EFFECTS OF UNNATURAL PAUSE ON SPEECH INTELLIGIBILITY

Sumiko Takayanagi, Jody Kreiman, and Donald D. Dirks
Division of Head and Neck Surgery, UCLA School of Medicine, Los Angeles, USA

ABSTRACT

Effects of unnatural pausing on spoken sentence intelligibility by native and non-native listeners of American English were studied in a framework of 6 different signal-to-noise ratios. Performance between pause and non-pause sentences differed significantly for native listeners at 0 dB and –2 dB signal-to-noise ratios, but no significant difference was obtained from non-native listeners. The temporal speech flow distortion of prosodic cues negatively affected auditory processing for the population who can fully utilize these cues, whereas smaller declines were shown in listeners who have less ability to exploit them. Pauses appear to provide distinctive speech information and serve as perceptual units that are critical to understanding spoken language. Further implications of suprasegmental cues for foreign language acquisition and speech training are discussed.

1. INTRODUCTION

Prosodic information has a significant role in human speech communication. Many types of suprasegmental cues, such as pausing, stress, intonation, and rhythmic variations, contribute to speech perception by helping listeners to disambiguate the meaning of the speech without changing the segmental information [1-3].

Speech communication is an interactive process in which the talker's forthcoming acoustic information should match the listener's expectations in order to be understood, and any deviations on either the segmental or suprasegmental levels of speech affect speech intelligibility [1,4]. Native speakers are generally aware of the most prosodic cues, and can differentiate meanings of the spoken utterance based upon the prosodic information given in the situation [1]. Typically, they can reject utterances which have rhythmic mismatches as created by differential boundary patterns [5].

Extensive research has already confirmed that pauses influence speech comprehension and are a distinct part of the information needed to obtain correct interpretations of speech [6-9]. Even infants as young as 7 months showed a sensitivity to the prosodic appropriateness of pause locations when both natural and unnatural pauses were inserted into speech [10], and most adults strongly preferred pauses that separated constituents, rather than inappropriately interrupting them [11]. A cross-linguistic study of native speakers of American English and German found a significant advantage for correct pause detection in the native language. However, German-speaking subjects who also spoke English performed the task on the English passage as well as the American subjects, while subjects who spoke only English showed a significant difference between the German and the English passages [12]. Several studies have shown that the addition of interword pauses improves the intelligibility of speech samples of dysarthric patients and deaf children [13-15]. However, such improvements did not occur in other research when only syntactically appropriate pauses were used [16].

The present study investigated the effects of prosodically and syntactically inappropriate pauses on speech intelligibility. Because listeners received identical segmental information with or without pauses, variations in speech intelligibility can be attributed solely to mismatches between perceptual strategies and the available acoustic information at suprasegmental level. Sentences were presented to subjects at six signal-to-noise levels in order to construct performance functions for each condition. Additionally, native versus non-native listener status was included as a factor in order to compare the effects of inappropriate pauses on these two groups.

2. METHODS

2.1. Subjects: Prior to the experimental session, the auditory thresholds (at octave frequencies from 250 to 8000 Hz) of all subjects were tested, and the subjects who had pure tone threshold of 15 dB hearing level or better were included in this study. They were 20 UCLA students and staff, and their average age was 31. Ten listeners were native speakers of American English, and 10 were non-native speakers who had intermediate to advanced levels of English listening comprehension skills in their self-reports. There were two native speakers of each of the following languages: Chinese, Italian, Japanese, Korean, and Spanish.

2.2. Stimuli Construction: A copy of the original recording of sentences from the HINT (Hearing In Noise Test) [17] was used as the speech material. The original test contained 250 sentences equated for 50% correct performance. The background noise is a pseudo random noise shaped to match the frequency response characteristics of the average sentences.

A subset of 120 experimental and 6 practice sentences that contained a potential unnatural boundary for a pause insertion were chosen from the original HINT sentences as experimental stimuli. The sentences were digitized via a 16 bit analog-to-digital converter operating at a sampling rate of 48kHz. The sentences were edited by using programs from a Kay Elemetric Computer Speech Laboratory System. A 500 msec. pause was inserted into a syntactically and prosodically inappropriate interword boundary position in each sentence. The pauses were inserted into 43 subject noun phrases, 61 object noun phrases, and 16 verb phrases.

