asked to report their stress judgments when given only the written text (with no tape recordings). These judgments with no speech were also obtained in three repetitions, to test their repeatability.

Speech texts used in this study included a paragraph of the Rainbow Script read by six talkers, a script composed only of monosyllabic words read by two talkers, 31 spontaneous sentences intended for man-computer interaction, which involved recordings by ten different talkers at several contractors within the ARPA Speech Understanding Research program. With the several repetitions by several listeners, this yielded over 17,000 judgments of stress levels for syllables in connected texts spoken by sixteen talkers.

In the next slide, we see plots of majority votes about the stress level for each syllable in several portions of texts. The majority vote from a listener's three repetitions of the listening test was first found. For example, on two trials he may perceive the work "strikes" as stressed, while on the third he hears that syllable as unstressed. His majority vote is then stressed. Then the results for all three listeners were pooled, by plotting a stress score as the number of listeners whose majority vote says the syllable is stressed, minus the number whose majority vote says the syllable is reduced. Unstressed judgments were assigned a value of zero. Thus, a plus 2 score for syllables like the word "strikes" indicates that two of the listeners heard the syllable as stressed, while the third listener perceived that syllable as unstressed. A companion study, reported on in another paper at this meeting (Lea, 1973), showed that about 85% of the syllables perceived as stressed by two or more listeners (that is, those which had a stress score of +2 or +3) were correctly found by an algorithm for locating stressed syllables from acoustic data.

Listeners obviously did not always agree about the stress level of a syllable. The next slide shows plotted, for each pair of listeners and each text, the percentages of majority stress judgments that differ from one listener to another. Listeners MFM and TES disagree about the stress levels they assign to about 50% to 60% of all syllables in each of the texts. The percentages of listener-to-listener confusions are not drastically affected by the talker or text. Even the percentages

of confusions with NO speech don't differ much from those with speech. (I should emphasize at this point that listener TES is quite unusual; most pairs of listeners have exhibited more like the 20 to 30% confusions between listeners WAL and MFM.)

In the next slide, the confusions between stressed and unstressed levels of perception have been separated from the confusions between unstressed and reduced syllables, for listeners WAL vs MFM. About 5% of all syllables are confused between stressed and unstressed by the two listeners WAL vs MFM, as shown by the cross-hatched bars, while 15 to 25% of all syllables were confused between unstressed and reduced categories, as shown by the blank bars. Thus, these two listeners agree quite well about which are the stressed syllables, while they do not as consistently agree about which are the reduced syllables.

How a listener's perceptions differ from time to time is shown in the next slide. As shown by the cross-hatched bars, listener MFM confused about 1 to 5% of the syllables between levels of stressed and unstressed from one trial to another. His confusions between unstressed and reduced levels were much more frequent.

The next slide shows that confusion between the majority judgments of a listener with speech and his majority judgments without the speech were more frequent if the speech was spontaneously spoken, such as the ARPA sentences were. Particularly for spontaneous speech, then, stress locations from acoustical correlates can be judged more reliably from stress perceptions obtained with speech recordings than from simple judgments based only on orthographic transcriptions.

The 31 ARPA Sentences involve declarative sentences, commands, questions requiring yes/no answers, and questions with interrogative WH-words (who, where, which, what). The next slide shows that confusions (from trial to trial) were more frequent in questions than in declaratives or commands, with yes/no questions yielding the most confusions, and declaratives yielding the fewest confusions.

We may conclude from these studies that while the stress perception methods used here are generally quite consistent from time to time and listener to listener, they will not consistently judge the effectiveness of stressed syllable location from acoustic data to any precision better than about 5% tolerance. Then, if a stressed syllable location algorithm could located 95% of all syllables perceived as stressed by majority votes of two or more listeners, it would be doing as well as one repetition of the perception tests would do for predicting the perceptions from a second repetition of the experiment. It would also be doing as well as one listener would do in comparison to another listener. Our "standard" thus has on the order of a 5% tolerance and, when using this standard, we can demand no better precision in stressed syllable location from acoustic data.

STRESS PERCEPTIONS ON ARPA SENTENCES

Listener MFM Date 2/20/73

LS21:

S	U	R	S	U	R	S	R	U	S
Who	is	the	own	er	of	utt	er	ance	eight?

LM13:

U	S	U	U	S	U	S	U	U	S	U	S	U	U
Dis	play	the	pho	ne	mic	la	bels	a	bove	the	spec	tro	gram.

