

Word intelligibility decrements and the comprehension of time-compressed speech

GARY W. HEIMAN, RAPHAEL J. LEO, and GLENN LEIGHBODY
State University College, Buffalo, New York

and

KATHLEEN BOWLER
University of Ottawa, Ottawa, Ontario, Canada

The extent to which decreased comprehension of time-compressed messages results from decreased word intelligibility was investigated. Experiment 1, in which 500-word messages were temporally interrupted, demonstrated that when 60% of the signal is deleted, comprehension is reduced even without the temporal limitations of compression. Experiment 2, which employed a backward masking procedure, demonstrated that individually compressed words are less intelligible when presented within the limited time constraints of a compressed message. It was concluded that decreased comprehension of compressed messages is the result of decreased word intelligibility.

In producing time-compressed speech, a recorded message is first temporally interrupted, resulting in a signal that is rapidly turned on and off, with "off-periods" of approximately 30 msec. The off-periods are deleted and the remaining on-periods are abutted to form a continuous signal. The rate of compression equals the percentage of the signal deleted during interruption, which is determined by the frequency of the off-periods. The final word rate in a compressed message is also a function of the original spoken word rate of the message. For example, a compressed message of 250 words per minute (wpm) can be produced by 50% compression of a message recorded at 125 wpm or by 33% compression of a message recorded at 188 wpm.

Comprehension of a compressed message is measured as the number of correctly answered questions concerning the contents of a lengthy passage. In an extensive review of the literature, Foulke and Sticht (1969) concluded that as word rate is increased from 175 wpm to approximately 275 wpm, comprehension declines at a moderate rate and that with increases beyond 275 wpm, comprehension declines at an accelerated rate. On the other hand, word intelligibility (the accuracy with which a word can be reported after it is presented in isolation) does not drop below 90% with less than 50% compression (e.g., Garvey, 1953). Foulke and Sticht concluded that the rapid decline in comprehension with rates beyond 275 wpm was not attributable to intelligibility decrements alone, but rather to some type of processing overload in short-term memory (STM).

Foulke and Sticht (1969) minimized the importance of intelligibility decrements by arguing that: (1) based on individual word intelligibility data, intelligibility decrements in a connected message are minimal with less than 50% compression; (2) syntactic/semantic cues in a connected text compensate for the minor intelligibility decrements produced by compression; and (3) while intelligibility decrements are linked to the 50% compression rate, comprehension declines beyond 275 wpm regardless of the compression rate employed.

However, individual word intelligibility scores cannot be used to infer the intelligibility of words presented in an extended message. First, in the standard intelligibility test, a subject is given relatively long periods of time in which to recognize a word following its presentation. This is not the case when subjects must continuously process the words of an extended compressed message. Second, even minimal decrements in the intelligibility of every word in a connected message may have cumulative effects on comprehension. Also, contrary to Foulke and Sticht, if degraded words are misperceived, they will provide misleading syntactic and semantic cues for subsequent words and thus actually compound subsequent intelligibility decrements.

It is also incorrect to argue that comprehension declines dramatically beyond 275 wpm regardless of the compression rate employed. Indeed, *large* decrements in comprehension tend to occur only with greater than 50% compression, regardless of the final word rate. For example, Fairbanks, Guttman, and Miron (1957) found that comprehension declined significantly only after a message was compressed by 50% (resulting in 282 wpm). Also, Foulke (1968), Foulke, Amster, Nolan, and Bixler (1962), and Foulke and Sticht (1967) reported severe comprehension

G. W. Heiman's mailing address is: Department of Psychology, State University College, 1300 Elmwood Ave., Buffalo, NY 14222.

decrements only when more than 50% compression was employed. Although word rates over 325 wpm were produced using compression rates approaching 50%, these resulted in relatively small, although statistically significant, comprehension decrements. Finally, de Haan (1977, 1982) found the threshold for "understanding" compressed speech to be at approximately 52% compression. Thus, the cause of comprehension decrements may be intelligibility decrements that result when the underlying compression rate reaches or surpasses 50%.

A final rate of 275 wpm can be produced with less than 50% compression, but only if the initial spoken word rate is relatively high. Recall that the final word rate of a message is a function of both the compression rate and the initial spoken word rate of the message. When the original message is spoken rapidly, it will involve changes in the relative duration of consonants, vowels, and pauses (Kozhevnikov & Chistovich, 1966) that may produce intelligibility decrements when subsequently compressed. Thus, 275 wpm is achieved by approximately 50% compression of a normal, intelligible message, or by less compression of a potentially less intelligible message.