2.3. Design and apparatus: Two sets of experimental sentences were created by interchanging pause and non-pause sentences in order to minimize any differences in intelligibility due to specific sentence. A total of 120 sentences, (60 with pauses and 60 without pauses), were randomized and grouped into 12 blocks. Each block had equal numbers of pause and non-pause sentences. Signal-to-noise levels were presented in a random

order within each block. Sentences were played from a DAT recorder, with speech delivered to one channel of an attenuator-amplifier complex and noise to a second channel. Noise and speech were mixed and presented monaurally via earphones. The noise preceded the speech by approximately 300 msec. and acted as an alerting signal. The noise also continued for 300 msec. after the termination of the speech signal. Sentences were separated by 4-5 seconds. The level of the speech was specified with a calibration tone which matched the long term RMS level of the averaged speech. The sound pressure level (SPL) of the output of the earphones was measured in a NBS-9A 6-cm^3 coupler with a Larson-Davis sound level meter. The speech was presented at six signal-to-noise ratios from –6 dB to +4 dB in two dB steps. This range provided estimates of intelligibility that encompassed scores from 0 to 100% for normal hearing listeners.

2.4. Procedures: Stimuli were presented at 66.5 dB SPL to a subject seated in a double-walled sound attenuation chamber. The noise was varied to obtain the experimental signal-to-noise ratios. The subjects were instructed to repeat each sentence. Sentences were scored as correct if the whole sentence was repeated correctly. Approximately one hour was required to test each subject.

3. RESULTS

The percentage correct mean scores and the standard deviations in pause and non-pause sentences for native and non-native subjects at each signal-to-noise ratio is shown in Table 1 and Figure 1. Overall performances are increased significantly with signal-to-noise ratio (repeated measures ANOVA: native-nonpause $F=641.71$, $p<.0001$; pause $F=156.58$, $p<.0001$; non-native-non pause F-138.28, $p<.0001$; pause $F=27.96$, $p=.001$). For both pause and non-pause conditions, performance of native subjects was higher than for non-native subjects (between-subject ANOVA: $F=101.58$, $p<.0001$). A Tukey's HSD post hoc multiple comparison test showed that performance on pause sentences differed significantly from non-pause sentences for native listeners at –2 dB and 0 dB levels, but no such differences were observed for non-native listeners. The data were arcsine transformed to stabilize the variance, and the analysis was repeated with the same results.

4. DISCUSSION

Effects of unnatural pausing on speech intelligibility were investigated for native and non-native listeners of American English at 6 different signal-to-noise ratios. Inappropriate pauses reduced speech intelligibility in both groups of subjects, except at the ceiling and floor levels. This study supports the idea that pauses are distinctive suprasegmental units and that their appropriateness in location and duration are critical for speech intelligibility.

As expected, overall performance level of native listeners was significantly higher than non-native listeners for both sentences with or without pauses. This suggests that processing abilities of non-native speech materials for non-native listeners is substantially lower than native listeners. Similarly, Black [18] found that non-natives were especially susceptible to noise that reduced speech cues substantially and impaired communication processing.

The key finding of this experiment was the significant performance difference between pause and non-pause sentences in native listeners and its absence in non-native listeners (see Figure 1). Since durational patterns provide cues to speech recognition [5,19], and pausal phenomena within an individual's speech pattern are relatively stable over different conditions [7], native listeners' perceptual strategies in terms of durational patterns can be presumed to be reasonably consistent. This suggests that distortion of prosodic cues negatively influences auditory processing for the native listener population who can fully utilize these cues, whereas a smaller decline was shown in listeners who have less ability to exploit such information. Thus, inconsistencies between signals and listeners expectations may have affected speech intelligibility. Non-native listeners, however, do not have such expectations, and they are less sensitive to the subtleties of speech flow, so only a modest decline in performance was observed [12]. Accordingly, native listeners' compensatory tactics for recovering speech in the unnatural pause sentences were not adequate when masking noise became intense enough to make the speech recognition task difficult.