B27:

U	U	R	S	U	U	S	S	U
Do	an	y	sam	ples	con	tain	troi	lite?

B10:

S	U	R	S	R	U	U	S	U	U	S	S	U	U	U	R	S	R	S	U
What	is	the	av	(er)	age	u	ra	ni	um	lead	ra	tio	for	the	lu	nar	sam	ples?	

RB6:

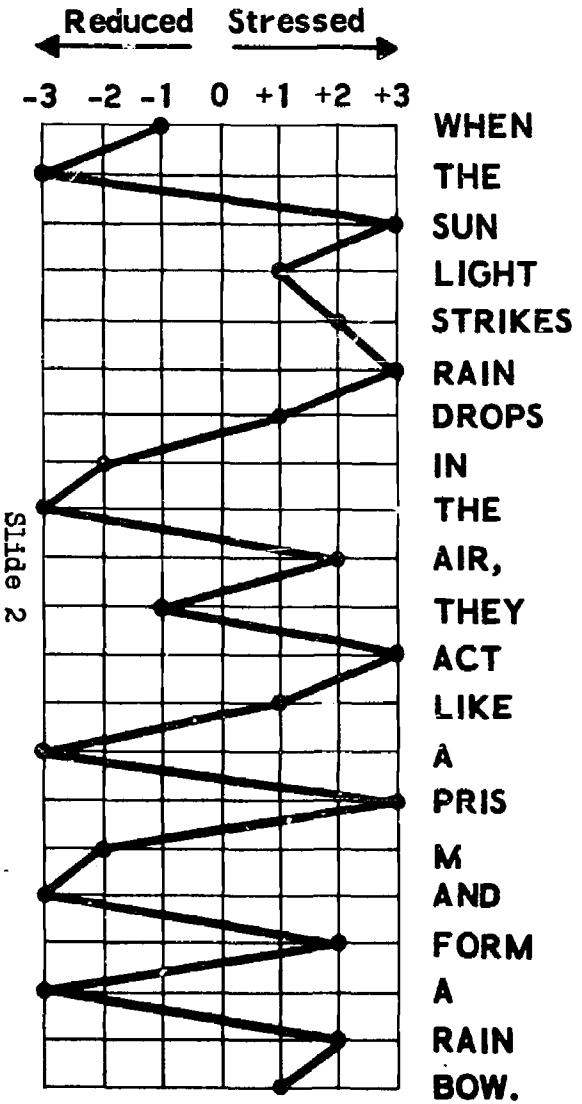
S	S	S	U	R	S	S	S	R	U
Do	you	have	an	y	right	square	box	es	left?

RB16:

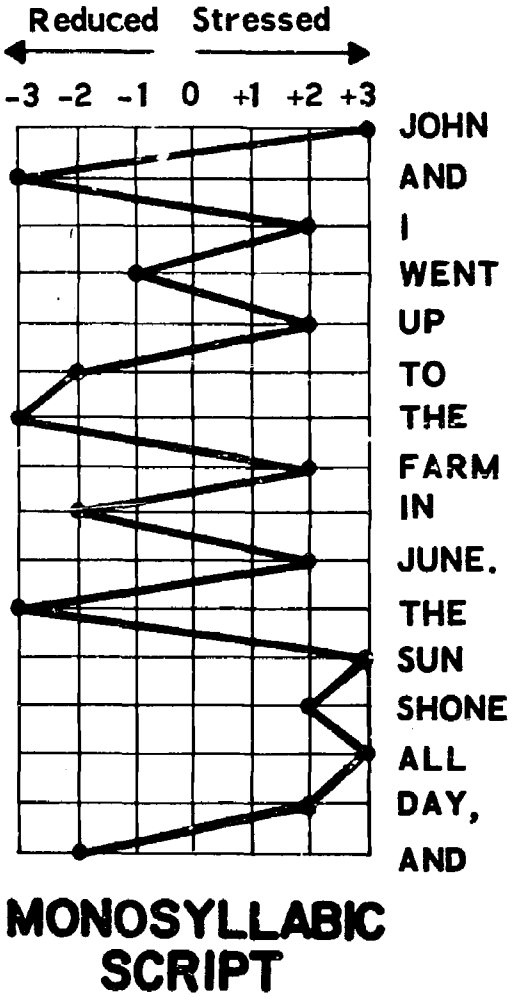
S	U	S	U	S	S	S	U	S	S
Put	the	oth	er	red	block	on	the	red	block.

