Perceptual Adjustment to Highly Compressed Speech: Effects of Talker and Rate Changes

Emmanuel Dupoux  
Ecole des Hautes Etudes en Sciences Sociale–Centre National de la Recherche Scientifique and France Telecom

Kerry Green  
University of Arizona

This study investigated the perceptual adjustments that occur when listeners recognize highly compressed speech. In Experiment 1, adjustment was examined as a function of the amount of exposure to compressed speech by use of 2 different speakers and compression rates. The results demonstrated that adjustment takes place over a number of sentences, depending on the compression rate. Lower compression rates required less experience before full adjustment occurred. In Experiment 2, the impact of an abrupt change in talker characteristics was investigated; in Experiment 3, the impact of an abrupt change in compression rate was studied. The results of these 2 experiments indicated that sudden changes in talker characteristics or compression rate had little impact on the adjustment process. The findings are discussed with respect to the level of speech processing at which such adjustment might occur.

It is well known that the relationship between the speech signal and underlying phonological representations is extremely complex. This complexity is the result of several different factors. First, no one-to-one mapping between cues in the acoustic signal and the underlying representations have been uncovered so far. Instead, many different acoustic cues distributed in the signal have been found to map onto the same representation. Second, the nature of these cues varies as a function of many different factors including surrounding phonetic context, syllabic position, changes in talker, speaking rate, speaking style, accent, and stress. Despite this variation, speech perception is amazingly accurate. This is comparable to visual perception, in which widely different retinal stimulations give rise to perception of the same object, a phenomenon called perceptual constancy. To achieve constancy, the perceptual system presumably uses different normalization mechanisms that compensate for the variations induced by various situations such as overall lighting conditions, distance, orientation, and others (Epstein & Broo, 1986; Jolicouer, 1988).

Previous studies have suggested that, in speech, similar normalization mechanisms exist that adjust perceptual criteria in the different contexts. For example, consider the problem of mapping acoustic parameters onto vowel categories across speaking contexts. Peterson and Barney (1952) demonstrated that formant frequencies corresponding to a single vowel category vary extensively across different speakers. Moreover, the formant values for different vowel categories show considerable overlap, creating a difficult problem for theories of vowel perception. Several studies have shown that vowel perception is influenced by the pitch (approximately, the F0) of the talker’s voice (Bladon, Henton, & Pickering, 1984; Johnson, 1990; Nearley, 1989). Because F0 correlates with vocal tract size, some researchers have proposed that, during vowel perception, there is a stage of processing in which the formant values extracted from the acoustic signal are normalized with respect to the talker (Darwin, McKown, & Kirby, 1989; Johnson, 1990; Ladefoged, 1967, 1989; Ladefoged & Broadbent, 1957; Remez, Rubin, Nygaard, & Howell, 1987).

Additional evidence for normalization has been found with respect to the rate at which speech is articulated. It has been shown that talkers make large and frequent changes in their speaking rate during a conversation (Miller, Grosjean, & Lomanto, 1984). These changes have a dramatic impact on the realization of temporal acoustic cues such as voice onset time (VOT), transition duration, closure duration, and vowel duration (Miller, Green, & Reeves, 1986; Miller & Baer, 1983; Summerfield, 1981; Port, 1979). Experiments...
in speech perception have demonstrated that the perceptual system alters its criteria for judging such cues in relation to the rate at which the speech was produced. For example, Miller and Liberman (1979) examined the stop-semivowel distinction of /ba/ versus /wa/ as cued by transition duration. They found that deleting the last 216 ms of the vowel caused the critical transition that separated the /b/-/w/ categories to shift toward a shorter duration. Miller and Liberman interpreted their results as demonstrating that analysis of the transition duration was done in relation to the overall speaking rate of the test syllable. In support of this notion, Miller, Aibel, and Green (1984) found that modifying /ba/-/wa/ syllables by deleting the final vowel had systematic influences on the judged speaking rate of the test syllables. Tokens with shorter overall syllable durations were judged to be produced at a faster speaking rate. Those with shorter transitions were identified as /wa/ more frequently at the fast rate than at the slow rate. Similar results have been found for a voiced-voiceless distinction cued by VOT and closure duration (Green & Miller, 1985; Green, Stevens, & Kuhl, 1994; Port & Dalby, 1982; Summerfield, 1981). Studies with young infants show that they also demonstrate sensitivity to overall syllable duration when discriminating speech tokens (Eimas & Miller, 1980; Miller &

The purpose of the current study was to address these two questions by examining how listeners adjust to compressed speech. Compressed speech is an ideal vehicle for studying normalization mechanisms for a number of reasons. First, it provides a quantifiable way of manipulating the speech signal to create speech that is clearly outside the bounds of normal experience. This is important because it is only when speech exceeds such bounds that evidence is obtained for gradual adjustment over time. This does not mean that such adjustments do not take place with more typical signals, but that typical signals make it difficult to obtain evidence for gradual adjustments due to ceiling effects in the performance of the listener. Second, present-day signal processing techniques enable one to compress the signal without actually deleting portions of it or creating discontinuities as occurred with older compressors (Foulke &
General Method

Participants

Participants were undergraduate students at the University of Arizona. Some were given course credit for their participation; others were paid. All were native speakers of English. None of the participants reported any history of a speech or hearing disorder.

