A Study of Rate Discrimination of Time-Compressed Speech

Lisa J. Stifelman
Speech Research Group, MIT Media Laboratory

Abstract

The purpose of this study is to determine the just noticeable differences (JNDs) for speech rate. The results are intended to be used for the design of an interactive speech speed control. The JND at three different speech rates was determined using the psychophysical method of constant stimuli. Speech stimuli were compressed using the SOLA time-compression technique. The findings show a significant increase in the JND as speech rate increases. However, a non-linear increase was found rather than a linear one as predicted by Weber's law. In addition, JNDs were much smaller than initially expected. The results indicate the need for a fine granularity, possibly non-linear, speech speed control.

Introduction

A variety of techniques currently exist for time-compressing speech without changing the pitch. Studies of time-compressed speech to date have focused on intelligibility and comprehension of isolated and connected speech using a variety of time-compression techniques (for a summary see [3, 11, 12]). In addition, considerable attention has been given to the determination of an intelligibility threshold [5, 7, 8, 19], and the effects of exposure and training [14, 21, 30].

Although several psychophysical studies have been performed regarding the perception of speech rate (see Related Research), no study of rate discrimination has been performed[1]. The purpose of this study is to determine the just noticeable differences (JNDs) for the rate of time-compressed speech (i.e., the amount of change in rate that a listener can reliably discriminate).

The results of this study are intended to be used for the design of an interactive speed control that allows a user to adjust the speed of playback while listening to digitally recorded speech. The development of an interactive speech speed control is important for use in speech interfaces and applications (e.g., [2, 28, 29]). Due to the slow and serial nature of speech, interactive control of playback speed is important in order to make listening more efficient.

There are two important findings that can be applied to the design of a speech speed control. First, the size of the JNDs can be used in determining the "granularity" of control. For example, if a user cannot perceive the difference between a normal speed of 1.0 and 1.25 times the original speed, it is unnecessary to provide continuous control between these two speeds. Second, the change in the size of the JND as a function of compression rate can be used to determine whether a linear or non-linear control should be provided.

---

[1]The only rate discrimination study found focused on the influence of pitch contours on the perception of rate [22]. In this study, the fundamental frequency rather than the speech rate was varied.

Related Research

Perception of Duration

Lehiste provides a summary of several studies measuring the minimum differences in duration that a listener is capable of discriminating (i.e., JNDs for duration) [22]. Among the studies cited, there was much variability in the findings. For stimuli in the range of 30 to 300 ms, the just noticeable differences were between 10 and 40 ms. The maximum standard duration cited was 600 ms with a JND value of 69 ms. Weber's ratio for duration discrimination, based on the results of these studies, does not remain constant. Since these studies employed non-speech stimuli (e.g., a 1,000 Hz tone), the findings may not be directly applicable to speech perception. Lehiste found a strong influence of the fundamental frequency pattern on the perception of duration of speech stimuli [23]. The results of this study indicated that listeners may perceive speech segments with changing fundamental frequency as longer than their actual length.

Fujisaki [13] studied the perception of duration of both speech and non-speech stimuli. For a standard non-speech stimulus (500 Hz tone) in the range of 50 to 300 ms, JND values were between 7.6 and 23.1 ms—somewhat shorter than findings summarized by Lehiste. Discrimination of synthetic speech stimuli given a standard duration of 200 ms, ranged from 7.1 to 16 ms depending on the manner of articulation (vowel, fricative, plosive, or nasal) and the context (word versus sentence). These findings indicate a finer discrimination for the duration of speech than for non-speech stimuli.

Nooteboom and Doodeman studied the perception of vowel duration[2] in spoken sentences and reported an average JND of 5 ms for a standard duration of 90 ms. Again, this value is shorter than the JND values cited in a similar range for non-speech stimuli.

Perception of Rate

Cartwright and Lass [4] studied the relationship between measured and perceived rate of time-compressed continuous speech using the psychophysical method of magnitude estimation. As expected, there was not a one-to-one relationship between measured and perceived rate. Perceived rate was found to be a function of measured rate raised to a power.

Grosjean studied both naturally and mechanically altered speech rate and points out some important differences between the two [17, 18]. When speech rate is naturally altered, articulation rate, and the number and duration of pauses are varied. However, speech that is mechanically time-compressed varies only in articulation rate and the duration of pauses, while the number of pauses remains constant. Grosjean and Lane conclude that the perception of changes in rate (whether mechanically or naturally altered) is not based on linguistic decoding (e.g., number of phrases) but on articulation and pause rate. This conclusion is based in part on the fact that no differences were found in the perception of rate using speech stimuli spoken in French by subjects that were native speakers of French and subjects that did not speak any French.

Rietveld found a significant effect of pitch contour on the perception of speech rate [24]. This is similar to the findings by Lehiste on the influence of fundamental frequency patterns on the perception of duration [23].

