Journal of Experimental Psychology:
Human Perception and Performance
1997, Vol. 23, No. 3, 914–927


0096-1523/97/$3.00

# 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 & Broota, 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; Nearey, 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, McKeown, & 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

Emmanuel Dupoux, Ecole des Hautes Etudes en Sciences Sociale–Centre National de la Recherche Scientifique and France Telecom, Paris, France; Kerry Green, Departments of Psychology and Cognitive Science, University of Arizona.

This research was supported in part by the Fyssen Foundation, Institut National de la Recherches en Informatique et Automatique, Centre National de la Recherches en Telecommunications, and the Human Frontier Science Project.

We thank France Telecom for providing the algorithm used to compress the stimuli. We also thank Tom Bourgeois for his help in creating the stimuli and Kathy Fohr, Jonathan Beory, and Zoe Forester for collecting, scoring, and summarizing all the data. Both authors contributed equally to the project, and the order of authorship was determined by the flip of a coin.

Correspondence concerning this article should be addressed to either Emmanuel Dupoux, Laboratoire de Science Cognitive et Psycholinguistique, 54 Boulevard Raspail, 75006 Paris, France, or Kerry Green, Cognitive Science, Psychology Building, Room 312, University of Arizona, Tucson, Arizona 85721. Electronic mail may be sent via Internet to Emmanuel Dupoux at dupoux@lscp.msh-paris.fr or to Kerry Green at kgreen@u.arizona.edu.

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 & Eimas, 1983). Thus, adjustment to rate-induced variation occurs at a very early stage in development. However, it is not yet known whether such adjustment involves a speech-specific mechanism, because similar results can be obtained in primates (Stevens, Kuhl, & Padden, 1988) and with nonspeech signals (Diehl & Walsh, 1989; Fowler, 1990; Pisoni, Carrell, & Gans, 1983).

Normalization appears to occur immediately in response to local variation in the speech signal. However, the question arises as to whether short-term mechanisms are the only ones responsible for perceptual constancy in speech, or whether there exist other mechanisms with different properties. For example, there are indications that perceptual adjustments occur in response to changes in speaking rate over longer intervals than a syllable such as a precursor phrase or sentence provides (Kidd, 1989; Miller, Green, & Schermer, 1984; Summerfield, 1981). However, even this phenomenon is fairly local, occurring within a few seconds of the changes in the immediately preceding context.

On a more anecdotal level, there is the experience of listening to a new talker with a heavy accent or dialect who is difficult to comprehend at first. After a couple of minutes, however, comprehension improves and it becomes possible to understand the talker. Schwab, Nusbaum, and Pisoni (1985) found a similar improvement in comprehension of synthetic speech. This suggests that the adjustment to a new speaking style occurs gradually, with recognition improving over several minutes of experience for a particular type of speech signal. A similar situation appears to occur when listeners are presented with highly compressed speech. For example, Voor and Miller (1965) found that comprehension of compressed speech increased significantly over the first 8 to 10 min of listening, with little increase after that. Two questions can be raised with respect to these findings. First, is the slow improvement the result of perceptual normalization mechanisms that are used to achieve perceptual constancy? Or does it reflect the operation of more cognitive strategies such as learning to guess words based on partial information? Second, if the locus of these effects is perceptual (as opposed to cognitive), what is the relationship between such a mechanism and the rate normalization mechanisms that have been previously examined in phonetic perception?

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 & Sticht, 1969; de Haan, 1982; Heiman, Leo, Leighbody, & Bowler, 1986; Schmitt, Moore, & Lass, 1986). Thus, in the current study, the compressed speech signal is smooth and its spectral characteristics are unchanged. Finally, compressing speech affects the perceived rate at which the speech was produced. It is therefore possible to compare the normalization that occurs to highly compressed speech with the normalization that occurs with more typical changes in speaking rate. Such a comparison is important for providing a complete understanding of the adjustment mechanisms involved in phonetic perception.

The three experiments in the current study address the nature of the perceptual adjustments to highly compressed speech. The first experiment uses a carefully constructed set of speech tokens along with a counterbalanced design to demonstrate that adjustment to compressed speech can occur. More important, this experiment also establishes the amount of exposure needed to obtain improved recall scores of the compressed sentences. This information is necessary for the next two experiments, which examine the effect of changing talkers and the preceding rate context on the adjustment process. Both of these variables have already been investigated with respect to their impact on local rate normalization during phonetic perception (Diehl, Souther, & Convis, 1980; Kidd, 1989; Summerfield, 1981). Experiment 2 examines whether the improvement that occurs generalizes to new talkers. Experiment 3 investigates whether the performance is affected by the intervention of uncompressed materials. These experiments constitute an initial examination of the locus of adjustments to compressed speech during processing and its relationship to previous findings in rate normalization. We first describe the stimuli and the general methodology used in the experiments.