Materials

A set of 40 sentences was constructed using the following criteria: Each sentence consisted of 10 words with seven content words and three function words; no compound, rare, or foreign words were used; the overall length of each sentence was between 14 and 16 syllables; each sentence fit into one of five different syntactic frames; and none of the sentences was semantically anomalous. The semantic plausibility of the sentences was examined in a pilot experiment. The 40 sentences were presented on a single sheet of paper to a group of 14 participants who were asked to rate the plausibility of each sentence on a scale from 1 to 7.¹

The 34 sentences with the highest plausibility ratings (average of 5.4 to 6.7) were chosen for additional screening. These 34 sentences were spoken by a male talker, recorded onto audiotape, digitized into a lab computer (at 20 KHz sampling rate, 12 bits quantization) and compressed down to 35% of their original duration. These compressed sentences were presented to a second group of pilot participants for recall. Recall was assessed by having each participant listen to a sentence, and at the end of the sentence to write down every word that could be recalled. Guessing was encouraged. The 20 sentences obtaining a recall accuracy of approximately 50% (a total of 5 words with at least 3 content words) were selected for use in the experiments. The range for the 20 sentences was 39% to 62% correct. This screening procedure enabled us to select sentences that were roughly comparable in their recall scores (being neither particularly easy or difficult to recall at a compression rate of 35%). The list of 20 sentences is presented in the Appendix.

A male and a female talker were recorded (Electrovoice microphone: RE16, Tascam cassette deck: 122 MKII) while reading these sentences at different rates of speech in a quiet, sound-attenuated room. Each participant sat about 4 feet in front of the loudspeakers. The 34 sentences with the same partitioning being used for both talkers. The 20 sentences were partitioned into four sets of five sentences with the same partitioning being used for both talkers. The presentation order of these four sets of sentences was varied across different groups of participants according to a Latin square design. However, the presentation order within a set did not vary. The sentences were presented individually to participants who were asked to listen to a sentence and write down as much of it as they could recall, taking as much time as necessary. At the start of the experiment, each participant was presented with a single practice sentence (not one of the original set of 20) compressed to either 38% or 45% of its original duration. The participants were asked to listen to the practice sentence to get an idea of what the sentences would be like.

The 20 sentences were converted into four sets of five sentences with the same partitioning being used for both talkers. The presentation order of these four sets of sentences was varied across different groups of participants according to a Latin square design. However, the presentation order within a set did not vary. The sentences were presented individually to participants who were asked to listen to a sentence and write down as much of it as they could recall, taking as much time as necessary. At the start of the experiment, each participant was presented with a single practice sentence (not one of the original set of 20) compressed to either 38% or 45% of its original duration. The participants were asked to listen to the practice sentence to get an idea of what the sentences would be like.

The sentences were low-pass filtered (9.8 KHz), amplified (Yamaha AX 630), and then presented to individual participants over loudspeakers (Realistic minimus) in a small, sound-attenuated room. Each participant sat about 4 feet in front of the loudspeakers. The sentences were presented at a comfortable listening level (approximately 74 dB SPL).

¹ Plausibility, as used in this article, refers to how likely or ordinary an event is that is described by a sentence. Thus a plausible sentence will describe a very ordinary event that has a high probability of occurring in everyday life, whereas an implausible sentence will describe a very bizarre or unexpected event that is not very likely to occur.
Data Analysis

For each participant, the number of content and function words that were correctly recalled was determined for each sentence. The data described in the current study only report the percentage of content words that were correctly recalled. A narrow form of scoring was used, with even slight deviations from what was said being counted as errors, with the following exceptions. First, any form of the noun or verb was counted as correct. Second, added words were not counted as errors. 2

