Internal Use only
非公開

TR-H-236 0014

Shrinkage in the Perceived Duration of Speech and Tone by Acoustic Replacement.

Minoru TSUZAKI, Hiroaki KATO
and Masako TANAKA

1998.1.19

ATR人間情報通信研究所
〒619-0288 京都府相楽郡精華町光台2-2 TEL: 0774-95-1011

ATR Human Information Processing Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan
Telephone: +81-774-95-1011
Fax : +81-774-95-1008

Shrinkage in the Perceived Duration of Speech and Tone by Acoustic Replacement*

Minoru Tsuzaki
Hiroaki Kato
Masako Tanaka

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

Phone: +81-774-95-1021

e-mail: tsuzaki@hip.atr.co.jp

Running Head: DURATIONAL SHRINKAGE BY ACOUSTIC REPLACEMENT

---

* This manuscript was submitted to "*Perception and Psychophysics*" on January 19, 1998.

Abstract

To investigate mechanisms for perceiving the duration of an auditory event, an effect of perceptual grouping upon perceived durations was studied psychophysically. In the first experiment, the perceived duration of a spoken word was measured under three conditions of acoustic continuity, i.e., (a) intact, (b) noise-replaced, and (c) gap-replaced, as a function of the duration of the target stimulus. Under the noise-replaced condition, a portion of the target stimulus was physically replaced with a noise burst. Under the gap-replaced condition, the replacement was made with a gap. The gap-replacement resulted in a prominent shrinkage of the perceived duration. In the case of noise-replacement, the amount of shrinkage was moderate but highly significant, although the word employed was perceived to be phonetically intact. Independent of this effect of replacement, the amount of shrinkage was affected by the physical duration of the target stimulus. The second experiment tested an effect of noise replacement on the perceived duration of a tone burst. In this case, the noise replacement also shrunk the perceived duration of the non-speech stimulus. This noise-induced shrinkage could be regarded as having occurred at a stage of auditory signal processing that functions for a wide range of acoustic stimuli. The phenomenon is discussed in relation to a revised model for perceived durations.

Shrinkage in the Perceived Duration of Speech and Tone by Acoustic Replacement

When we listen to various acoustic events, e.g., speech, music, or environmental sounds, it is critical that we perceive their temporal structure to describe their characteristics. Time intervals between successive sub-events relate the rates of activities involved, and the time interval between the starting point and the ending point of a certain activity will indicate the period of existence of the activity. If one aligns the time markers of multiple events along a single time axis, that person will get a sequence of randomly distributed markers coming from different sound sources and representing different natures, i.e., onsets or offsets. These markers should be perceptually grouped to properly represent the temporal structure of a certain coherent event. This paper deals with the issue of perceptual decomposition and its relation to the perceived duration of an auditory event.

The literature on time perception has repeatedly stated that it is important to distinguish two types of temporal structures, i.e., the filled interval and the empty interval (James, 1890; Goldfarb & Goldstone, 1963; Goldstone & Godfarb, 1963; Grondin, 1993). As a general principle, a filled interval tends to be perceived longer than an empty interval (Goldfarb & Goldstone, 1963; Goldstone & Godfarb, 1963; Craig, 1973), and several models have been proposed to explain this tendency by assuming a systematic shift of offset markers relative to onset markers in the auditory representation of events (Todd, 1996; Penner, 1975; Allan, 1979, for review). If we consider a response of the peripheral auditory system, it is predictable that an onset is processed with a faster time constant than an offset. Physiological micro-electrode recordings have shown a prominent onset responses in post stimulus time histograms of the cochlea nucleus (Pickles, 1982). In addition, the shape of the ear's temporal window, derived from the results of non-simultaneous masking experiments, has shown a temporal asymmetry (Moore, Glasberg, Plack, & Biswas, 1988).

However, it is not so easy to define points of onsets or offsets nor to predict whether intervals are filled or empty in realistic situations. For example, spoken words are perceived as coherent sounds, although we know that they include multiple segments of different characteristics, i.e., consonants and vowels. The phenomenon of phonemic restoration is a compelling evidence that we hear spoken words as coherent sounds. When a segment of a spoken word is replaced with a noise burst, what we hear is not a speech-noise-speech sequence, but an intact whole speech with a concurrent noise burst (Warren, 1970).