## 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. These tokens were low-pass filtered at 9.8 KHz and then digitized into a lab computer (NEC 386-20) at a 20 KHz sampling rate (12 bits quantization). The overall durations of the sentence tokens were measured for each talker, and sentences were selected that had similar durations across the set of 20 sentences and the two talkers (see Table 1). This was to reduce any potential interactions of speaking rate and compression. Each of the 20 sentences for the two talkers was then copied into its own file. Finally, these two sets of 20 sentences were compressed to 38% and 45% of their original durations using a computer program developed by French Telecom based on the PSOLA algorithm (Charpentier & Stella, 1986). The program operates by first segmenting the signal into consecutive pitch periods. Unvoiced portions of speech are blindly segmented into chunks equal to the size of the average pitch period. Next, adjacent pitch periods are averaged in the time domain according to a scheme that depends on the compression rate. The averaging function is slightly longer than the width of each pitch period with smoothing functions on either side of the averaging window. The end result is a signal with fewer pitch periods than the original. For example, a signal compressed at a 50% compression rate will end up with only half the number of pitch periods as the original signal. However, because the information is averaged across adjacent pitch periods and not simply deleted, the signal retains many of the brief acoustic events such as release bursts that are important to phonetic perception.

Table 1
*Duration Values and Articulation Rates of the 20 Sentences Spoken by the Male and the Female Talker*

| Sentence no. | No. of syllables | Duration (seconds) | | Articulation rate (syllables/second) | |
|---|---|---|---|---|---|
| | | Male | Female | Male | Female |
| 1 | 14 | 2.42 | 2.44 | 5.79 | 5.74 |
| 2 | 15 | 2.54 | 2.44 | 5.91 | 6.15 |
| 3 | 14 | 2.36 | 2.26 | 5.93 | 6.19 |
| 4 | 15 | 2.50 | 2.54 | 6.00 | 5.91 |
| 5 | 14 | 2.45 | 2.42 | 5.71 | 5.79 |
| 6 | 16 | 2.67 | 2.64 | 5.99 | 6.06 |
| 7 | 15 | 2.55 | 2.57 | 5.88 | 5.84 |
| 8 | 16 | 2.43 | 2.36 | 6.58 | 6.78 |
| 9 | 14 | 2.22 | 2.20 | 6.31 | 6.36 |
| 10 | 16 | 2.53 | 2.48 | 6.32 | 6.45 |
| 11 | 14 | 2.74 | 2.99 | 5.11 | 4.68 |
| 12 | 15 | 2.78 | 2.83 | 5.40 | 5.30 |
| 13 | 14 | 2.26 | 2.24 | 6.19 | 6.25 |
| 14 | 15 | 2.56 | 2.65 | 5.86 | 5.66 |
| 15 | 15 | 2.66 | 2.71 | 5.64 | 5.54 |
| 16 | 15 | 2.56 | 2.48 | 5.86 | 6.05 |
| 17 | 15 | 2.49 | 2.43 | 6.02 | 6.17 |
| 18 | 14 | 2.55 | 2.50 | 5.49 | 5.60 |
| 19 | 14 | 2.72 | 2.78 | 5.15 | 5.04 |
| 20 | 14 | 2.66 | 2.60 | 5.26 | 5.38 |

### Procedure

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 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 talker as well as a Spanish–Catalan bilingual.

The purpose of Experiment 1 was to replicate this finding using the speech of two monolingual English talkers and to provide a baseline for Experiments 2 and 3. In addition, this first experiment extended the study of Mehler et al. by documenting in more detail the gradient of adaptation as a function of the amount of experience with compressed speech. Finally, two different compression rates were used to determine the generality of the results.

### Method

*Participants.* One hundred sixty undergraduate students at the University of Arizona participated in the experiment. Each participant was randomly assigned to 1 of 16 groups of 10 participants.

*Materials.* The materials were the four sets of five sentences spoken by a male and female talker, compressed to 38% and 45% of their original durations.

*Procedure.* Eight groups of participants listened to the sentences produced by the male talker, and eight groups heard the sentences of the female talker. For each talker, half of the groups heard sentences at the 38% compression rate and the remainder at the 45% compression rate. Finally, the presentation order of the four sentence sets was counterbalanced across the four groups receiving a particular compression rate, using a Latin square design.

### 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. 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. Such an improvement would provide evidence that experience with the compressed speech resulted 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 for both talkers at both compression rates, $F(3, 432) = 76.9, p < .0001$. Set Position interacted with Compression Rate, $F(3, 432) = 3.42, p < .02$, because, whereas the recall scores for the 38% compression rate increase from the first to the second and from the second to the third sentence sets, for the 45% compression rate, there is only an increase from the first to the second sentence set. To examine this interaction further, separate three-way ANOVAs were conducted on the two compression rates. Because our specific interest was whether there was any improvement in recall scores from one sentence set to the next, planned comparisons were carried out on pairwise set means for each compression rate.

At the 38% compression rate, the effect of Set Position was significant, $F(3, 216) = 26.19, p < .0001$. However, the planned comparisons indicated significant differences only between the cell means for Set 1 and Set 2 ($p < .0001$) and between Set 2 and Set 3 ($p < .05$). There was also a significant

---

[2] That is, missing or added plurals on the noun or verbal inflections were not counted as errors, and additions such as "every one" versus "every" were scored as correct.