Experiment 1

As reported earlier, there is some indication, both anecdotal and experimental, that perceptual adjustment can occur to compressed speech. On the anecdotal side, participants often report that compressed speech sounds unnaturally fast at first and difficult to understand. With exposure, they find it easier to understand and report that it sounds less compressed. In addition, when switched back to uncompressed sentences, listeners report that the uncompressed speech sounds abnormally slow. Such anecdotal reports are suggestive of some kind of adjustment to rapid rates of speech. On the experimental side, there is the finding that, with practice, perceptual performance on highly compressed speech shows some measurable increase (Voor & Miller, 1965). This is expected if some type of adjustment to the compressed signal takes place. However, these earlier studies used a compression technique that involved deletion and subsequent concatenation of portions of the acoustic signal. Such effects tend to disrupt the speech signal. Using the less disruptive techniques, Mehler et al. (1993) showed that performances on compressed speech improve over time for a French-English bilingual. Such an improvement would provide evidence that experience with compressed speech results in perceptual adjustment or adaptation. As can be seen in Table 2, there was an increase in the recall scores across the first four sentence sets. To examine whether there was any evidence of an increase in recall rates across the different sentence sets, planned comparisons were conducted on the two compression rates. Because our interest was whether there was any improvement in adjustment or adaptation, the main factors were Compression Rate, Talker (male vs. female), and Set Position (sentence sets presented in Position 1, 2, 3, and 4). If there were any differences in the recall scores across participant groups, these were counterbalanced. Unless otherwise noted, F values not discussed in the text were not significant with p > .05. The cell means collapsed across participant groups are presented in the upper half of Table 2. An examination of this table reveals several things. First, there is an overall difference in the percentage of content words that were correctly recalled in the two compression rates. Not surprisingly, the sentences that were compressed to 45% of their original duration produced much higher accuracy in the recall scores of the content words. This resulted in a significant effect of Compression Rate,  F(1, 144) = 655.4, p < .0001.

Second, there is a small difference in the recall scores for the two different talkers with the female talker producing slightly higher recall scores,  F(1, 144) = 10.67, p < .005. This difference occurred at both compression rates; there was no interaction between Compression Rate and Talker,  F(1, 144) = 2.45, p > .12. It is not clear why the female talker’s utterances produced slightly higher recall scores. However, it is probably due to the female talker’s utterances being more formally produced than the male’s.

Of primary interest is whether there is any evidence of an increase in recall rates across the different sentence sets.

Results

The percentages of content words correctly recalled from each sentence were determined for each participant. These values were averaged over the five sentences of each set to yield four set scores for each participant. The set scores were then submitted to a single four-way analysis of variance (ANOVA) with Compression Rate (38% vs. 45%), Talker (male vs. female), Set Position (sentence sets presented in Position 1, 2, 3, and 4) and Counterbalancing (the 4 counterbalanced participant groups) as the main factors. Table 2. An examination of this table reveals several things. First, there is an overall difference in the percentage of content words that were correctly recalled in the two compression rates. Not surprisingly, the sentences that were compressed to 45% of their original duration produced much higher accuracy in the recall scores of the content words. This resulted in a significant effect of Compression Rate,  F(1, 144) = 655.4, p < .0001.

Second, there is a small difference in the recall scores for the two different talkers with the female talker producing slightly higher recall scores,  F(1, 144) = 10.67, p < .005. This difference occurred at both compression rates; there was no interaction between Compression Rate and Talker,  F(1, 144) = 2.45, p > .12. It is not clear why the female talker’s utterances produced slightly higher recall scores. However, it is probably due to the female talker’s utterances being more formally produced than the male’s.

Of primary interest is whether there is any evidence of an increase in recall rates across the different sentence sets.
Table 2
Average Percentage of Content Words Correct for Four Different Sentence Sets Across Two Talkers and Sentence Compression Rates

<table>
<thead>
<tr>
<th>Talker</th>
<th>First</th>
<th>Second</th>
<th>Third</th>
<th>Fourth</th>
<th>Talker M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Compression rate 38%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>26.9</td>
<td>31.9</td>
<td>34.9</td>
<td>35.0</td>
<td>32.2</td>
</tr>
<tr>
<td>Female</td>
<td>31.7</td>
<td>40.2</td>
<td>42.9</td>
<td>42.3</td>
<td>39.3</td>
</tr>
<tr>
<td>Set M</td>
<td>29.3</td>
<td>36.1</td>
<td>38.9</td>
<td>38.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Compression rate 45%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>61.5</td>
<td>74.7</td>
<td>74.2</td>
<td>78.6</td>
<td>72.2</td>
</tr>
<tr>
<td>Female</td>
<td>67.6</td>
<td>77.7</td>
<td>77.0</td>
<td>76.6</td>
<td>74.7</td>
</tr>
<tr>
<td>Set M</td>
<td>64.5</td>
<td>76.2</td>
<td>75.6</td>
<td>77.7</td>
<td></td>
</tr>
</tbody>
</table>

In summary, the results of this first experiment revealed a significant effect of Set Position, which indicated that recall scores only between the first and the second sets. This result is consistent with the findings of Mehler et al. (1993) and indicates that they are not due to the specific acoustic-phonetic characteristics of a particular talker. However, the overall performance of the participants listening to the two different talkers is very similar probably because the durations of the sentences for the talkers were closely matched before compression. Second, these results show that the form of the improvement depends on the nature of the experience. When the sentences are compressed to 38% of their original duration, they are extremely difficult to understand, and the adjustment process requires more time than when they are compressed to 45% of their original durations.