If 275 wpm is the critical factor, then it is actually the temporal constraints imposed by compression which result in comprehension decrements. If 50% compression is the critical factor that produces intelligibility decrements, then it is actually the amount of the speech signal deleted by the underlying interruption that results in comprehension decrements. In Experiment 1, these alternatives were tested in the following ways: First, if temporal constraints are crucial, major comprehension decrements will result with word rates greater than approximately 275 wpm regardless of the compression rate used. If the amount deleted is crucial, major comprehension decrements will result only with compression rates beyond 50%, regardless of the final word rate. Second, if temporal constraints are crucial, there should be no effect of increasing the percentage of the signal deleted by temporal interruption if the interrupted message is not compressed. If amount deleted is crucial, there should be decreased comprehension of interrupted messages with greater than 50% interruption. Third, simply stated, a comprehension overload implies that individual words are largely intelligible but that the message cannot be comprehended as rapidly as it is presented. To whatever extent preceding portions of a message are comprehended, or simply because contextually important words are intelligible and accurately perceived, contextual cues that aid in the accuracy and speed of comprehension of subsequent portions of the message will be available. Thus, if temporal constraints are crucial, comprehension scores might increase over successive portions of a message. If amount deleted is crucial, and words are not intelligible, initial words will be misperceived, providing inaccurate and misleading cues for subsequent degraded words. This will result in declining comprehension scores over successive portions of a message.

EXPERIMENT 1

Method

Subjects. Two hundred students served as subjects for credit in their introductory psychology course.

Materials and Apparatus. Two grammatically simple 500-word narratives were recorded by a male at 175 wpm. For each narrative, 20 short-answer comprehension questions were developed. None were answered correctly by pilot subjects who had not heard the narrative. The serial order of questions on the response sheet corresponded to the serial order in which the answers were presented in the narrative. An additional 200-word narrative and 10 questions were used in a practice session.

Interrupted speech was produced by playing the original recording from a Wollensack audio recorder through an electronic chopping circuit. The circuit monitored a frequency timing generator and a switch-selectable "on-off time" square-wave decoder. Switching was clean, and the output was close to being a square wave. Each time the original recording was played, the device interrupted the audio signal using a 30-msec off-period. The frequency of off-periods (varied between 0 and 20/sec) determined the percentage of the signal deleted—0%, 33%, 43%, 50%, or 60%.

Corresponding compression rates with 0%, 33%, 43%, 50%, or 60% of the signal deleted were produced by playing the original recording on a VSC brand A-6 variable speech compressor. This operates by first increasing playback speed, using a variable-speed tape drive, which, in turn, speeds the message and increases all component frequencies of the signal. The signal is then interrupted between 0 and 20 times per second with an off-period corresponding to a 30-msec real-time off-period. Each on-period is then passed through a clock-controlled shift register, which slows the signal so that, at output, it contains the normal frequency range and is of the original temporal duration. Each on-period is abutted in time with adjacent on-periods to form a continuous compressed signal.

The compressed or interrupted speech was presented over monophonic earphones located in sound-attenuating subject booths.

Procedure. One hundred subjects were presented compressed messages and 100 were presented interrupted messages. For each rate (0%, 33%, 43%, 50%, or 60%), 20 subjects were presented the practice narrative followed by both test narratives. Following presentation of a narrative, the subjects answered the comprehension questions, in writing, using one or two words. After responding to a question, the subjects moved a covering sheet downward, exposing the next question. The subjects were not to return to a question after having proceeded to the next. The subjects were tested in groups of 10 or less, so that the data for each condition were based on at least two compressed/interrupted presentations of the narratives. Order of narrative presentation was counterbalanced across subjects.

Results and Discussion

The number of correctly answered questions per narrative per subject was entered into an analysis of variance for the factors of interruption versus compression, order of narrative presentation, and percentage deleted (0%, 33%, 43%, 50%, and 60%). The effect of interruption versus compression was significant [$F(1,180) = 233.00, p < .01$] (accounting for approximately 18% of the total variance using η^2 ; Linton & Gallo, 1975), as was percentage deleted [$F(4,180) = 117.94, p < .01, \eta^2 = .36$]. Significant effects were found for narratives, order of narrative presentation, the interaction between order and narratives, and the interaction between type of dis-

Table 1
Mean Percent Correct and Slopes of Regression as a Function of Percentage Off Values and Interruption/Compression

	Percentage Off Values				
	0%	33%	43%	50%	60%
	Percent Correct				
Interruption	92.5	89.0	89.75	87.0	82.75
Compression*	90.75	83.67	78.13	70.62	24.25
	Slopes				
Interruption	-.08 ()	-.29 (7%)	-.31 (8%)	-.38 (9%)	-.42 (11%)
Compression	-.16 ()	-.61 (21%)	-.68 (17%)	-.87 (27%)	-.95 (31%)

Note—Values in parentheses indicate percentage of variance accounted for (r^2). Missing values indicate a nonsignificant correlation ($p > .05$). *Corresponding word rates for each percentage off value were 175, 262, 303, 350, and 437 wpm, respectively.

ruption and narratives, but each accounted for less than 1% of the total variance.