Slide 1
5

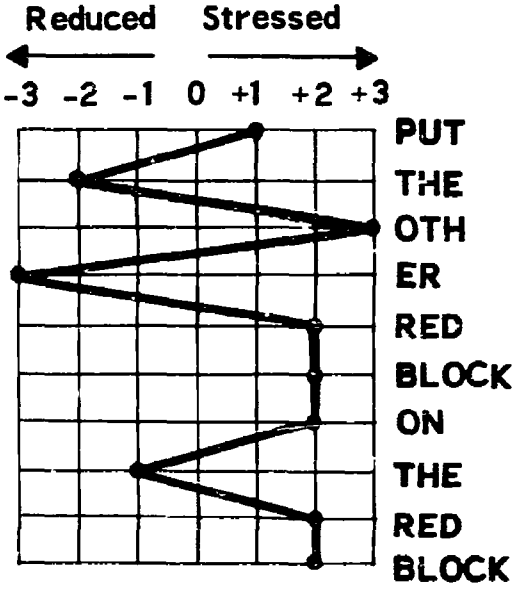
STRESS SCORES OBTAINED FROM PERCEPTIONS OF THREE LISTENERS



RAINBOW SCRIPT



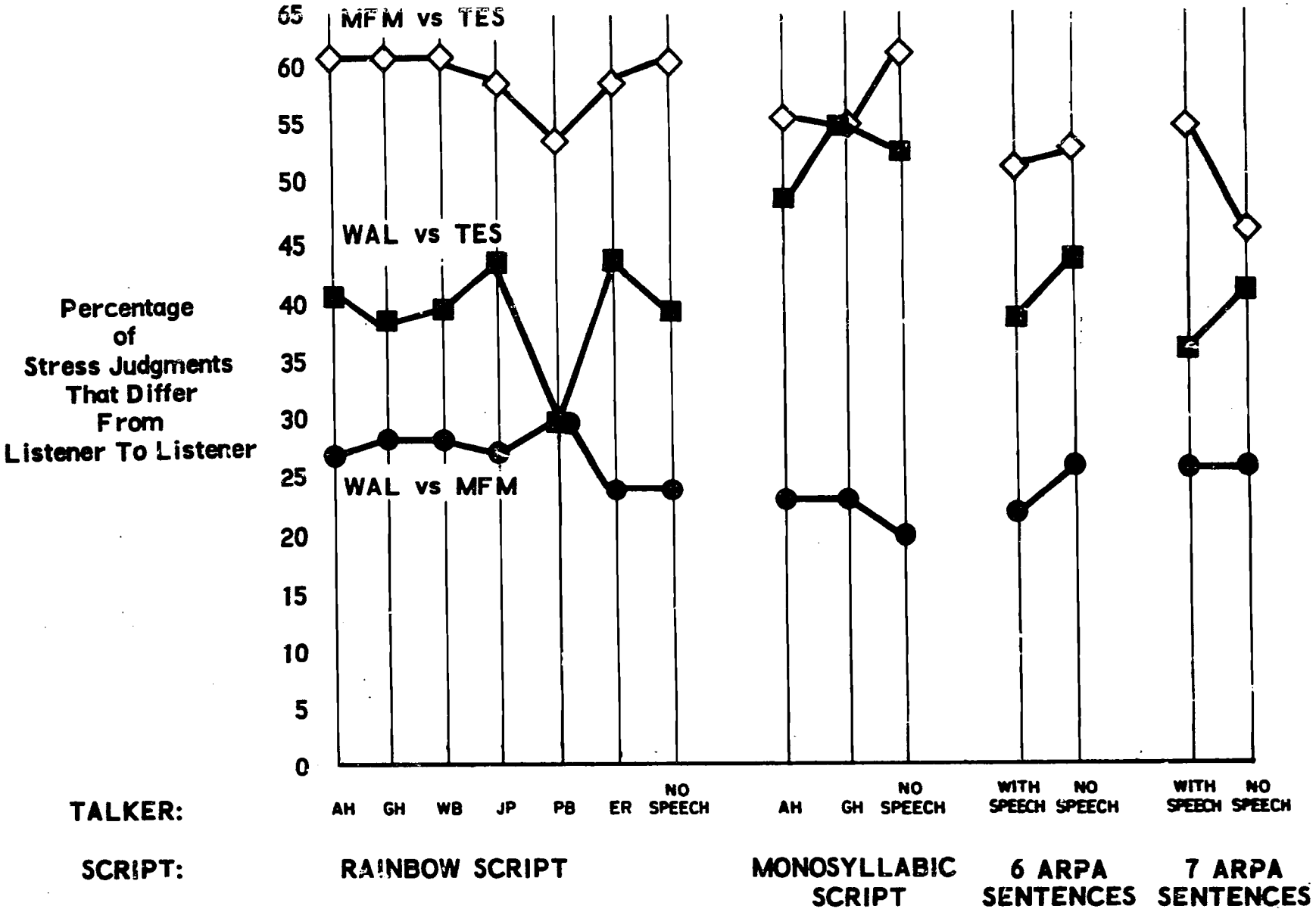