One important implication of this finding is that listeners' performance did not improve over the course of the experiment simply due to practice with the task itself (e.g., memorizing and recalling sentences under a time constraint). If this had been the only reason for their improvement, then similar rates of improvement should have been obtained for the two different compression rates.
A one-way ANOVA revealed a significant difference between the compressed condition and the other three comparisons (Neuman-Keuls) revealed a significant difference between the two compression rates, $F(3, 36) = 7.95$, $p < .001$. Pairwise comparisons (Neuman-Keuls) revealed a significant difference between the compressed condition and the other three ($p < .01$). None of the other groups differed significantly.

The results of this control experiment clearly demonstrate that the improvements measured in this first experiment were not due to general practice effects with the experimental task. Rather, they reflect adjustments specific to the compressed speech.

How then do we interpret the difference in rate of learning between the two compression rates? One possibility is that the ability to comprehend a message has an impact on the amount of improvement. If a large number of words can be recognized initially, then it is quite easy to engage in a strategy that reconstructs the missing words. Alternatively, if only a few words are recognized, a strategy based on reconstructing or “guessing” the missing words is much less effective. According to this notion, improvement in recognizing compressed speech would primarily reflect adjustments in higher level guessing strategies. The rate of adaptation should therefore depend on the number of recognizable words.

An alternative possibility is that the amount of exposure required is greater for the higher (38%) compression rate, because the tokens are more extreme examples of rapid speech. When perceivers encounter speech produced at different rates, they retune certain perceptual criteria (cf. Miller & Liberman, 1979). This retuning takes time, and the more extreme the speaking rate, the more time is required for the retuning process to be completed. Note that such a retuning process would be independent of the number of words initially recognized in the speech signal. More important, the shape of the recall function should depend on the compression rate of the speech and not on a particular participant’s overall performance level. Thus, the gradient of adjustment has a steeper slope for the 45% group than for the 38% group not because participants find the sentences compressed at 45% easier to recognize than the sentences compressed to 38%. Instead, the more highly compressed sentences require more time for complete normalization to occur.

To examine this possibility, overall performance across the 38% and 45% conditions was equated to determine whether similar learning curves occur. For each participant, the two sentences on which they performed best and the two sentences on which they performed worst were identified in each of the four sentence sets. Next, the participants’ best responses in the 38% compression condition and the worst responses in the 45% compression condition were extracted for further analysis. The mean percentage of content words correct in each of the four sentence sets obtained from these data are displayed in Figure 2. As shown in the figure, the performance level on these sentences in the two compression conditions is approximately equal. However, the shape of the function differs for the two compression rates. For the 38% condition, improvement is gradual over the first three sets. In contrast, a substantial improvement occurs between the first and second set in the 45% condition, with little change between the remaining sets. Analyzes similar to that performed in the Results section confirms this evaluation. Even though the performance in the two compression conditions was no longer significantly different ($F < 1$), there was still a significant Rate $\times$ Set Position interaction, $F(2, 288) = 9.29$, $p < .0002$. Separate ANOVAs were again performed on the data for each compression rate with planned comparisons between the individual set means. Again, there was a significant increase between Set 1 and Set 2 for both compression rates ($p < .001$); however, the increase between Set 2 and 3 was only significant for the 38% compression rate ($p < .001$).

These results demonstrate that the improvement in recall is primarily determined by compression rate and not by...
The percentage correct of content words for those participants with best recognition scores in the 38% compression group and the worst recognition scores in the 45% compression group. The data are collapsed across the two different talkers.

Experiment 2

The results of the first experiment support the claim that normalization of highly compressed speech does occur, that maximum normalization takes some time, and that the rate at a point at which performance was already significantly above baseline. The results of Experiment 1 indicate that not much improvement occurs after 10 sentences have been presented. Therefore, in this experiment a switch in talker was made after the presentation of the first 10 sentences.

To look at the effect of a switch in talker on the adjustment to compressed speech, three analyses were performed. The first examined whether there was a drop in performance due to the switch in talker. The second set of analyses compared the performance after the switch in talker to the performance of other groups of participants that had no switch but had varying amounts of experience with compressed speech, ranging from no prior experience to three full sets of sentences. Finally, to determine whether a switch in talker had a local effect, we ran similar analyses but restricted the data to the first two sentences after the switch. Note that the counterbalancing of the experiment was done only in sets of five sentences, not in terms of single sentences or pairs of sentences. This means that the interpretation of these more local analyses is tentative.

Method

Participants. A total of 60 new undergraduate students at the University of Arizona participated in the experiment. The participants were randomly assigned to one of six experimental groups.

Materials. The materials were the same four sets of five sentences spoken by the male and female talkers of Experiment 1. The sentences were compressed to 38% of their original duration.