---

[2]Note that Dutch vowel sounds were used for the stimuli in this study.

Method

Subjects

An initial pilot study of 4 subjects, 3 female and 1 male, was performed. Thirteen subjects, 6 female and 7 male, participated in the actual experiment. Nine of the subjects were students from the Media Laboratory at the Massachusetts Institute of Technology. Four of the 13 subjects had previous experience listening to time-compressed speech.

Apparatus

The experiment was run on a Macintosh IIfx computer, using stereo headphones and a small keypad for the subject's response entry.

Time-Compression Technique

There are several techniques for time-compressing speech without changing the pitch (see [1] for a summary). Two implementations of such time-compression techniques were available for use in this experiment.

One of these implementations is based on the Fairbank's sampling method [9] of time-compression and runs in real time on a Macintosh computer. This isochronous sampling technique operates in the time domain removing segments of speech at regular intervals (Figure 1). The algorithm performs a linear cross-fade at interval boundaries to minimize distortion. The sampling technique used in this implementation "blindly" removes portions of the speech signal without knowledge of the contents. More sophisticated selective sampling techniques (Figure 1) attempt to locate and remove redundant portions of the speech signal [6, 26].

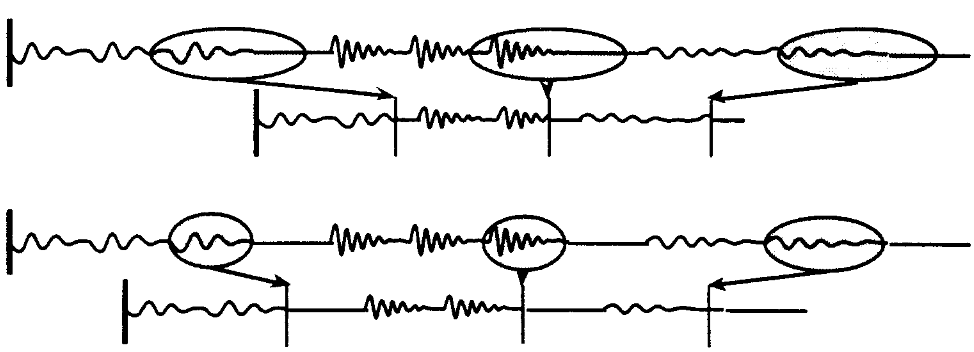

Figure 1: Conceptual representation of time-compression techniques. Top: Isochronous sampling—segments of speech are removed at regular intervals without knowledge of the contents. Bottom: Pitch-synchronous selective sampling—redundant pitch periods are selectively removed from the speech signal.

The second implementation available for use in this experiment uses a synchronized overlap add (SOLA) time-compression technique [25] and runs on a Sun SparcStation. This technique attempts to overlap and add portions of the speech signal at points of highest cross-correlation. The result is effectively a selective removal of speech, although no specific attempt is made to locate pitch periods (as ideally depicted in Figure 1).

Sampling techniques for time-compression "result in discontinuities in the transitions between inserted or deleted segments" ([20], p. 3). Algorithms like SOLA attempt to minimize these effects "by improving the splicing or windowing adjoining segments"

([20], p. 3). The SOLA implementation was selected for this experiment since there is expected to be less distortion at interval boundaries, resulting in a higher quality signal.

Auditory Stimuli

The phrase "Human Factors Engineering" was recorded by a female speaker using a MacRecorder connected to a Macintosh IIfx. The speech was recorded at a sampling rate of 7.4 kHz using 8 bit linear[3] coding. The original sound file was 1.38 seconds in length. This sound file was then transferred to a Sun SparcStation and time-compressed using the SOLA technique. The resulting sound files were then transferred back to a Macintosh on which the experiment was run.

Experimental Design

The psychophysical method of constant stimuli [16, 27] was used to determine the subject's differential sensitivity to the rate of time-compressed speech. This involves the repeated use of a single stimulus that varies in only one dimension. In this experiment the stimuli consist of the same segment of speech time-compressed at a total of 27 different rates (Table 1). The subject is asked to compare pairs of stimuli (a standard stimulus against a comparison stimulus) and to judge which produces a sensation of greater magnitude. In this case the subject judged whether one stimulus was "faster" or "slower" than the other. A forced-choice paradigm was used—the subject must answer either "faster" or "slower" ("equal" is not a valid response). Since speech stimuli must be presented in sequence rather than in parallel, a time error can result due to the time delay between the presentation of the two stimuli. The memory of the first stimuli may fade during this delay interval causing a possible bias in response (e.g., the first stimuli may tend to be judged as "faster" than the second or vice versa). To compensate for this time error, the order of the stimuli in each pair is counterbalanced such that on half of the conditions the standard stimulus is presented first, and on the other half the comparison is presented first.