The interaction between interruption/compression and percentage deleted (shown in the upper portion of Table 1) was also significant [$F(4,180) = 68.92, p < .01, \eta^2 = .21$]. For interrupted narratives, Tukey post hoc tests indicated that the difference between means for 0% and 60% deleted was significant ($p < .05$). Thus, interruptions of more than 50% produced decreased comprehension, presumably due to decreased word intelligibility. Notably, even in the absence of the temporal constraints of compression, syntactic and contextual cues derived from the message could not fully compensate for the effects of the interrupted speech signal. With less than 60% deleted, word intelligibility may also have been slightly degraded, and the nonsignificant trend toward decreasing comprehension from 0% to 50% interruption supports this interpretation. For compressed narratives, there were significant differences between the means for 50% and 60% compression and for all nonadjacent rates. With 33%, 43%, and 50% compression, word rates were 262, 306, and 350 wpm, respectively. Clearly, there was no exceptional effect once a rate of 275 wpm was exceeded, although there was a substantial effect once 50% compression was exceeded.

Comprehension was significantly lower for compressed narratives than for correspondingly interrupted narratives at 43% deleted or greater. Comprehension of interrupted messages was relatively high, probably because the temporal duration of the original narratives was preserved, and thus there was ample time for subjects to attempt to process each word. Comprehension of compressed messages was much lower because a compressed signal is, first, an interrupted signal and presumably interruption produces decreased intelligibility. When the interrupted message is presented within the limited temporal constraints imposed by compression, there is not ample time for subjects to process each word and intelligibility decrements produced by the underlying interruption are exacerbated. Indeed, Garvey (1953) demonstrated that, with greater than 50% deleted, individually presented compressed words were substantially less intelligible than interrupted words. In the present study, even in the con-

text of a connected message, a similar finding was obtained.

Comprehension was also examined as a function of serial position within the narratives. Responses per subject per condition were grouped into four blocks of five questions each: Block 1 contained the total correct for Questions 1 through 5 of both narratives, Block 2 contained the total correct for Questions 6 through 10, and so forth. A regression analysis indicated that the best fitting function across blocks in each condition was linear.

As shown in the lower portion of Table 1, for interrupted narratives, performance decreased over the length of the narratives, with increasingly negative slopes with increasing percent deleted. This is inconsistent with a comprehension overload proposal, because, without the temporal constraints of compression, no overload could occur. It is also inconsistent with the notion that syntactic/semantic cues compensate for the intelligibility decrements produced by interruption/compression. However, it is consistent with the proposal that intelligibility decrements produce comprehension decrements: If initial words in a message are misperceived, they provide misleading cues for the perception of subsequent words, resulting in greater misperception and less accurate comprehension of subsequent portions of the message.

We have argued that a compressed message should be viewed as a temporally constrained interrupted message. The greater slopes with compression relative to interruption supports the notion that the temporal constraints of compression exacerbate the intelligibility decrements presumably produced by the underlying interruption. Furthermore, across the compression conditions, there were increasingly negative slopes with increasing percentages deleted. This is inconsistent with an STM overload explanation, because each block would represent comprehension of approximately 125 words. It is not likely that STM, with its limited capacity, would be increasingly overloaded by additional segments containing 125 words. However, the negative slopes are consistent with a cumulative effect of intelligibility decrements.

To determine whether the regression functions might be due to the order in which the comprehension questions were answered, the above compression conditions were

presented to 80 additional subjects. The procedure was identical to that of Experiment 1, except that the serial order in which the questions were answered was reversed. Both the overall levels of comprehension and the negatively sloping regression functions were not different from those of the preceding study.

EXPERIMENT 2

We have argued that intelligibility decrements produced by interruption are exacerbated by the temporal constraints imposed by subsequent compression. If compressed words are to remain intelligible, each reduction in the signal produced by deleting an off-period must be accompanied by a corresponding reduction in the amount of time subjects need to process the signal. If this does not occur, then processing of the word will overlap with presentation of the next word in the message, with each subsequent word acting as a backward masking stimulus for each preceding word. Word intelligibility decrements will be exacerbated because subjects must identify each word on the basis of an incomplete perceptual analysis of a signal already degraded by interruption.