Procedure. The participants in this experiment heard two sets of five sentences from one talker, and then a third set of sentences from the second talker. Despite the attempts to make the sentences comparable, there was some variability in their "recognizability" after compression. Therefore, it was necessary to counterbalance for the set of sentences presented as the third set. This was accomplished by comparing across four groups of participants. The first two groups (Groups A and B) were presented with Sets 1, 2, and 3, with the change in talker coming between Sets 2 and 3. The second two groups of participants (Groups C and D) were presented with Sets 1, 2, and 3, with the change in talker occurring
Table 3
Average Percentage of Content Words Correct for Two Different Sentence Sets as a Function of the Preceding Talker Context

<table>
<thead>
<tr>
<th>Initial talker</th>
<th>Set position</th>
<th>Second</th>
<th>Third</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Different talker</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>27.2</td>
<td>33.1</td>
<td>30.1</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>32.2</td>
<td>29.7</td>
<td>30.9</td>
<td></td>
</tr>
<tr>
<td>Set M</td>
<td>29.7</td>
<td>31.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Also presented in the table is performance on the first two sentences presented in the set. These data are discussed below.

Performance on the third set of sentences is lower for those participants for which there was a change in talker between the second and third sets (different talker) than for participants hearing the same talker throughout (same talker and complete context). However, performance in the different talker group was higher than that of the group of participants with no prior exposure to compressed speech. A three-way ANOVA was used to compare the means across
on the adjustment to compressed speech which is then compensated for very rapidly by the perceptual system.

Discussion

The results of this experiment found support for the idea that there are considerable savings on recall over a shift in talker when dealing with highly compressed speech. After a change in talker, participants' performances do not return to baseline. Indeed, the overall performance of groups experiencing talker change did not differ significantly from those of groups experiencing the same talker. Finally, a marginal decline in performance was seen in the first two sentences after a change in talker, indicating that changing talkers may have an immediate impact on the adjustment to compressed speech. However, the perceiver is able to recover from this within a couple of sentences and performance at the end of the set is comparable to that of the group of participants presented with just a single talker.

The results of this experiment contrast with others showing that changing talkers has a detrimental effect on many perceptual and memory tasks (Goldinger, Pisoni, & Logan, 1993; Martin, Mullennix, Pisoni, & Summers, 1989; Pisoni & Mullennix, 1989, 1990). This difference may mean that the adjustment to compressed speech is accomplished by mechanisms that are not tapped by these other tasks. The fact that there is considerable savings from one talker to another might indicate that adjustment to compressed

<table>
<thead>
<tr>
<th>Preceding context</th>
<th>Before change</th>
<th>All 5 sentences</th>
<th>First 2 sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different talker</td>
<td>29.7</td>
<td>31.4</td>
<td>23.9</td>
</tr>
<tr>
<td>Same talker</td>
<td>31.1</td>
<td>34.2</td>
<td>29.8</td>
</tr>
<tr>
<td>No preceding context</td>
<td>24.3</td>
<td>25.5</td>
<td>17.1</td>
</tr>
<tr>
<td>(no normalization)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete context</td>
<td>35.5</td>
<td>29.8</td>
<td></td>
</tr>
<tr>
<td>Same talker (full normalization)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Preceding Talker Context</th>
<th>After change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before change</td>
</tr>
<tr>
<td>Different talker</td>
<td>29.7</td>
</tr>
<tr>
<td>Same talker</td>
<td>31.1</td>
</tr>
<tr>
<td>No preceding context</td>
<td>24.3</td>
</tr>
<tr>
<td>(no normalization)</td>
<td></td>
</tr>
<tr>
<td>Complete context</td>
<td>35.5</td>
</tr>
</tbody>
</table>

Table 4
Average Percentage of Content Words Correct for Two Different Sentences as a Function of the Preceding Talker Context

The results of this experiment found support for the idea that there are considerable savings on recall over a shift in talker when dealing with highly compressed speech. After a change in talker, participants' performances do not return to baseline. Indeed, the overall performance of groups experiencing talker change did not differ significantly from those of groups experiencing the same talker. Finally, a marginal decline in performance was seen in the first two sentences after a change in talker, indicating that changing talkers may have an immediate impact on the adjustment to compressed speech. However, the perceiver is able to recover from this within a couple of sentences and performance at the end of the set is comparable to that of the group of participants presented with just a single talker.

The results of this experiment contrast with others showing that changing talkers has a detrimental effect on many perceptual and memory tasks (Goldinger, Pisoni, & Logan, 1993; Martin, Mullennix, Pisoni, & Summers, 1989; Pisoni & Mullennix, 1989, 1990). This difference may mean that the adjustment to compressed speech is accomplished by mechanisms that are not tapped by these other tasks. The fact that there is considerable savings from one talker to another might indicate that adjustment to compressed

on the adjustment to compressed speech which is then compensated for very rapidly by the perceptual system.

Discussion

The results of this experiment found support for the idea that there are considerable savings on recall over a shift in talker when dealing with highly compressed speech. After a change in talker, participants' performances do not return to baseline. Indeed, the overall performance of groups experiencing talker change did not differ significantly from those of groups experiencing the same talker. Finally, a marginal decline in performance was seen in the first two sentences after a change in talker, indicating that changing talkers may have an immediate impact on the adjustment to compressed speech. However, the perceiver is able to recover from this within a couple of sentences and performance at the end of the set is comparable to that of the group of participants presented with just a single talker.