Standard	**Comparison Rates**										
1.0	0.875	0.900	0.925	0.950	0.975	1.000	1.025	1.050	1.075	1.100	1.125
1.5	—	1.30	1.35	1.40	1.45	1.50	1.55	1.60	1.65	1.70	—
2.0	—	—	1.7	1.8	1.9	2.0	2.1	2.2	2.3	—	—

Table 1: Standard and comparison compression rates tested. Note that the original sound file was also processed by the algorithm at a compression rate of 1.0 (no change in speed) to ensure that any noise artifacts introduced by the time-compression algorithm would be present in all stimuli.

In this experiment, three standard rates were selected—1.0, 1.5, and 2.0, in order to find the differential sensitivity to the change in speed at each of these points. The standard rates were selected in a range expected to be used by a speech speed control. A compression rate of 2.0 times normal speed has been cited as the threshold above which comprehension begins to drop off rapidly [10, 12]. Heiman states that compression greater than 50% (2.0 times as fast) "presents too little of the signal in too little time for a sufficient number of words to be accurately perceived" ([19], p. 411). However, this threshold varies depending on the amount of subject exposure to time-compressed speech and the time-compression technique employed [7, 14]. A maximum standard rate of 2.0 was selected for this experiment, with a maximum comparison rate of 2.3 times normal speed.

---

[3] Since the Macintosh uses linear rather than logarithmic coding, the resolution is 8 bits rather than the 12 bits achieved with μ-law encoded speech. A dynamic range larger than 8 bits would have been preferable.

Each standard rate is compared with a selected number of comparison rates at constant increments above and below the standard. The comparison rates tested for each standard are shown in Table 1. Note that each standard is also compared with itself so that in some trials the standard and comparison stimuli will be of equal rate. Based on the results of a pilot study, the comparison rates were chosen such that the comparison stimulus with the fastest rate is almost always judged as faster than the standard (e.g., 1.125 vs. 1.0) and the comparison stimulus with the slowest rate is almost always judged as slower than the standard (e.g., 0.875 versus 1.0). An attempt was also made to select an increment between stimuli (e.g., a 0.025 increment is used between comparison stimuli for the 1.0 standard) that is a smaller change than can be reliably discriminated.

Procedure

Each subject was tested in a quiet room for approximately 30 minutes. Subjects were instructed that they would hear a series of pairs of speech segments and were to judge whether the second segment was either "faster" or "slower" than the first. The subject pressed one of two buttons (F or S) on a small keypad. The stimuli were presented diotically (i.e., the same signal played to both ears) over headphones. There was a 0.5 second delay between stimuli, and a 1.0 second delay between trials. The test was "self-paced" in that the next trial did not begin until 1.0 second after the subject made a response.

The experiment was composed of three tests, one for each standard compression rate. Each subject participated in all three tests. The tests were counterbalanced for order and sequence effects. Thirteen subjects were tested such that the six possible orders of the tests were used at least twice. For each of the three tests, the order of presentation of the pairs of stimuli was completely randomized. Following each test the subject was briefly interviewed and given a short break.

For the 1.0 standard rate, 11 pairs of stimuli were compared (Table 1) in the order standard-comparison and comparison-standard, for a total of 22 pairs. Each pair was presented 5 times for a total of 110 trials. For the 1.5 standard, 9 stimulus pairs were compared in two different orders, 5 times each[4] for a total of 90 trials. For the 2.0 standard, 7 stimulus pairs were compared in two different orders, 5 times each for a total of 70 trials. An additional five trials were added at the beginning of each test for practice. These trials were not considered in the final analysis.

Results

Pilot Test

An initial pilot study was run for the standard compression rates of 1.0 and 1.5. The main goal of the pilot study was to test the experimental design and adjust the range of comparison rates for the actual experiment. The maximum and minimum rates compared with a standard of 1.0 were adjusted since the pilot subjects were able to discriminate the stimuli on 100% of the trials at rates below the maximum and above the minimum tested (Figure 2).

---

[4]Further experiments with a larger number of repetitions should be performed. However, it was difficult to maintain the subject's attention throughout the task, and adding many more trials would have increased this problem (see Discussion section).

For the standard rate of 1.5, two ranges of comparison rates were tested—one set at increments of 0.1 above and below the standard, and one set at 0.05 increments. As can be seen in Figure 3, a 0.1 increment resulted in a polarized psychometric function—the subject was almost always able to detect when the comparison stimulus was "faster" or "slower". On the basis of these results, it was determined that a 0.05 increment would allow a more accurate measurement of the JND in this range[5]. In addition, the pause time between stimuli and trials was adjusted during the pilot phase of testing.