Although the present study examined the effects of a single prosodic cue, suprasegmental corrections might improve speech intelligibility for populations who need speech modifications, such as dysarthric and non-native speakers [20]. Suprasegmental training as well as correct production of segmental contrasts should be emphasized in foreign language education and speech rehabilitative training.

ACKNOWLEDGMENTS

We thank Dr. Patricia Keating for her insightful suggestions, and advice for token selections and development. This work was supported by NIH training grant 5-T32-DC-00029-07 to the Department of Linguistics, UCLA, and also by Veterans Administration Rehabilitation Research and Development Grants RCTR 597-0160 and C2225.

REFERENCES

[1] Price, P. J., Ostendorf, M, Shattuck-Hufnagel, S. and Fong, C. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustic Society of America, 90*, 6, 2956-2970.

[2] Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. and Price, P. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America, 91*, 3, 1707-1717.

[3] Shattuck-Hufnagel, S. and Turk, A. E. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*, 2, 193-247.

[4] Liss, J. M. and Spitzer, S. 1998. Syllabic strength and lexical boundary decision in the perception of hypokinetic dysarthric speech. *Journal of the Acoustical Society of America, 104*, 4, 2457-2466.

[5] Smith, M. R., Cutler, A., Butterfield, S. and Nimmo-Smith, I. 1989. The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech and Hearing Research, 32*, 912-920.

[6] Osgood, C. D. 1954. Psycholinguistics, (Ed.), *Journal of Abnormal Social Psychology, 49*, Supplement, 4, part 2.

[7] Goldman-Eisler, F. 1951. The measurement of time sequences in conversational behavior. *British Journal of Psychology, 42*, 355-362.

[8] Suci, G. J. 1967. The validity of pause as an index of units in language. *Journal of Verbal Learning and Verbal Behavior, 6*, 26-32.

[9] Gee, J. P. and Grosjean, F. 1983. Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15*, 411-458.

[10] Hirsh-Pasek, K., Nelson, D. G. K., Jusczyk, P., W., Cassidy, K., W., Druss, B. and Kennedy, L. 1987. Clauses are perceptual units for young infants. *Cognition, 26*, 269-286.

[11] Pilon, R. 1981. Segmentation of speech in a foreign language. *Journal of Psycholinguistic Research, 10*, 2, 113-122.

[12] Shuckenberg, A. and O'Connell, D. D. 1988. The long and short of it: Reports of pause occurrence and duration in speech. *Journal of Psycholinguistic Research, 17*, 1, 19-28.

[13] Maassen, B. 1986. Marking word boundaries to improve the intelligibility of the speech of the deaf. *Journal of Speech and Hearing Research, 29*, 227-230.

[14] Gutek, J. M. and Rochet, A. P. 1996. Effects of insertion of interword pauses on the intelligibility of Dysarthric Speech. In Robin, D. A., Yorkston, K M., and Beukelman, D. R. (Eds.), *Disorders of Motor Speech: Assessment, Treatment, and Clinical Characterization.* Baltimore, Maryland: Paul H. Brookes Publishing Company.

[15] Yorkston, K. M. and Beukelman, D. R. 1981. Ataxic dysarthria: Treatment sequences based on intelligibility and prosodic considerations. *Journal of Speech and Hearing Disorders, 46*, 398-404.

[16] Hammen, V. L., Yorkston, K. M. and Minifie, F. D. 1994. Effects of temporal alterations on speech intelligibility in parkinsonian dysrthria. *Journal of Speech and Hearing Research, 37*, 244-253.

[17] Nilsson, M., Soli, S. D. and Sullivan, J. A. 1994. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95*, 2, 1085-1099.

[18] Black, J. W. 1964. Language varriers and language traning. *Communication Processes: Proceedings of a Symposium* held in Washington, Pergamon Press, Oxford, 101-165.

[19] Nakatani, L. H. and Schaffer, J. A. 1978. Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America, 63*, 234-245.

[20] Maassen, B. and Povel, D. 1985. The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech. *Journal of the Acoustic Society of America, 78*, 3, 887-886.