In the present study, a backward masking procedure was applied to individually presented words. If word intelligibility is decreased by backward masking in a compressed message, then an individual compressed word's intelligibility will also be decreased by a subsequent masking stimulus. If a compressed word can be perceived within the temporal constraints of compression, then there should be no effect of the mask.

Method

Subjects. Forty-four students, similar to those in Experiment 1, served as subjects.

Materials and Procedure. A list containing 45 high-frequency one-, two-, and three-syllable words was recorded by a male speaker. Four re-recordings were generated from the initial recording: two for playback at 0% compression and two for playback at 50% compression. Each recording was such that, when played on the speech compressor at the appropriate rate, the list was presented at a rate of one word every 20 sec. One 0% compression recording and one 50% compression recording also contained a white-noise mask. In the masking conditions, when each word was played at the appropriate rate, 500 msec of mask occurred 100 msec after word offset. In producing the recordings, a Lafayette voice-activated relay in conjunction with a Colburn modular system detected stimulus word offset, timed the stimulus-mask interval, and produced the mask.

Eleven subjects were presented the list at 0% compression with mask or with no mask or at 50% compression with mask or with no mask. No attempt was made to control which portions of words were deleted by the compressor. Subjects were tested in groups of 4 or fewer, so that the data in each condition were based on at least three compressed presentations of the stimuli.

Immediately after the presentation of each word, the subjects recorded the word in writing or indicated if it was completely unintelligible by responding with an X.

Results and Discussion

The number of words correctly identified was entered into an analysis of variance for the factors of number of

syllables, compression, and mask/no mask. The effect of compression was significant [$F(1,40) = 189.65, p < .01, \eta^2 = .65$], as was the effect of mask/no mask [$F(1,40) = 13.87, p < .01, \eta^2 = .05$]. The only significant effects of word length occurred in the interactions with compression and with compression and mask/no mask, accounting for a total of 3% of the variance. Two-syllable words were slightly more intelligible than one- and three-syllable words, although it is unclear why this was so.

The interaction between compression and mask/no mask was also significant [$F(1,40) = 6.31, p < .01, \eta^2 = .02$]. Tukey post hoc comparisons indicated that at 0% compression, there was no significant difference between the mean correct for mask (91.71%) and for no mask (94.54%). Thus, initial perceptual processing of normal words was largely completed 100 msec after the offset of the word. At 50% compression, intelligibility was significantly lower in the mask condition (53.73%) than in the no-mask condition (68.28%). Thus, processing of compressed words was frequently not completed 100 msec after the offset of the word. There is apparently no reduction in subjects' word-processing time that corresponds to the temporal and acoustic reduction in the speech signal produced by compression. Furthermore, previous research regarding the intelligibility of individually presented compressed words probably presents an overestimate of the intelligibility of words in a compressed message. In fact, the actual masking occurring in a connected message might be much greater than that suggested by the present study, because words in a compressed message are not consistently separated by a 100-msec interval.

In the 50% compression condition, when subjects did not correctly report a word, in only 18% of the cases did they indicate that the word was unintelligible. This is consistent with our interpretation of the regression data of Experiment 1: few words in a message are probably totally unintelligible. Rather, words are misperceived and thereby generate inaccurate cues which increase the probability of misperceiving subsequent words.

GENERAL DISCUSSION

The effects of interruption in Experiment 1 were obtained under optimum temporal constraints, in that each narrative was presented at a normal rate. The effects of masking in Experiment 2 were obtained under relatively favorable temporal constraints in that backward masking was delayed by 100 msec and there was no possibility of forward masking. Yet, in both studies, there was decreased performance, presumably due to decreased word intelligibility. Under the limited temporal constraints of a continuous compressed message, the potential for much larger intelligibility decrements is very clearly present.

There are several factors that would cause the temporal constraints of compression to compound intelligibility decrements produced by interruption. Obviously, an interrupted signal loses some intelligibility because speech

information is deleted. Yet there is often sufficient information remaining in the signal after interruption, with the "off-periods" having deleted what is, in effect, redundant information. However, a stimulus presented in a non-redundant context requires greater processing time than does one presented in a redundant context (e.g., Reicher, 1969), and as redundancy decreases, the amount of time required increases (e.g., Tulving & Gold, 1963). Interruption decreases the acoustic redundancy in the speech signal, thereby decreasing the speed with which it can be perceived. An interrupted signal remains relatively intelligible because off-periods provide time during which preceding on-periods can be processed. Compression eliminates the processing time provided by the off-periods, necessitating that each portion be perceived in *less* time than normal if subjects are to keep pace with the incoming signal. Each unit of speech is presented in a less redundant context, however, so that *more* time per unit is required.