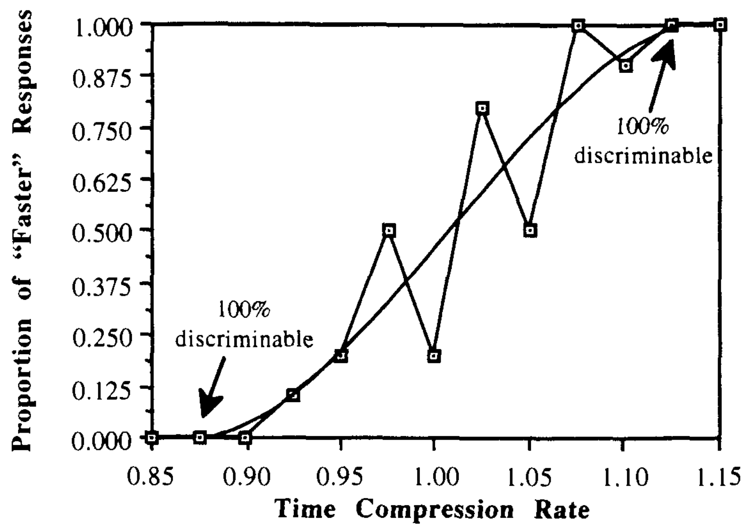

Figure 2: Results for pilot subject #1 for a 1.0 standard compression rate.

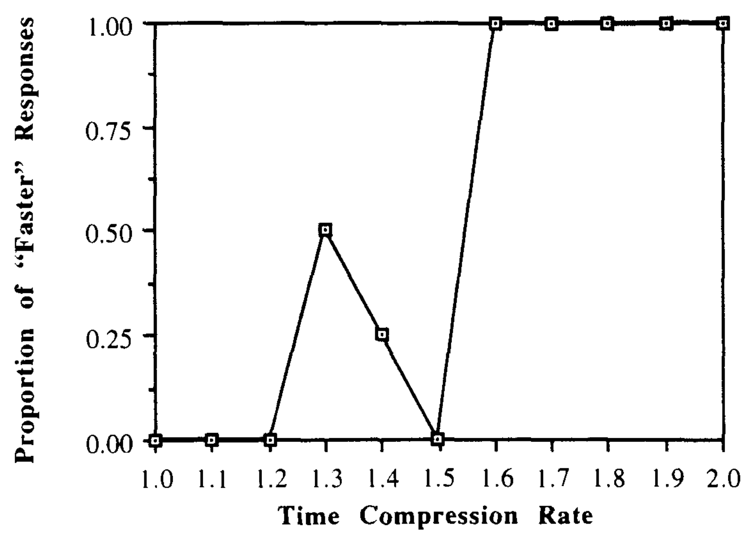

Figure 3: Results for pilot subject #1 for a 1.5 standard compression rate.

---

[5]Note that the same step size was not used for each standard. In order to use the same size, the smallest increment (0.025) would have had to be selected, greatly increasing the number of comparisons for the 1.5 and 2.0 standards (e.g., increasing the number of comparison rates from 6 to 24 for the 2.0 standard). Since it was impractical to quadruple the length of the test for each subject, appropriate step sizes were chosen for each standard based on the results of the pilot study.

Test Results

For each subject, psychometric functions were plotted (Figure 4). This function shows the percentage of "faster" responses for each comparison stimulus. When 50% of the subject's responses are "faster" and 50% are "slower" this is known as the point of subjective equality (PSE). Note that the PSE does not necessarily correspond exactly to when both stimuli are physically equal. Two difference limen (DL), an upper and a lower, can then be determined. The upper DL is equal to the point where 75% of the subject's responses are "faster" minus the PSE. The lower DL is equal to the PSE minus the point where 25% of the subject's responses are "faster". These two values, the upper DL, and the lower DL, can then be averaged to produce a single JND value for each standard rate.

Upper DL, lower DL, PSE, and JND values were calculated for each subject, for each of the three standard rates. Figures 4 through 6 show the results for one of the subjects. A summary of the results across all subjects is given in Table 2 and Figure 7.

The mean JND values for subjects previously experienced in listening to time-compressed speech were compared to the means for subjects without previous experience (Table 3). The differences between the means was not significant at the $p < .05$ level for any of the standard rates. Therefore, the data was evaluated across all subjects.

The mean JNDs across all subjects for standard compression rates of 1.0, 1.5, and 2.0 were 0.053, 0.079, and 0.133 respectively. A one-way analysis of variance indicates a significant difference between these means[6], $F(2, 32) = 4.53$, $p < .05$. Weber's ratio has also been calculated based on the mean JND values for each standard compression rate (Table 4).