Table 2
*Average Percentage of Content Words Correct for Four Different Sentence Sets Across Two Talkers and Sentence Compression Rates*

| Talker | Set order: First | Second | Third | Fourth | Talker *M* |
|---|---|---|---|---|---|
| | Compression rate 38% | | | | |
| Male | 26.9 | 31.9 | 34.9 | 35.0 | 32.2 |
| Female | 31.7 | 40.2 | 42.9 | 42.3 | 39.3 |
| Set *M* | 29.3 | 36.1 | 38.9 | 38.7 | |
| | Compression rate 45% | | | | |
| Male | 61.5 | 74.7 | 74.2 | 78.6 | 72.2 |
| Female | 67.6 | 77.7 | 77.0 | 76.6 | 74.7 |
| Set *M* | 64.5 | 76.2 | 75.6 | 77.7 | |

effect of Talker, $F(1, 72) = 13.24$, $p < .001$, but no Set Position by Talker interaction, $F(3, 216) = .84$, $p > .45$.

In the 45% compression condition, there is an increase in recall scores only between the first and the second sets. Specifically, there is a significant effect of Set Position, $F(3, 216) = 56.49$, $p < .0001$. However, the planned contrasts showed that only the difference between Set 1 and Set 2 was reliable ($p < .0001$). There was also a significant Talker × Set Position interaction, $F(3, 216) = 4.24$, $p < .01$. The nature of this interaction can be seen in Table 2. The male talker's sentences in the first set were harder to recall than the female's (61.5 versus 67.5 for the male and female talker, respectively). However, recall reaches the same level for both talkers by the end of the second set.[3]

In summary, the results of this first experiment revealed a significant increase in the number of content words that were correctly recalled across the different sentence sets. For the 38% compression rate, the increase occurred over the first three sets whereas in the 45% compression rate the increase only occurred between the first two sentence sets. Thus, as shown in Figure 1, the two compression rates produce different rates of improvement with the 45% rate approaching a plateau faster than the 38% rate. Furthermore, although the total number of words recalled differed for the two talkers, the pattern of improvement across sentence sets is remarkably similar in both compression rates for the two talkers. This similarity demonstrates the generality of the findings and indicates that they are not due to the specific acoustic–phonetic characteristics of a particular talker.

### *Discussion*

The results from this experiment indicate that improvement in recall of compressed speech can occur with exposure to only 5 or 10 compressed sentences. These results replicate the earlier findings of Mehler et al. (1993) and provide empirical support to the anecdotal reports of those who have listened to compressed speech. In addition, this experiment extends the findings of Mehler et al. in two ways. First, similar findings were obtained for two new talkers of English. Moreover, 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.

To rule out definitively that task familiarity might play some role in performance, we conducted an additional control experiment. Four new groups of 10 participants each were presented with a set of five test sentences compressed to 38%. Before the critical set of test sentences, each group of participants heard a set of context sentences. The first group of participants heard the remaining three sentence sets compressed to 38%. We call this the compressed condition, and it serves as a replication of one of the earlier participant groups. The second group of participants heard the remaining sentence sets uncompressed. These sentences were easy to comprehend and provide a context in which participants can practice the recall task without getting any exposure to compressed speech. We call this group the uncompressed condition. The third group of participants heard the uncompressed sentences in white noise at a signal-to-noise ratio of −1dB.[4] Pilot testing had determined that this signal-to-noise ratio resulted in an overall intelligibility for the uncompressed speech that was comparable to that of the 38% compressed speech. This condition was included to determine whether the adaptation is simply the result of learning how to deal with speech that is degraded in some fashion. This group is called the noise condition. Finally, a fourth group was included that had no prior experience with either compressed or uncompressed speech. This group provided a baseline level of performance on the critical sentence set without any previous context. This group is called the baseline condition.

If practice with the task is sufficient to account for improvement, then the first three conditions should result in significantly better performance on the critical sentence set over the baseline condition. However, if improvement is specific to experience with compressed speech, performance in the compressed condition only should be above baseline.

The results of this control experiment are that the recall scores in the compressed condition (26%) are higher than

[3] The analyses of both compression rates revealed significant interactions between Counterbalancing, Set Position, and Rate ($p < .0001$ for both analyses). These higher order interactions are difficult to interpret; however, they appear to reflect the fact that the different sentence sets had different recognition performance for which the counterbalancing in the experiment controlled.

[4] We thank Arthur Samuel for suggesting this control condition.