The above interpretation also suggests why the crucial parameter is at approximately 50% compression. It has been argued that any 50% of the speech signal is redundant, given the remaining 50% (Fodor, Bever, & Garrett, 1974). Such a description is probably inaccurate for any single presentation of an interrupted or compressed message, because the amount of information deleted depends on the exact placement of the off-periods. However, on the basis of the large body of research dealing with compressed speech, it is reasonable to propose that, on average, 50% compression removes virtually all redundant acoustic information. With greater than 50% compression, critical nonredundant speech information is also deleted. Thus, 50% compression is the crucial parameter, because greater than 50% compression presents too little of the signal in too little time for a sufficient number of words to be accurately perceived.

Compression may also have a secondary effect by degrading the remaining acoustic information contained in the interrupted signal. Miller and Licklider (1950) inserted white noise in the off-periods of very rapidly interrupted speech and found that the intelligibility of words with noise inserted was decreased to the level of uninterrupted words presented in a continuous noise background. Thus, the noise of the off-periods was perceptually integrated with the speech of the on-periods so that the signal was the functional equivalent of one presented in continuous noise. In a compressed message, what would be an off-period is filled not with white noise, but with the adjacent on-period of speech. Thus, any moment of compressed speech may be the functional equivalent of an interrupted signal presented in a speech-produced noise background.

Furthermore, although perceptual integration of adjacent sounds occurs in normal speech, the adjacent sounds in the compressed signal are sounds that often were originally separate. Thus, when these sounds are integrated, they may not be perceived as the appropriate speech

sounds. For example, in normal speech, if voice onset time is delayed by approximately 30 msec, stop consonants are perceived as voiceless, whereas with little or no delay, they are perceived as voiced (see Miller, 1981). Should an off-period remove the delay of a voiceless stop, the compressed version will be perceived as voiced. Thus, compressed speech may result in the integration of sounds that alter the original perceptual characteristics of the signal and thereby further decrease word intelligibility.

In summary, the compressed word rate of 275 wpm appears to be of little significance. The present research suggests that there are several ways in which compression and its underlying temporal interruption decrease word intelligibility, and that it is decreased word intelligibility that results in decreased comprehension.

REFERENCES

- DE HAAN, H. J. (1977). A speech-rate intelligibility threshold for speeded and time-compressed connected speech. *Perception & Psychophysics*, **22**, 366-372.
- DE HAAN, H. J. (1982). The relationship of estimated comprehensibility to the rate of connected speech. *Perception & Psychophysics*, **32**, 27-31.
- FAIRBANKS, G., GUTTMAN, N., & MIRON, M. S. (1957). Auditory comprehension of repeated high speed messages. *Journal of Speech & Hearing Disorders*, **22**, 20-22.
- FODOR, J. A., BEVER, T. G., & GARRETT, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.
- FOULKE, E. (1968). Listening comprehension as a function of word rate. *Journal of Communication*, **18**, 198-206.
- FOULKE, E., AMSTER, C. H., NOLAN, C. Y., & BIXLER, R. H. (1962). The comprehension of rapid speech by the blind. *Exceptional Children*, **29**, 134-141.
- FOULKE, E., & STICHT, T. G. (1967). The intelligibility and comprehension of time compressed speech. In E. Foulke (Ed.), *Proceedings of the Louisville Conference on Time Compressed Speech*, October 19-21, 1966 (pp. 21-28). Louisville, KY: University of Louisville, Center for Rate Controlled Recordings (Perceptual Alternatives Laboratory).
- FOULKE, E., & STICHT, T. G. (1969). Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin*, **72**, 50-62.
- GARVEY, W. D. (1953). The intelligibility of speeded speech. *Journal of Experimental Psychology*, **45**, 102-108.
- KOZHEVNIKOV, V. A., & CHISTOVICH, L. A. (1966). *Speech: Articulation and perception* (Trans. No. JPRS 30, 543, 1965). Washington, DC: U.S. Department of Commerce, Joint Publications Research Service.
- LINTON, M., & GALLO, P. S. (1975). *The practical statistician: Simplified handbook of statistics*. California: Wadsworth.
- MILLER, G. A. (1981). *Language and speech*. San Francisco: W. H. Freeman.
- MILLER, G. A., & LICKLIDER, J. C. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, **22**, 167-173.
- REICHER, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimuli material. *Journal of Experimental Psychology*, **81**, 275-280.
- TULVING, E., & GOLD, C. (1963). Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology*, **66**, 319-327.