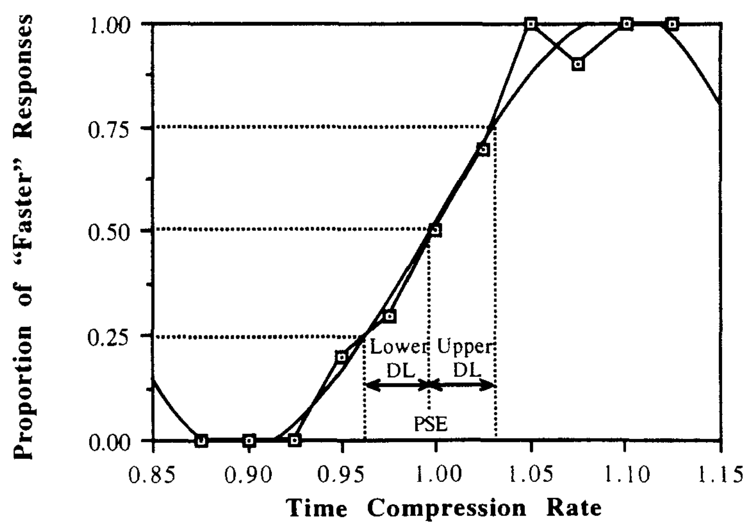

Figure 4: Results for subject #10 for a 1.0 standard rate of time-compression. The PSE is 0.997, the upper DL is 0.034, the lower DL is 0.034, and the JND is 0.034.

---

[6]Note that JND values could not be calculated in four cases since the subject's responses did not reach either the 25% or the 75% level. These subjects found the task especially tedious and did not remain attentive throughout the test.

JND	Standard Compression Rate 1.0	1.5	2.0
Mean	0.053	0.079	0.133
Median	0.045	0.059	0.123
SD	0.026	0.036	0.057

Table 2: Summary of JND values for all subjects.

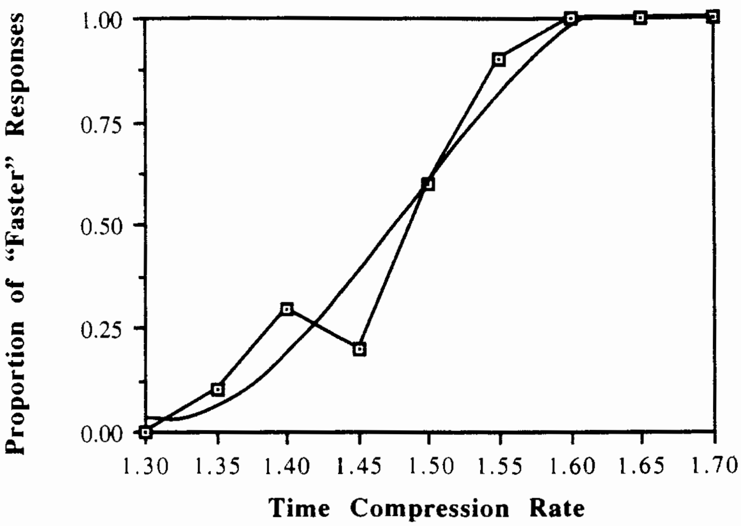

Figure 5: Results for subject #10 for a 1.5 standard rate of time-compression. The PSE is 1.47, the upper DL is 0.056, the lower DL is 0.059, and the JND is 0.058.

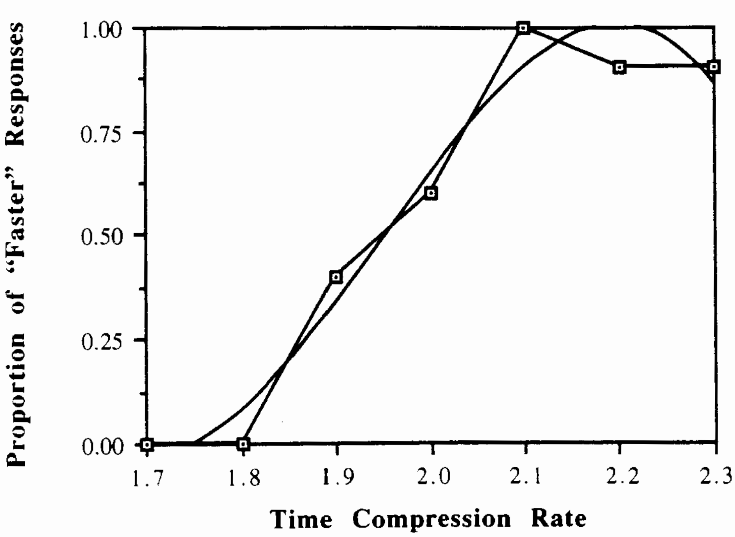

Figure 6: Results for subject #10 for a 2.0 standard rate of time-compression. The PSE is 1.953, the upper DL is 0.084, the lower DL is 0.083, and the JND is 0.083.