MONOSYLLABIC SCRIPT

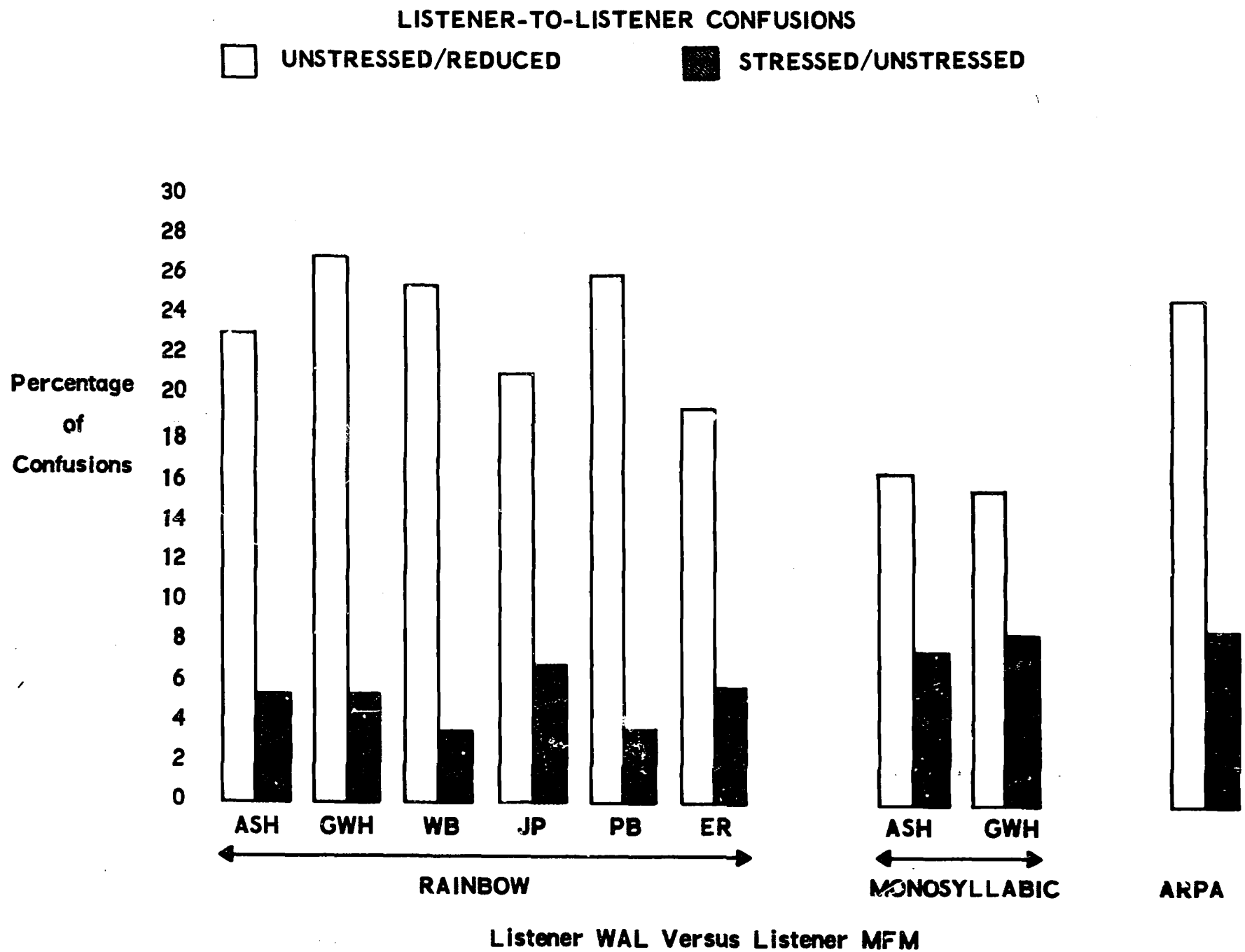


ARPA SENTENCES

Slide 2

LISTENER-TO-LISTENER CONFUSIONS



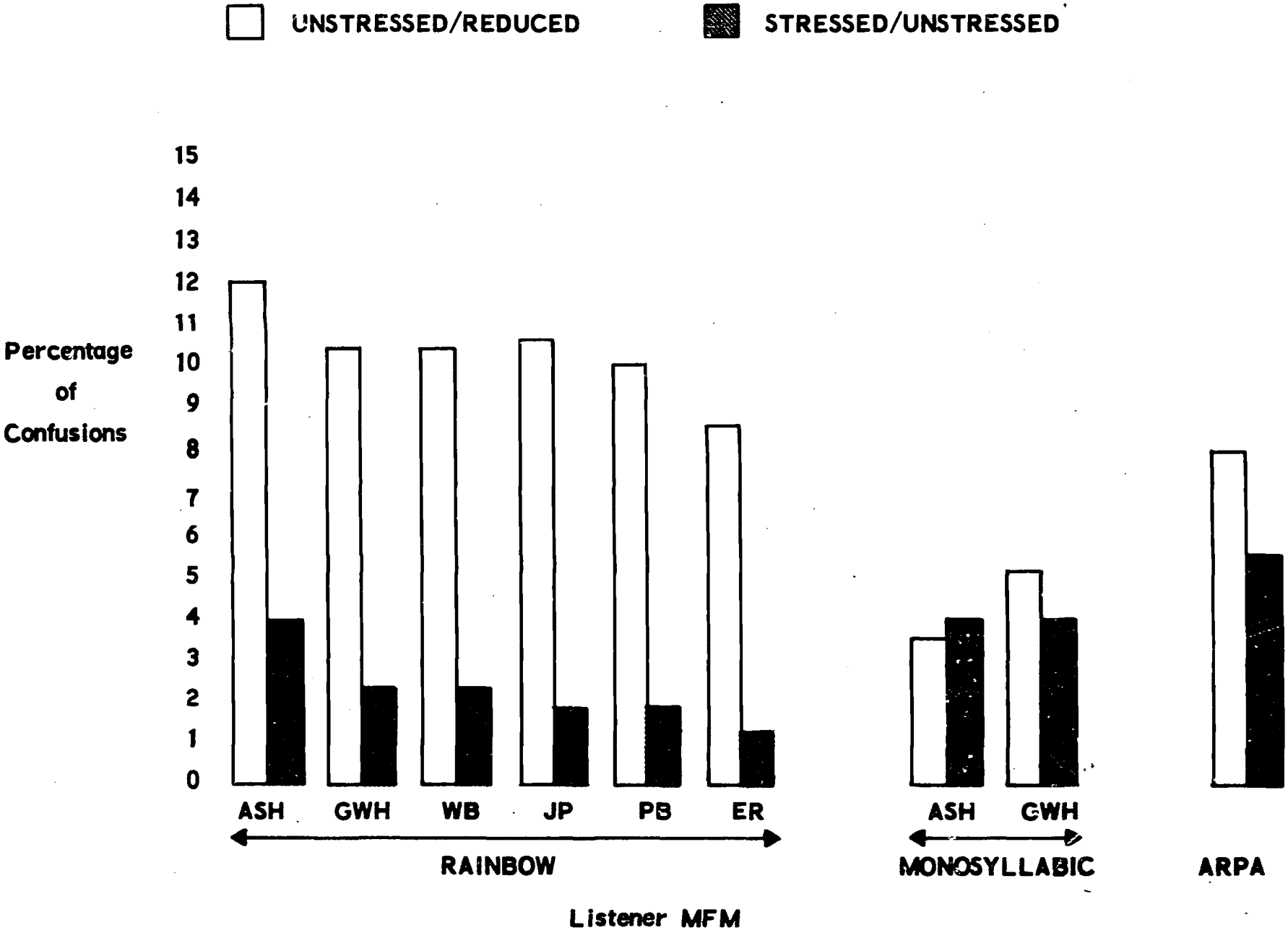