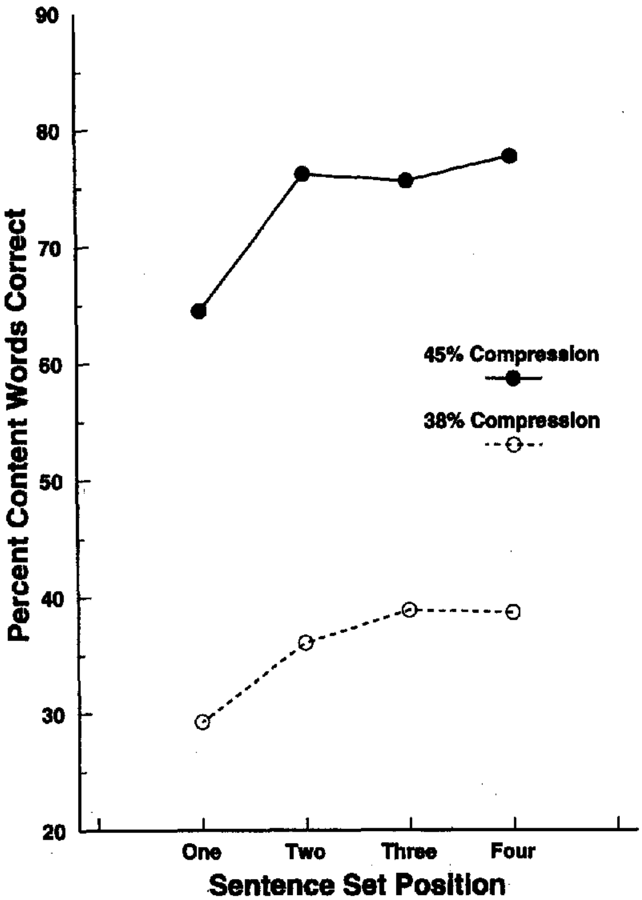


*Figure 1.* The percentage correct of content words for the four different sentence sets in the 38% and the 45% compression groups. The data are collapsed across the two different talkers.

the scores in the remaining three conditions (11%, 12%, and 13% for the uncompressed, noise, and baseline conditions, respectively). A one-way ANOVA revealed a significant effect of condition, $F(3, 36) = 7.95$, $p < .001$. 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. Analyses 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 × 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 ($p < .001$).[5]

These results demonstrate that the improvement in recall is primarily determined by compression rate and not by

[5] In the first set of five sentences, the mean sequential position of the best 38% was 3.6, and for the worst 45% sentences, it was 2.4. In the other sets of five sentences, the mismatch was in the same direction, but did not exceed 1 sentence position. It is unlikely that this small difference in sentence position could account for the pattern of improvement across compression rates.

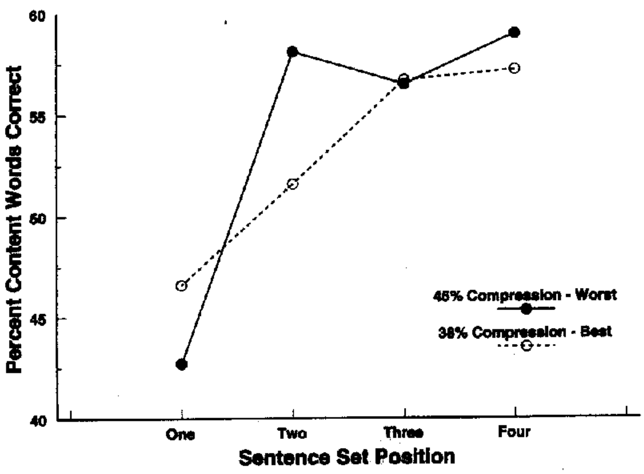


*Figure 2.* The percentage correct of content words for those participants with best recognition scores in the 38% compression group and the worst recognition scores in the 45% compression group. The data are collapsed across the two different talkers.

overall performance level. Even when performance levels are equated, the same difference in the slope of the adjustment function occurs. This finding is inconsistent with the view that high-level guessing strategies account for the differences between the two compression rates. However, the results are also consistent with the idea that at least some of the improvement in the ability to recognize compressed speech is due to perceptual mechanisms involved in the normalization for speaking rate.

## 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 of improvement is a function of the compression rate. The purpose of the second experiment was to explore the nature of that normalization. Specifically, this experiment addressed the question of whether experience with one talker transfers to a different talker. The issue of transfer is important because studies have found that changes in speaker identity have an impact on the early levels of speech processing. Green, Tomiak, and Kuhl (in press) have shown that the underlying dimensions of talker and speaking rate are processed in an integral manner. Studies examining how the speaking rate of a precursor phrase influences the phonetic identification of a test syllable indicate that little or no transfer occurs when the precursor phrase and test utterance are spoken by different talkers (Diehl, Souther, & Convis, 1980). It is therefore possible that the kind of normalization measured in Experiment 1 may be talker specific. If so, then participants who have listened to one talker's sentences at a highly compressed rate should drop to the level of their original performance (or even lower) when switched to a different talker.

To test this hypothesis, it was necessary to switch talkers 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, 3, and 2, with the change in talker occurring between Sets 3 and 2. Participant Groups A and C heard the male sentences first, followed by the female sentences. Participant Groups B and D heard the female sentences first.