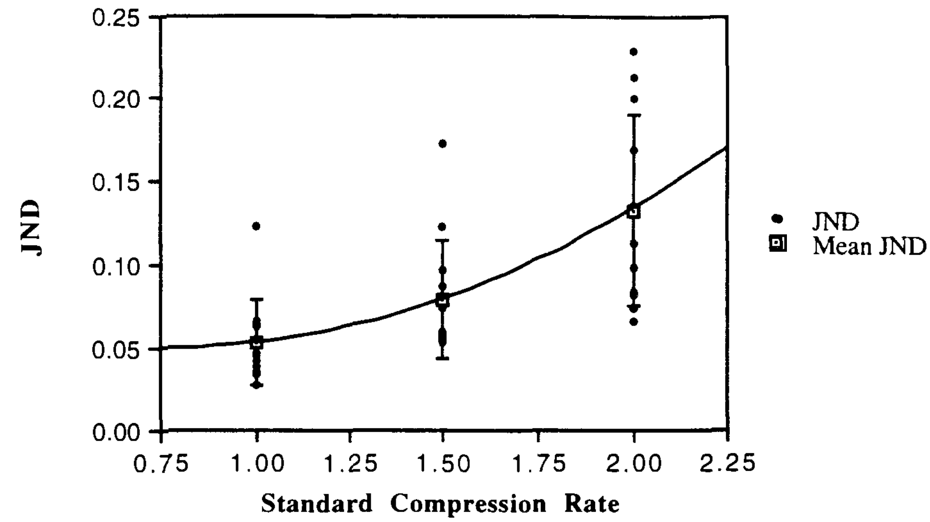

Figure 7: JND values for each subject and mean JND values across subjects for each of the three standard compression rates tested. The error bars show the standard deviation.

	Standard Compression Rate					
	1.0		1.5		2.0	
JND	**Exp**	**No-Exp**	**Exp**	**No-Exp**	**Exp**	**No-Exp**
Mean	0.050	0.055	0.071	0.084	0.143	0.128
Median	0.053	0.045	0.066	0.059	0.134	0.123
SD	0.018	0.031	0.020	0.043	0.071	0.054

Table 3: Summary of JND values for subjects with previously experience listening to time-compressed speech (Exp, N=4) versus those subjects with no previous experience (No-Exp, N=9).

Standard Rate	Weber's Ratio
1.0	0.053
1.5	0.053
2.0	0.067

Table 4: Weber's ratio ($\Delta R/R$) calculated for each standard rate based on the mean JND values listed in Table 2.

Note On Compression Rates

After the experiment was performed, it was discovered that the implementation of the SOLA algorithm outputs sound file lengths that are always multiples of 192 bytes. Therefore, the rates given in Table 1 are not exact when based on the actual length of the resulting sound file. A listing of the rates calculated based on the lengths of the sound files is given below (Table 5).

Standard	**Comparison Rates**										
1.0	0.870	0.899	0.914	0.947	0.982	1.000	1.019	1.039	1.081	1.104	1.127
1.51	—	1.29	1.36	1.39	1.47	1.51	1.55	1.60	1.65	1.70	—
2.0	—	—	1.7	1.8	1.88	2.0	2.1	2.19	2.29	—	—

Table 5: Compression rates based on actual file lengths.

INTERVIEW RESULTS

Following the experiment, each subject was briefly interviewed. Subjects were questioned about the difficulty of the tasks and were asked to report on any strategies employed. Two of the subjects reported that when they were unsure of an answer, they always chose "faster" (or "slower") indicating a possible response bias. The method of constant stimuli was designed with the expectation that when subjects are unsure of their response, the resulting performance will be at 50% on the psychometric function (i.e., 50% "faster" responses, 50% "slower"). No bias was evident, however, when reviewing the results.

Several strategies were reported by the subjects. Most subjects used the length of the vowels in comparing the rate of the two stimuli, especially the /ae/ phoneme in the word "factors." In addition, subjects also attempted to judge the rate of articulation, especially at the beginning and endpoints of the phrase (e.g., "engineering" sounded more like "and gin eering" when played at a slower rate). Another strategy was to determine the range of speeds being presented, and judge the first stimulus as being one of the "faster" or "slower" ones prior to hearing the second stimulus. Subjects reported that this strategy was mostly employed when it seemed obvious that the first stimulus was the "fastest" or the "slowest" in the range being tested. One subject reported noticing faster rates due to shorter pause lengths. Some subjects noticed pitch changes for the fastest and slowest rates tested, and used this information to distinguish the rate.

Subjects were also asked to compare the difficulty of each test (1.0 versus 1.5 versus 2.0). There was no consistency in the subject's responses. Some thought the fastest rates (standard of 2.0) were more difficult to distinguish than the slower rates, while others reported more difficulty for the rates compared with the 1.0 standard. Since the order of the tests was counterbalanced, subjects experienced a different amount of practice prior to each test. For example, subjects receiving the 2.0 test last may have found it easier, having been well practiced by this point in the experiment.

Discussion

The results indicate a much finer discriminatory capability than was expected prior to the experiment. As expected, however, the JND increases as the standard rate of compression increases. The results of this experiment indicate a non-linear relationship between the JND and the standard rate rather than linear as predicted by Weber's law[7]. Although Weber's ratio did remain constant at the 1.0 and 1.5 standard rates, it increased for the 2.0 standard. Further testing is needed with a larger number of standard rates to verify this finding.