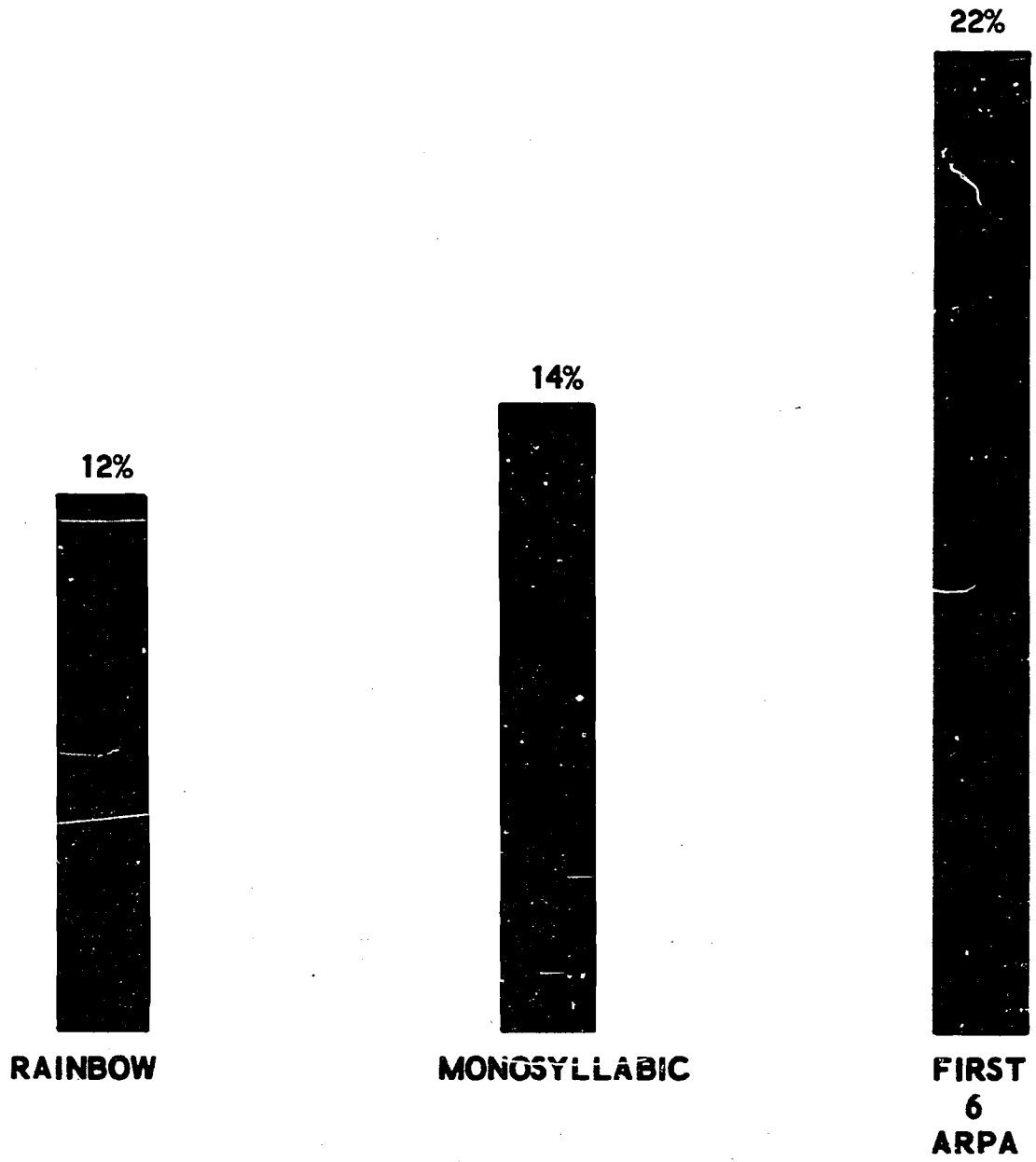
Slide 4
8

Listener WAL Versus Listener MFM

Slide 5
9

REPETITION-TO-REPETITION CONFUSIONS





CONFUSIONS BETWEEN NO-SPEECH AND SPEECH CONDITIONS
(Results Pooled for All Talkers and Listeners)

Slide 7

PERCENTAGES OF TOTAL REPETITION-TO-REPETITION CONFUSIONS
FOR VARIOUS TYPES OF SENTENCES
(Pooled for All 3 Listeners)