In addition, four control groups of participants were needed for comparison purposes. These groups were presented with the sentences in the same order but with no change in talker. Two of these groups heard the sentence sets in the order 1, 2, 3. One group was presented with the female sentences and the other group with the male sentences. These data were taken from Experiment 1. In addition, two new control groups were run. These participants heard the sentence sets in the order 1, 3, 2. Again, one group was presented with the male sentences and the other group with the female sentences. As in Experiment 1, each group consisted of 10 participants.

### Results

Table 3 shows the mean percentage of content words correctly reported for the second and third sentence sets for those participants who were presented with a change in talker between these two sentence sets (the "different

Table 3
*Average Percentage of Content Words Correct for Two Different Sentence Sets as a Function of the Preceding Talker Context*

| Initial talker | Set position: Second | Third | *M* |
|---|---|---|---|
| | Different talker | | |
| Male | 27.2 | 33.1 | 30.1 |
| Female | 32.2 | 29.7 | 30.9 |
| Set *M* | 29.7 | 31.4 | |
| | Same talker | | |
| Male | 25.9 | 31.6 | 28.7 |
| Female | 36.3 | 36.8 | 36.5 |
| Set *M* | 31.1 | 34.2 | |

talker" condition), and for those participants in the control groups in which there was no change in talker (the "same talker" condition). As can be seen in the table, changing talkers appears to have little or no impact on the listeners' performance on these highly compressed sentences. These data were analyzed using a four-way ANOVA with Change Condition (change in talker, no change), Talker (male or female), and Counterbalancing as between-subject factors and Set Position as a within-subject factor. As in the first experiment, the effect of Set Position by itself was significant, $F(1, 72) = 4.03$, $p < .05$. This result confirms the earlier finding that improvement in recalling these highly compressed sentences occurs between the second and third set of sentences. Although there is a numerical trend toward less improvement in the different talker condition versus the same talker condition, neither the effect of Change Condition, $F(1, 72) = 1.18$, $p > .28$, nor the Change Condition × Set Position interaction was significant, $F(1, 72) = .414$, $p > .52$.

Although there was no significant effect of Talker, $F(1, 72) = 3.29$, $p > .07$, there was a significant interaction of Talker and Change Condition, $F(1, 72) = 4.98$, $p < .03$, and Change Condition by Talker by Set Position, $F(1, 72) = 8.24$, $p < .001$. These results indicate a difference between the two talkers as a function of a change in talker. Specifically, changing from the female talker to the male resulted in a slight drop in recall whereas a change from the male talker to the female produced a small increase. These effects can be partially accounted for by the fact that the sentences of the female talker were easier to recognize than the sentences by the male (see Experiment 1). The main effects turn into interactions due to the counterbalanced design. Finally, the remaining interactions, as well as the Counterbalancing factor, were not significant, $Fs < 1.18$, $p > .27$.

A second analysis compared the performance in the set of sentences after a change in talker (the Different Talker condition) to three other conditions taken from Experiment 1: a no preceding context condition, a same talker condition in which the sentences were preceded by two sentence sets by the same talker, and a complete context in which the sentences were presented after three sentence sets of the same talker. The means for these conditions are presented in 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 the four experimental conditions with Condition (4 levels), Talker, and Group as between-subjects factors. The ANOVA revealed a significant effect of Talker, $F(1, 144) = 11.23$, $p < .001$, again due to the female talker producing higher recall than the male talker sentences. The ANOVA also revealed a significant effect of Condition, $F(3, 144) = 9.42$, $p < .0001$. None of the other factors or interactions was significant (all $Fs < 1.42$, $p > .23$).

A pairwise comparison (Newman-Keuls) between the different talker condition and the no preceding context condition was significant, indicating that a change in talker resulted in performance that was still better than baseline. Comparisons among the different talker, same talker, and complete context conditions were not significant. Thus, performance after a change in talker is not worse than performance after no change in talker.

It is possible, however, that there is a more local or immediate decline in the recall of compressed speech that only occurs in the first couple of sentences following the change in talker. To examine this possibility, recall performance was examined for the first two sentences in the set presented right after the change in talker (see Table 4).[6]

In the different talker condition, performance on the first two sentences after the change (24%) was lower than the average performance on the set just before the change (30%). Moreover, performance on these first two sentences was also lower than the performance on the same two sentences in the same talker condition (30%), which showed little decline relative to the preceding set (31%). Thus, it appears that there is a local decline in performance immediately after a change in talker, which then shows rapid recovery over the last few sentences in the set. To examine this further, paired *t* tests were calculated between the sentence set before a change in talker and the average of the first two sentences in the set immediately after a change in talker for the different talker and same talker conditions. These tests indicated that the decline in performance during the first two sentences after a change in talker was nearly significant, $t(19) = 2.56$, $p < .03$ using a corrected alpha level of .025, whereas the corresponding test on the same talker condition was not, $t(19) = .55$, $p > .58$. Thus, changing talkers seems to have a small, immediate impact

[6] We decided to average over the first two sentences to reduce the influence that a single sentence might have. In addition, to provide a good baseline comparison of the participants' performance before the change in rate, the average of the five sentences in the set presented before the rate change was used.