The results of this experiment contrast with others showing that changing talkers has a detrimental effect on many perceptual and memory tasks (Goldinger, Pisoni, & Logan, 1993; Martin, Mullennix, Pisoni, & Summers, 1989; Pisoni & Mullennix, 1989, 1990). This difference may mean that the adjustment to compressed speech is accomplished by mechanisms that are not tapped by these other tasks. The fact that there is considerable savings from one talker to another might indicate that adjustment to compressed

on the adjustment to compressed speech which is then compensated for very rapidly by the perceptual system.

Discussion

The results of this experiment found support for the idea that there are considerable savings on recall over a shift in talker when dealing with highly compressed speech. After a change in talker, participants' performances do not return to baseline. Indeed, the overall performance of groups experiencing talker change did not differ significantly from those of groups experiencing the same talker. Finally, a marginal decline in performance was seen in the first two sentences after a change in talker, indicating that changing talkers may have an immediate impact on the adjustment to compressed speech. However, the perceiver is able to recover from this within a couple of sentences and performance at the end of the set is comparable to that of the group of participants presented with just a single talker.

The results of this experiment contrast with others showing that changing talkers has a detrimental effect on many perceptual and memory tasks (Goldinger, Pisoni, & Logan, 1993; Martin, Mullennix, Pisoni, & Summers, 1989; Pisoni & Mullennix, 1989, 1990). This difference may mean that the adjustment to compressed speech is accomplished by mechanisms that are not tapped by these other tasks. The fact that there is considerable savings from one talker to another might indicate that adjustment to compressed
sentences. This condition was added as an alternative way of examining the generalizability of the compression adjustment. In the third condition, listeners heard just one speaker, and the shift was from less compressed (50%) to compressed speech rather than from uncompressed to compressed speech. This condition was added to see whether any influences due to an abrupt change in compression rates would also be found with a less dramatic change in rate. Each of these groups was compared to a group of participants who heard the same sentences with no intervening uncompressed or 50% compressed sentences. This group was drawn from Experiment 1.

As in Experiment 1, drop rates of analysis were computed. In the first set, a potential drop in performance was tested by comparing scores before the change in compression rate to scores after the change. In the second set of analyses, the performance after the change was compared to conditions with no change in compression rate across the four sets and various amounts of prior exposure to compressed materials. Finally, we again looked for evidence of an immediate drop in performance on the first two sentences right after the change in compression rates.

Method

Participants. Sixty undergraduate students at the University of Arizona participated in the experiment. These participants were randomly assigned to one of six groups of 10 participants. They had not participated in Experiments 1 or 2.

Materials. Three of the four sets of sentences spoken by the male talker and compressed to 38% of their original duration (from Experiment 1) were used in this experiment. In addition, the fourth set of sentences spoken by the male and the female talker in their original uncompressed form, and the fourth set of sentences spoken by the male talker compressed to 50% of their original duration, served as stimuli in the study.

Procedure. The same general procedure used in Experiments 1 and 2 was followed in this experiment. Each participant heard two sets of five sentences compressed to 38% spoken by the male talker, followed by one set of alternate rate sentences (either uncompressed or 50% compressed), followed by a last set of sentences compressed to 38%, spoken by the male talker. The first two groups of participants were presented with the male talker's uncompressed sentences as the alternate rate set (henceforth, the "male--uncompressed" condition). The last two groups of participants talkeder, uncompressed--different talker, 50% compressed--same talker, 38% compressed--same talker) and Counterbalancing as between-subject factors and Set Position as a within-subject factor. As in the last experiment, the effect of Set Position was significant, $F(1, 72) = 9.66, p < .005$, reflecting the overall improvement in recall scores from the sentence set before the change in compression rate to the sentence set after the change in rate. Although there was a tendency for there to be less improvement in the conditions with a change in compression rate, this did not result in either a significant effect of Change Condition or a Change Condition $\times$ Set Position interaction (both $F$s $< 1.37, p > .25$). Thus, this first analysis indicated that an intervening rate had little effect on the recall of compressed sentences.

The second analysis compared the recall of the sentence set after a change in compression rate to three no change conditions (no preceding context, no change context, full context) using a two-factor ANOVA with Condition (six levels) and Counterbalancing (two levels) as between-subject factors. The ANOVA revealed a significant effect of Condition, $F(5, 108) = 4.04, p < .005$. Neither the effect of Counterbalancing nor its interaction with Condition was significant (both $F$s $< 1.48, p > .22$). Pairwise comparisons between the individual Condition means (Newman-Keuls test) revealed that the group having no preceding context (Number 5) recalled significantly less than all of the other five groups ($p < .01$), including those with a set of intervening sentences at a different compression rate. Thus, an intervening compression rate did not cause performance to drop to levels equal to the unadjusted state. None of the other contrasts reached significance ($p > .1$), indicating that performance after a change in compression rate did not differ from a condition with no change in rate.
Table 5
Average Percentage of Content Words Correct for the Sentence Sets Immediately Before and After the Intervening Alternate Rate Set and the Means for the Same Sentence Set for the Control Groups With No Preceding or Intervening Alternate Context