	Native Listeners			Non-Native Listeners		
	S/N Ratios	Mean	Standard Deviation	S/N Ratios	Mean	Standard Deviation
Non-Pause Sentences	-6 dB	4	5.16	-6 dB	0	0
	-4 dB	17	14.18	-4 dB	8	9.19
	-2 dB	55	14.34	-2 dB	22	15.49
	0 dB	85	8.50	0 dB	40	23.09
	2 dB	92	7.89	2 dB	55	26.35
	4 dB	97	4.83	4 dB	72	14.75
Pause Sentences	-6 dB	2	4.22	-6 dB	0	0
	-4 dB	14	9.66	-4 dB	5	8.50
	-2 dB	37	19.47	-2 dB	15	15.09
	0 dB	61	17.29	0 dB	29	18.53
	2 dB	83	14.18	2 dB	48	22.01
	4 dB	90	11.55	4 dB	67	17.03

Table 1. Percent correct scores for HINT sentences at six signal-to-noise ratios.
Results are shown for native and non-native listeners and for sentences with and without unnatural pauses.

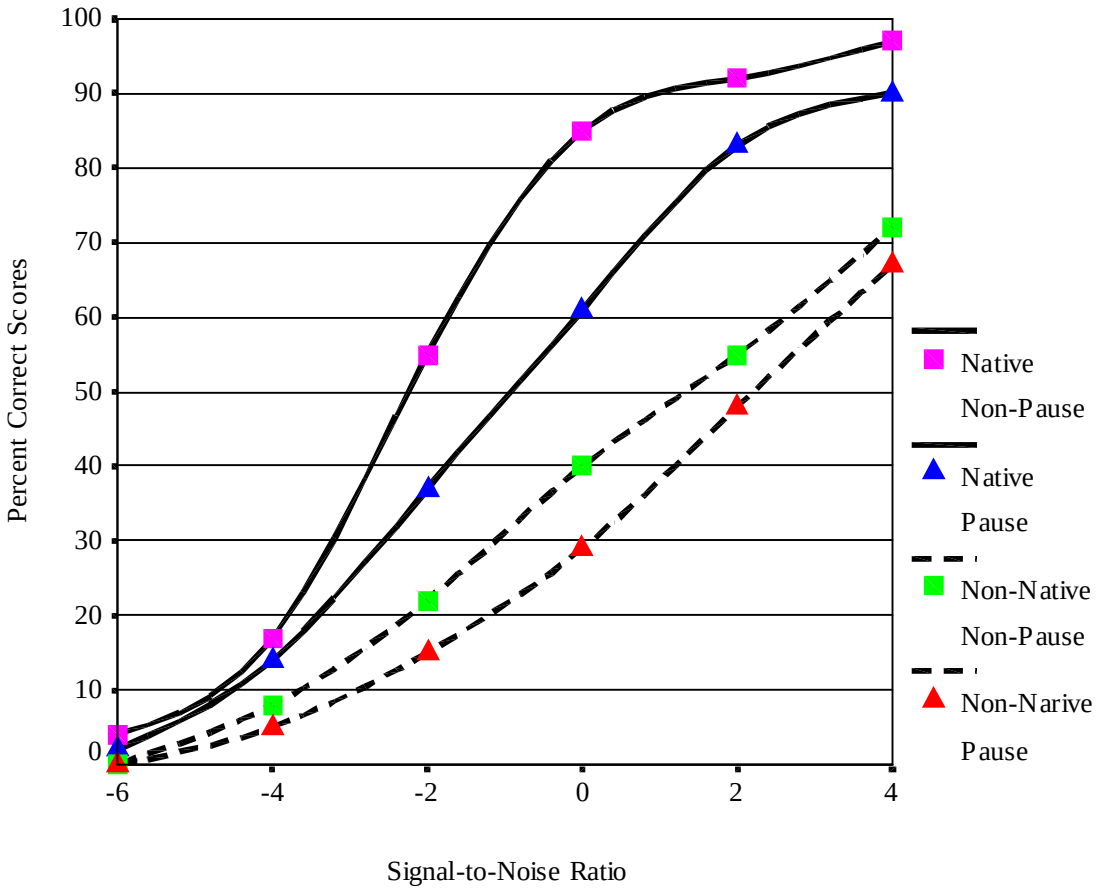

Figure 1. Performance functions for native and non-native listeners with pause and non-pause sentences