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

| | | After change | |
|---|---|---|---|
| Preceding context | Before change | All 5 sentences | First 2 sentences |
| Different talker | 29.7 | 31.4 | 23.9 |
| Same talker | 31.1 | 34.2 | 29.8 |
| No preceding context (no normalization) | | 24.3 | 17.1 |
| Complete context | | 35.5 | 29.8 |
| Same talker (full normalization) | | | |

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 speech is performed at a rather abstract level, that is, a level at which detailed acoustic characteristics of the signal no longer play a role. An alternative account, which we favor, is that at least some of the adjustment to compressed speech does involve perceptual mechanisms involved in rate normalization. Moreover, switching talkers does produce an immediate decline in performance, but the effect is compensated for very rapidly by the perceiver. However, further investigation of the time course of adjustment on a sentence-by-sentence basis may be necessary to show such immediate effects.

## Experiment 3

The previous two experiments examined adjustment as a function of exposure to highly compressed speech and explored whether the adjustment transfers from one talker to a new talker. However, they did not address the durability of the adjustment, that is, once a plateau in performance is attained, whether it is maintained despite the need to adjust to other rates of speech. This question is of interest because it relates to the nature of the underlying mechanism. If the phenomenon is due to local rate normalization, then the adjustment should be affected by abrupt changes in the compression rates of the sentences. For example, the normalization might reset after the presentation of uncompressed materials. Alternatively, there may be some savings across different rates, in which case part of the adjustment process may be related to more permanent learning.

Consider an analogy from the visual modality, the perception of contrast with respect to dark adaptation. The perception of contrast is directly influenced by the sensitivity of the visual system to the ambient light level. However, we often experience large changes in the overall levels of ambient light in the course of a normal day. Therefore, our visual systems have to be able to operate over a wide range of lighting conditions with extreme sensitivity. The visual system meets these two requirements by adopting a narrow operating range for intensity with good contrast sensitivity, and then adjusting that operating range to match the level of illumination in the environment, through an adaptation mechanism (Goldstein, 1989; Hood & Finkelstein, 1986). When an observer moves from a dark environment to a light environment (or vice versa), adaptation takes place over a 5–10 min period of time. During that time, the visual system adjusts its operating range to the new light level, and the perception of contrast can be impaired. Once a state of adaptation has been attained, the perceptual system can quickly adjust to smaller amounts of illumination changes within a particular scene.

If adjustment to compressed speech is analogous, then after adjustment has occurred, presentation of uncompressed speech should cause the perceptual system to reset to the normal range of speech rates. A switch back to compressed speech should then trigger a readjustment. Alternatively, normalization to compressed speech may involve some kind of perceptual learning (e.g., Pisoni & McClaskey, 1983; Schwab et al., 1985) in which the phonetic parameters adequate for dealing with new or degraded stimuli are extracted and stored for later use. If so, some savings may be maintained through intervening uncompressed materials.

This last experiment asked whether normalization, once attained, can be affected by intervening speech that is either uncompressed or less compressed. Participants first heard 10 sentences compressed to 38% of their original duration. Next, five new uncompressed or less compressed sentences were presented, followed by five new sentences at the original 38% compression rate. As in Experiment 2, the participants were assigned to either of two groups, with the sentence material before and after the change in compression rates counterbalanced across the two groups. Three conditions were compared. In the first, all sentences, compressed and uncompressed, were produced by one talker. In the second, the talker changed during the uncompressed 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 2, three sets of analyses were conducted. 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 second two groups of participants were presented with the female talker's uncompressed sentences as the alternate rate set (henceforth the "female–uncompressed" condition). The last two groups of participants heard the male talker's sentences compressed to only 50% as the alternate rate set (the "male–50% compressed" condition). The order of the sentence sets appearing before and after the change in compression rates was counterbalanced across the two groups in each condition.

### Results

Table 5 presents the mean percentage of content words correctly reported from the set of 38% compressed sentences presented immediately before a change in compression rate (before change) and the set presented immediately after the alternate rate sentences (after change), for the three rate-change conditions. Also included in the table are the means from the three control groups of participants who heard the same sentence sets without any intervening alternate rate set (the "no change" conditions).

The table shows that an intervening change in compression has little impact on the listeners' recall of these highly compressed sentences. The first analysis compared the recall scores for the three conditions (Numbers 1–3 in Table 5) in which there was an intervening set of sentences at a different compression rate, with a control condition in which there was no change in compression rate across the sentence sets (Number 4 in Table 5). The data were analyzed using a three-way ANOVA with Change Condition (uncompressed–same talker, 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 × Set Position interaction (both $Fs < 1.37$, $p > .25$). 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 $Fs < 1.48$, $p > .22$). 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 also presents the data for the first two sentences in the set immediately following the change in compression rates. The table shows that performance on the first two sentences immediately following a change to an uncompressed rate (Numbers 1 and 2) is lower than on the same two sentences in the condition with no change in compression rate (Number 4). There is a decrease of 6.3 and 8.8 percentage points for the same and different talker conditions, respectively. This suggests the possibility that a change in compression rate resulted in a more immediate decrement in performance right after the switch back to the original 38% compression rate that was compensated for by the end of the sentence set. To examine this possibility