The level of discrimination found by this study is probably finer than would be expected under uncontrolled conditions. The subjects listened to the speech over headphones in a quiet room, with no disruptions of the task. In the test, judging the rate was the primary task, while in actual use of a speed control, adjusting the rate will be secondary to the task of searching for, or listening to, information in a speech database.

No difference was found between the mean JNDs for subjects with previous experience listening to time-compressed speech and those without previous experience. Although

---

[7] Weber's law says that the amount that must be added to a stimulus to produce a just noticable difference is a constant fraction of the stimulus intensity ($\Delta I/I = C$, where C is a constant) [16].

previous experience did not have an effect, it is suspected that practice could lead to significant improvement of the JND. Previous researchers have tried to examine the effects of exposure and training on comprehension of time-compressed speech. Voor found significant improvement in comprehension after a short amount of exposure to time-compressed speech [30]. However, Lass [21] reported that exposure did not significantly improve comprehension scores, although it did have a significant influence on preferred listening rate. Since no control group was used in the Voor study, it cannot be determined whether the effects reported are due to exposure to time-compressed speech or simply to practice of the comprehension task.

In evaluating the results of this experiment, a question arises regarding the subject's discrimination of duration and the discrimination of rate. Since each of the comparison stimuli differed from the standard in rate and duration of speech (e.g., a compression of 2.0 decreases a sound's length by one half), subjects may have been judging differences in duration rather than differences in rate, or perhaps a combination of the two. These findings for rate discrimination should therefore be validated in a further study that isolates the variable of rate from duration. This could be accomplished in one of two ways. One possibility is to use speech stimuli that differ in duration by an amount smaller than the JND for duration in the range tested. Given the range in duration of the stimuli used in this experiment (approximately 600–1380 ms), it is possible that the differences in duration were below the JND. However, a more robust design would completely randomize stimuli duration (e.g., a speech segment played at a rate of 2.0 would be randomly longer or shorter than a segment played at 1.9).

Another question regards the problem of using the same speech content for each of the trials. Based on the method of constant stimuli, only one dimension (i.e., rate) should be varied. This results in an extremely tedious task. The subject may therefore "tune out" after a certain number of trials, having heard the same speech segment spoken over and over again, only varying in rate. Therefore, care was taken to minimize the number of trials used in each test, and subjects were warned that the task might be tedious but to try to remain attentive. Weber's ratio can be reduced when an appropriate payoff function is employed (rewarding for correct judgments and penalizing for incorrect judgments) [15]. No explicit payoff function was used in this experiment and subjects did not receive any monetary reimbursement for their time.

Since the contents of the speech segment remained constant throughout this experiment, a future experiment could be performed to validate these results, by varying the contents over several conditions. In addition, it would be interesting to compare the results of this experiment, which used the SOLA time-compression technique, to one using an isochronous sampling technique.

Conclusion

The goal of this study was to determine the just noticeable differences in speech rate for the design of a speech speed control. Based on the JND values calculated, it appears that a very fine grained control would be required. In addition, since the JND increases as the compression rate increases, the granularity of control could vary based on the speech rate rather than remaining constant. For example, finer control could be provided at slower speeds (e.g., 1.0–1.3), since the JND is smaller in this range.

In attempting to apply these results to the design of a speed control, the question arises as to whether the just noticeable difference is the most appropriate measure for determining the granularity of control. Perhaps, the "just *notable* difference" is actually the value that needs to be determined. In other words, what amount of increase in speed is most useful (i.e., notable) for a given task such as searching through a large speech database. For example, even though a rate of 1.05 may be distinguishable from 1.0, the amount of speed increase may be insignificant for the task of searching.

Since there is no empirical test designed to reveal the just notable difference, an iterative design and testing approach may be employed to compare different types of speed controls (e.g., using joystick, thumb-controlled track ball, volume-like potentiometer, etc.) with different levels of granularity. The results of this study can be used as a basis for the initial design of such a control.

Acknowledgements

This study was performed as a project for Tom Sheridan's Human Factors Engineering class at the Massachusetts Institute of Technology. Thanks to Greg Zacharias for his assistance with planning this study. Thanks also to Barry Arons, Tony Brown, and Chris Schmandt for reviewing drafts of this paper and providing useful feedback.

References

1. B. Arons. Techniques, perception, and applications of time-compressed speech. In Proceedings of AVIOS '92, pages 169-177. American Voice I/O Society, 1992.

2. B. Arons. SpeechSkimmer: Interactively Skimming Recorded Speech. *In Proceedings of* the ACM Symposium on User Interface Software and Technology, pages 187-195. ACM, 1993.

3. D.S. Beasley and J.E. Maki. Time- and frequency-altered speech. In N.J. Lass, editor, *Contemporary Issues in Experimental Phonetics*, chapter 12, pages 419-458. Academic Press, 1976.