<table>
<thead>
<tr>
<th>Condition no.</th>
<th>Preceding context</th>
<th>Before change</th>
<th>After change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All 5 sentences</td>
<td>First 2 sentences</td>
</tr>
<tr>
<td>Change conditions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Uncompressed same talker</td>
<td>25.9</td>
<td>29.6</td>
</tr>
<tr>
<td>2</td>
<td>Uncompressed different talker</td>
<td>27.4</td>
<td>27.1</td>
</tr>
<tr>
<td>3</td>
<td>50% compressed same talker</td>
<td>27.2</td>
<td>32.4</td>
</tr>
<tr>
<td>No change conditions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>38% compressed same talker</td>
<td>25.7</td>
<td>31.7</td>
</tr>
<tr>
<td>5</td>
<td>No preceding context (no normalization)</td>
<td>20.4</td>
<td>12.9</td>
</tr>
<tr>
<td>6</td>
<td>Complete context same talker</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(full normalization)</td>
<td>32.8</td>
<td>23.9</td>
</tr>
</tbody>
</table>

Note. Also included are the means for: the first sentence, the first two sentences in the set following the alternate rate context.

Further, paired t-tests were calculated between the sentence set before a change in compression rate and the average of the first two sentences in the set immediately after a change in rate for the three change conditions and the 38% compression no-change condition (using a corrected alpha of .013). These tests indicated that a change to uncompressed speech produced a significant decline over the first two sentences of the set after a switch back to 38% compression, t(19) = 2.74, p < .013, and t(19) = 2.87, p < .01, for the same talker and different talker conditions, respectively. However, switching to only 50% compressed speech did not result in a significant decrease in performance, t(19) = 1.32, p > .20. Finally, the no-change condition also failed to show a decrease in performance over the same sentences, t(19) = .62, p > .54. Thus, a change in rate to normal speech also seems to have an immediate impact on the adjustment to compressed speech.

Discussion

Overall, these results indicated that adjustment to compressed speech did not cause a return to baseline performance when intervening uncompressed sentences were presented to the listener in either the same voice or a different voice. The presentation of compressed speech at an intermediate rate had a similar effect. However, there were indications of a small drop in performance due to the intervention of the uncompressed speech: The mean recall scores for the compressed sentences immediately following the uncompressed tokens were numerically lower than those of a comparable group of participants who did experience a change in compression rate. This small drop was due to lower performance on the first two sentences after the change, suggesting that a change in rate has a small and immediate impact on the adjustment process. However, this impact does not cause a complete resetting of adjustment parameters to baseline. This suggests that some kind of perceptual learning must be occurring. Finally, these results support the findings from Experiment 2 by again demonstrating that a change in talker has little effect on the adjustment to compressed speech even when that change is combined with a change in compression rates. Apparently, the perceptual mechanism that is responsible for compression adjustment, although influenced by large changes in compression rates, is not influenced by a change in talker characteristics.

General Discussion

The results of the first experiment established that performance improves with increased exposure to compressed speech. In addition, this improvement is specific to compressed speech in that prior exposure to uncompressed or noise degraded speech did not produce improved performance with compressed speech. The adjustment is not immediate, requiring experience with five or more sentences before full adjustment is obtained. Moreover, the time to reach a plateau in performance depends on the rate of compression: Highly compressed stimuli require more time for improvement than less compressed tokens. The second and third experiments extended this research by exploring how an abrupt change in either talker or compression rate or both influences improvement.

Experiment 2 demonstrated that, once listeners have adapted to the compressed speech of one talker, an abrupt switch to a different talker causes a small decline in performance immediately after the switch. Performance does not, however, drop to the level of participants having no prior experience with compressed speech. Moreover, the drop in performance is made up in the next few sentences, indicating that there are considerable savings across talkers. Experiment 3 investigated whether an abrupt shift in compression rates from highly compressed speech to uncompressed speech would influence the adjustment process. The results of the experiment revealed that an interruption by uncom-
pressed speech had little overall impact on the recall scores for compressed speech. There was some evidence of a decrement in performance in the first two sentences after the switch back to compressed speech. However, the decrement was small and by the end of the set, performance had fully recovered. Thus, the adjustment after a change in compression rate or a change in talker is much more rapid than observed on the first exposure to these compressed sentences. These two findings indicate that there are savings from an earlier exposure to compressed speech, although further experiments are necessary to determine the exact nature of such savings and how long it might last.