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*

| Condition no. | Preceding context | Before change | After change: All 5 sentences | After change: First 2 sentences |
|---|---|---|---|---|
| | Change conditions | | | |
| 1 | Uncompressed same talker | 25.9 | 29.6 | 19.6 |
| 2 | Uncompressed different talker | 27.4 | 27.1 | 18.6 |
| 3 | 50% compressed same talker | 27.2 | 32.4 | 22.1 |
| | No change conditions | | | |
| 4 | 38% compressed same talker | 25.7 | 31.7 | 25.0 |
| 5 | No preceding context (no normalization) | | 20.4 | 12.9 |
| 6 | Complete context same talker (full normalization) | | 32.8 | 23.9 |

*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 uncompressed 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 parts of the speech signal equally, the temporal relationship between cues such as VOT or closure duration to vowel duration are also distorted. Therefore, extraction of phonetic information from compressed speech is severely jeopardized, and it is remarkable that listeners can recognize anything at all. Second, once phonetic representations are determined, whether correct or not, lexical access must then operate with an unusually high rate of phonetic–phonological input. After lexical access has occurred, and barring errors, syntactic integration and semantic interpretation have to cope with the unusually high rate of syntactic–semantic input. Therefore, difficulties are created for the entire speech processing system and adjustments may take place at any of these levels, resulting in an improvement in the recall score of compressed speech over time. In the following discussion, several of these possibilities are addressed.

The present research ruled out the possibility that the improvement in performance reflects the operation of certain task-specific strategies such as the ability to recall the sentences or write down the items on the response sheet. Participants presented with uncompressed training sentences on which they performed the transcription task had no better performance on subsequently presented compressed sentences than participants with no prior exposure to the compressed tokens (see also Mehler et al., 1993). Moreover, no effect of these training sentences was observed even when the sentences were partially masked by white noise, making them as difficult to understand as the compressed versions. Improvement occurred only when the training sentences were compressed.

We also judge it unlikely that the improvement reflects improved guessing of unperceived words based on the words that were accurately perceived. Participants initially recalled fewer than 30% of the words at the 38% compression rate, which does not provide much of a context in which guessing strategies can operate. In addition, the results of Experiment 1 indicate that the compression rate influences the gradient of adjustment. The higher the compression rate, the more exposure is required before performance reaches a plateau. Importantly, this difference in gradient is maintained even when the number of recalled words is globally equated between the two rates. This indicates that it is compression rate, and not the amount of recalled words, that determines the gradient of adjustment. This finding is consistent with those of Mehler et al. (1993), who reported that monolingual Spanish speakers can benefit from hearing compressed sentences in Catalan, a language they do not know (and for which they can hardly understand anything). In other words, comprehension of the message may not be necessary for adaptation to compressed speech to occur (for a similar finding, see Altmann & Young, 1993). Therefore, although the possibility of improvement at late levels of processing cannot be ruled out, it is unlikely that they are the only factors in the adjustment to highly compressed speech.

Another possibility is that the improvement reflects adjustments occurring at early levels of speech processing in which the acoustic signal is mapped onto underlying phonetic representations. As described previously, past studies have demonstrated that certain perceptual criteria at the acoustic–phonetic level can be modified by variations in speaking rate. Given that compressing speech dramatically alters the physical and perceived rate of the sentences, it is plausible that some portion of the adjustment to compressed speech was the result of low-level tuning of the perceptual system such as occurs during rate normalization. Some of the results from the current study do seem to reflect rate adjustments in phonetic perception. For example, there was a small and immediate drop in performance as a result of changing talker, compression rate, or both, which is reminiscent of similar effects in rate normalization studies.

However, the results of the current study reveal some important differences between compression adjustment and rate normalization. First, the adjustment to compressed speech occurs over a number of sentences whereas rate normalization operates over a fairly small, local context. Most studies on rate normalization randomize their test trials such that the rate of either the precursor phrase or the syllable itself varies from trial to trial. If the system did not respond within a single trial, there would be little difference in the phoneme boundaries across the entire experiment.[7] Second, although there was evidence for a drop due to a change in talker and compression rate, this drop was small and did not reset performance to baseline. In other words, unlike in rate normalization studies, compression adjustment shows considerable savings through changes in talker characteristics. Moreover, once stable performance has been reached for compressed speech, the level of performance is

[7] However, we cannot rule out the possibility that the nonlocal character of our effects are due to the extreme range of compression rates that we are using.

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 processses 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

Altmann, G. T. M., & Young, D. H. (1993). Factors affecting adaptation to time-compressed speech. *Proceedings of Eurospeech, 9, Berlin, 1*, 333–336.

Bladon, R. A. W., Henton, C. G., & Pickering, J. B. (1984). Towards an auditory theory of speaker normalization. *Language & Communication, 4*, 59–69.