4. L.R. Cartwright and N.J. Lass. A psychophysical study of rate of continuous speech stimuli by means of direct magnitude estimation scaling. *Language and Speech*, 18:358-365, 1976.

5. R.P. Carver. Effect of increasing the rate of speech presentation upon comprehension. *Journal of Educational Psychology*, 65(1):118-126, 1973.

6. E.E. David and H.S. McDonald. Note on pitch-synchronous processing of speech. *Journal of the Acoustic Society of America*, 28(7):1261-1266, 1956.

7. H.J. deHaan. A speech-rate intelligibility threshold for speeded and time-compressed connected speech. *Perception and Psychophysics*, 22(4):366-372, 1977.

8. H.J. deHaan. The relationship of estimated comprehensibility to the rate of connected speech. *Perception and Psychophysics*, 32(1):27-31, 1982.

9. G. Fairbanks, W.L. Everitt and R.P. Jaeger. Method for time or frequency compression-expansion of speech. *Transaction of the Institute of Radio Engineers, Professional Group on Audio*, AU-2:7-12, 1954.

10. G. Fairbanks, N. Guttman and M.S. Miron. Effects of time compression upon the compression of connected speech. *Journal of Speech and Hearing Disorders*, 22:10-19, 1957.

11. E. Foulke. The perception of time-compressed speech. *In Proceedings of* Symposium of the Learning and Research Development Center, pages 79-119. 1971.

12. E. Foulke and T.G. Sticht. Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin*, 72(1):50-62, 1969.

13. H. Fujisaki, K. Nakamura and T. Imoto. Auditory perception of duration of speech and non-speech stimuli. In G. Fant and M.A.A. Tatham, editor, *Auditory Analysis and Perception of Speech*, pages 197-219. Academic Press, 1975.

14. P.A. Gade and C.B. Mills. Listening rate and comprehension as a function of preference for and exposure to time-altered speech. *Perceptual and Motor Skills*, 68:531-538, 1989.

15. E. Galanter. Textbook of Elementary Psychology. San Francisco, Holden-Day, Inc., 1966.

16. G.A. Gescheider. Psychophysics: Method, Theory, and Application. New Jersey, Lawrence Erlbaum Associates, 1985.

17. F. Grosjean and H. Lane. How the listener integrates the components of speaking rate. *Journal of Experimental Psychology: Human Perception and Performance*, 2(4):538-543, 1976.

18. F.H. Grosjean and N.J. Lass. Some factors affecting the listener's perception of reading rate in English and French. *Language and Speech*, 20(3):198-208, 1977.

19. G.W. Heiman, R.J. Leo, G. Leighbody and K. Bowler. Word intelligibility decrements and the comprehension of time-compressed speech. *Perception and Psychophysics*, 40(6):407-411, 1986.

20. D.J. Hejna. Real-Time Time-Scale Modification of Speech via the Synchronized Overlap-Add Algorithm. Masters Thesis. Massachusetts Institute of Technology, 1990.

21. N.J. Lass, E. Foulke, A.A. Nester and J. Commerci. The effect of exposure to time-compressed speech on subject's listening rate preferences and listening comprehension skills. *Journal of Auditory Research*, 3:179-186, 1974.

22. I. Lehiste. Suprasegmentals. Cambridge, MIT Press, 1970.

23. I. Lehiste. Influence of fundemental frequency pattern on the perception of duration. *Journal of Phonetics*, 4:113-117, 1976.

24. A.C.M. Rietveld and C. Gussenhoven. Perceived speech rate and intonation. *Journal of Phonetics*, 15:273-285, 1987.

25. S. Roucos and A.M. Wilgus. High quality time-scale modification for speech. *In Proceedings of* International Conference on Acoustics, Speech, and Signal Processing, pages 493-496. IEEE, 1985.

26. R.J. Scott and S.E. Gerber. Pitch-synchronous time-compression of speech. *In Proceedings of* Conference on Speech Communication and Processing, pages 63-65. IEEE, 1983.

27. S.S. Stevens. Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects. John Wiley & Sons, Inc., 1975.

28. L.J. Stifelman. VoiceNotes: An Application for a Voice-Controlled Hand-Held Computer. Master's Thesis. Massachusetts Institute of Technology, 1992.

29. L.J. Stifelman, B. Arons, C. Schmandt and E.A. Hulteen. VoiceNotes: A Speech Interface for a Hand-Held Voice Notetaker. *In Proceedings of* INTERCHI '93, pages 179-186. ACM SIGCHI, 1993.

30. J.B. Voor and J.M. Miller. The effect of practice upon the comprehension of time-compressed speech. *Speech Monographs*, 32:452-455, 1965.

JOURNAL OF THE AMERICAN VOICE I/O SOCIETY

VOLUME 16 **NOVEMBER, 1994**