The results of this study raise the question of the level of speech processing that is responsible for the improvement in the recall scores with highly compressed speech. Speech compressed at high rates presents a challenge to the listener's processing system at a number of different levels. First, the robustness of various acoustic cues in the signal is severely reduced due to their brief durations. Some cues that are extremely short to begin with (such as release bursts) may be too brief for the perceptual system to detect. In addition, the underlying temporal characteristics of cues such as VOT or transition duration are also dramatically altered. Because the compression algorithm compresses all aspects of the speech signal equally, the temporal relationship between cues such as VOT or closure duration to vowel information from compressed speech is severely jeopardized. Thus, the adjustment after a change in compression rate or a change in talker is much more rapid than observed on the first exposure to these compressed sentences. These two findings indicate that there are savings from an earlier exposure to compressed speech, although further experiments are necessary to determine the exact nature of such savings and how long it might last.

The results of this study raise the question of the level of speech processing that is responsible for the improvement in the recall scores with highly compressed speech. Speech compressed at high rates presents a challenge to the listener's processing system at a number of different levels. First, the robustness of various acoustic cues in the signal is severely reduced due to their brief durations. Some cues that are extremely short to begin with (such as release bursts) may be too brief for the perceptual system to detect. In addition, the underlying temporal characteristics of cues such as VOT or transition duration are also dramatically altered. Because the compression algorithm compresses all aspects of the speech signal equally, the temporal relationship between cues such as VOT or closure duration to vowel information from compressed speech is severely jeopardized. Thus, the adjustment after a change in compression rate or a change in talker is much more rapid than observed on the first exposure to these compressed sentences. These two findings indicate that there are savings from an earlier exposure to compressed speech, although further experiments are necessary to determine the exact nature of such savings and how long it might last.

Another possibility is that the improvement reflects adjustments occurring at early levels of speech processing in which the acoustic signal is mapped onto underlying phon
roughly maintained despite the presentation of intervening materials spoken at a very different rate.

We speculate that adjustment to compressed speech may be the result of two processes operating simultaneously: a short-term adjustment to local speech rate parameters (that would be related to rate normalization in phonetic processing) and a longer term adjustment reflecting a more permanent perceptual learning process. This long-term adjustment would operate on a level of representation abstract enough that the acoustic differences between talkers no longer matters. As in other cases of perceptual learning (e.g., Schwab et al., 1985), it would involve a long-term memory component.

What could be the purpose of such a long-term adjustment? One possibility is that it is used to compensate for the fact that talkers vary in their overall articulation rates (Miller, Grosjean, & Lomanto, 1984) and in the way in which they realize their articulatory targets according to speech rate, phonological context, and dialect. To operate under optimal conditions for all talkers, the perceptual system may extract a set of phonetic–phonological parameters that work best in a given situation and store them for later use.

In perception of compressed speech, some of the cues for a phonetic segment might be lost in the compression process or misapplied to another segment. Alternatively, even when phonetic information is recovered correctly, it could be inappropriate to the current speech rate. For example, the compression algorithm does not flap or spread nasality. However, the phonological–lexical system might expect these phonetic changes to be present at these speaking rates and attempt to overcorrect for them. Therefore, it may be that the phonological–lexical component has to adjust for what is essentially a new speech style, perhaps in the same way that it handles foreign accents or dialects. This hypothesis would account for the fact that, once the adjustment to fast speech has been learned for a given talker, there will be (a) considerable savings for a different talker (at least, of the same dialect) and (b) savings that persist over an extended period of time (the only cost being in retrieving the phonetic–phonological mapping rules appropriate for that particular input on later occasions).

In summary, adaptation to compressed speech is an interesting phenomenon that raises many questions related to speech perception and spoken language processing. The purpose of the current study was to establish the phenomenon and suggest some directions for future research. We expect that examining how listeners adjust to highly compressed speech will provide further insight into the mechanisms involved in spoken language processing.

References


Martin, L. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V.
ADJUSTING TO COMPRESSED SPEECH

927


Appendix

Materials

The 20 sentences used in the experiment:

1. Mother yelled at Billy because he never eats fresh fruit.
2. During home football games we drink lemonade and eat peanuts.
3. Sally wanted fresh fish but could only buy it frozen.
4. Every girl who phoned Mary heard exactly the same story.
5. After a hard day's work William goes directly to bed.
6. A man who loved fresh vegetables wanted to grow cucumbers.
7. Behind the wooden fence Sam watched the construction crew working.
8. The oranges that grow in Nancy's yard taste very bitter.
9. Randy has never left the country or even the state.
10. A painting that cost a million dollars was stolen.
11. Above Susan's head large black spiders crawled on the ceiling.
12. From disease and rot the large oak timbers finally collapsed.
13. The woman who won the lottery last week quit working.
14. An apple that is baked without cinnamon tastes very bland.
15. The famous author who opened the program told clever jokes.
16. Linda called home shortly after she got her big promotion.
17. The angry teacher made Tommy write sentences on the board.
18. Many passengers saw the cruise ship strike the fishing boat.
19. Mark only makes mistakes when the supervisor checks his work.
20. Near the old red building many different desert plants grow.