Charpentier, F. J., & Stella, M. G. (1986). Diphone synthesis using an overlap-add technique for speech waveforms concatenation. *Proceedings of the IEEE International Conference ASSP, Tokyo* (pp. 2015–2018).

Darwin, C. J., McKeown, J. D., & Kirby, D. (1989). Perceptual compensation for transmission channel and speaker effects on vowel quality. *Speech Communication, 8*, 221–234.

Diehl, R. L., Souther, A. F., & Convis, C. L. (1980). Conditions on rate normalization in speech perception. *Perception & Psychophysics, 27*, 435–443.

Diehl, R. L., & Walsh, M. A. (1989). An auditory basis for the stimulus–length effect in the perception of stops and glides. *Journal of the Acoustical Society of America, 85*, 2154–2164.

Eimas, P. D., & Miller, J. L. (1980). Contextual effects in infant speech perception. *Science, 209*, 1140–1141.

Epstein, W., & Broota, K. D. (1986). Automatic and attentional components in perception of size-at-a-distance. *Perception & Psychophysics, 40*, 256–262.

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

Fowler, C. A. (1990). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America, 88*, 1236–1249.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1993). On the nature of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 152–163.

Goldstein, E. B. (1989). *Sensation and perception.* Belmont, CA: Wadsworth.

Green, K. P., & Miller, J. L. (1985). On the role of visual rate information in phonetic perception. *Perception & Psychophysics, 38*, 269–276.

Green, K. P., Stevens, E. B., & Kuhl, P. K. (1994). Talker continuity and the use of rate information during phonetic perception. *Perception & Psychophysics, 55*, 249–260.

Green, K. P., Tomiak, G. R., & Kuhl, P. K. (in press). The interaction of rate and talker normalization processes during phonetic perception. *Perception & Psychophysics.*

de Haan, H. J. (1982). The relationship of estimated comprehensibility to the rate of connected speech. *Perception & Psychophysics, 32*, 27–31.

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

Hood, D. C., & Finkelstein, M. A. (1986). Sensitivity to light. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and performance, Volume 1: Sensory processes and perception* (pp. 5-1 to 5-66). New York: Wiley.

Johnson, K. (1990). Contrast and normalization in vowel perception. *Journal of Phonetics, 18*, 229–254.

Jolicouer, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology, 42*, 461–478.

Kidd, G. R. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 736–748.

Ladefoged, P. (1967). *Three areas of experimental phonetics.* London: Oxford University Press.

Ladefoged, P. (1989). A note on "Information conveyed by vowels." *Journal of the Acoustical Society of America, 85*, 2223–2224.

Ladefoged, P., & Broadbent, D. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29*, 98–104.

Martin, L. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V.

(1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 676–685.

Mehler, J., Sebastian, N., Altmann, G., Dupoux, E., Christophe, A., & Pallier, C. (1993). Understanding compressed sentences: The role of rhythm and meaning. In P. Tallal, A. M. Galabunda, R. R. Llinas, & C. von Euler (Eds.), Temporal information processing and the nervous system. *The Proceedings of the New York Academy of Science, 682,* 272–282.

Miller, J. L., Aibel, I. L., & Green, K. P. (1984). On the nature of rate-dependent processing during phonetic perception. *Perception & Psychophysics, 35,* 5–15.

Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of /b/ and /w/. *Journal of the Acoustical Society of America, 73,* 2124–2133.

Miller, J. L., & Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition, 13,* 135–165.

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica, 43,* 106–115.

Miller, J. L., Green, K. P., & Schermer, T. M. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Perception & Psychophysics, 36,* 329–337.

Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica, 41,* 215–225.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25,* 457–465.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America, 85,* 2088–2113.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24,* 175–184.

Pisoni, D. B., Carrell, T. D., & Gans, S. J. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Perception & Psychophysics, 34,* 314–322.

Pisoni, D. B., & McClaskey, C. L. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception & Psychophysics, 34,* 323–330.

Pisoni, D. B., & Mullennix, J. W. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85,* 365–378.

Pisoni, D. B., & Mullennix, J. W. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47,* 379–390.

Port, R. F. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics, 7,* 45–56.

Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics, 32,* 141–152.

Remez, R. E., Rubin, P. E., Nygaard, L. C., & Howell, W. A. (1987). Perceptual normalization of vowels produced by sinusoidal voices. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 40–61.

Schmitt, J. G., Moore, S. W., & Lass, N. J. (1986). Variable speech control vs. Varispeech: Does newer mean better? *Journal of Auditory Research, 26,* 183–190.

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27,* 395–408.

Stevens, E. B., Kuhl, P. K., & Padden, D. M. (1988). Macaques show context effects in speech perception. *Journal of the Acoustical Society of America, 84*(Suppl. 1), S77.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1074–1095.

Voor, J. B., & Miller, J. M. (1965). The effect of practice on the comprehension of worded speech. *Speech Monographs, 32,* 452–455.

## 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 yesterday.
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.

Received August 16, 1993
Revision received March 15, 1996
Accepted March 15, 1996 ■