Listeners are unable to notice that the speech segment is physically missing and cannot distinguish between noise addition and noise replacement if the overall loudness is carefully equalized (Samuel, 1981a; 1981b). Even though the acoustical characteristic of each segment differs from the others, our perceptual system is able to group those segments into a single entity; it also succeeds in excluding a portion of a noise burst as well. One important aspect of this phenomenon, in view of processing the temporal structure, is that the acoustic boundary between the pre-noise portion and noise does not function as the ending point of the speech stimulus but as the starting point of the noise burst. In the same manner, the acoustic boundary between the noise and post-noise portion does not function as a new starting point of the speech stimulus but as the ending point of the noise burst. One should also note that a description of the physical characteristics is not sufficient for predicting whether an interval is filled or empty. It has been reported that the replacing sounds must possess a sufficient potential to mask the replaced sounds to induce perceptual continuity (Warren, 1984; for review). Therefore, we may conclude that sound with a masking potential will function as a filler. In other words, it will function identically to the original missed segment in the perceptual judgment of whether the sound is continuous or not. It is, however, another empirical question as to whether such a filler sound works equivalently in the perception of the temporal structure. In other words, defining the starting and ending points, and defining the filled and empty intervals, cannot be done in an a priori manner. This might be a result of the perceptual grouping; at the very least, it is necessary to consider the interaction between the perceptual grouping and the process defining the filled and empty intervals.

As Bregman (1990) argued, one might consider phonemic restoration a good demonstration of the principle of disjoint allocation. Following the principle of disjoint allocation, an edge cannot be perceptually allocated to more than a single object at once. Applying this principle to the case of phonemic restoration, the edges between some speech portions and a noise burst are allocated to the noise burst and not to the speech. Thus, the perceptual experience of hearing an intact spoken word is consistent with this principle. This intact experience, however, only applies when we pay attention to the degree of perceptual continuity, and may not apply when we focus on temporal structure, such as a perceived duration.

In a preliminary report by Tsuzaki et al. (1994), it was pointed out that noise replacement applied to a portion of a spoken word reduces the subjective duration of the speech sound compared to a physically intact speech. This demonstrated the possibility of a noise-induced shrinkage of the perceived duration. They also reported that the shrinkage could be observed for all variations of the standard duration, and that the amount of illusion was also influenced by the standard duration itself under both the noise-replaced condition and the intact condition. They explaind that the latter effect of the standard duration could be considered to reflect a schema-based process, that is, the stored knowledge about the word functions as a perceptual anchor and changes the perceived image slightly.

The fact that the noise-induced shrinkage was observed independent of the effect of the standard duration might suggest that the shrinkage itself is not governed by a speech-specific process but by a process common to other auditory events. To make this observation less compelling, the noise-induced shrinkage was not so prominent when the listeners made adjustments from shorter comparison stimuli in their results, and the range of the tested standard duration was not very wide. These limitations have prevented researchers from arriving at a general conclusion. Thus, in the first experiment of this paper, their experiment was replicated with a wider range for the standard duration.

EXPERIMENT 1

The main purpose of Experiment 1 was to replicate the experiment by Tsuzaki et al. (1994) with a wider range for the standard duration. The point of subjective equality (PSE) for the standard duration was measured by the method of adjustment. In the experiment of Tsuzaki et al. (1994), five standard duration classes were prepared by modifying the segmental duration of one vowel in a Japanese spoken word with an analysis-synthesis technique. The duration was shortened or lengthened by 25 ms at most from the original duration. In this experiment , we extended the lengthening direction up to 75 ms.

The second purpose of Experiment 1 was to compare the effect of noise replacement to the effect of gap replacement. It has been known that phonemic restoration occurs only if the replacing sound has sufficient power for masking the to-be-replaced sound. If the replacement is made with a gap, phonemic restoration does not occur. If the results under the noise-replaced condition do not differ from those in the gap-replaced condition, it could be concluded that the perceptual processing of duration is not substantially related to the perceptual reconstruction of auditory events.

Method

Design

A four-way factorial design was applied. The first factor was the existence of acoustic replacement in the standard stimulus. There were three levels in this factor; i.e., (1) intact [I], (2) noise-replaced [N], and (3) gap-replaced [G]. The second factor was the duration of the standard stimulus. The duration of the standard stimulus was modified (shortened or lengthened) to one of five levels; that is, -25, 0, 25, 50, or 75 ms, relative to the original naturally spoken sample. The third factor was the series of adjustments, i.e., ascending or descending. The fourth factor was the presentation order. In the first presentation order, the standard stimulus preceded the comparison stimulus. In the second presentation order, the comparison stimulus preceded the standard stimulus. All of these factors were treated as a within-subject comparison.

Stimulus

As in the experiment of Tsuzaki et al. (1994), synthesized and duration-modified tokens based on a Japanese word, "shi.na.bi.ru", were used. To make a well-balanced stimulus set in terms of naturalness by modifying the segmental duration of the stimulus, we selected the original material carefully. A Japanese 4-mora word, "shi.na.bi.ru", which means "to wither" in English, was selected as the original material because the duration of its third vowel, "i", was judged to be sufficiently "natural" in a perceptual experiment (Kato, Tsuzaki, & Sagisaka, 1992; Kato, Tsuzaki, & Sagisaka, in press); the original token was selected from the ATR speech database (Sagisaka, Takeda, Abe, Katagiri, Umeda, & Kuwabara., 1990). The duration of this third vowel was determined to be 90 ms according to the phonetic labeling attached to the speech database.

Using the Log Magnitude Approximation analysis-synthesis technique (Imai & Kitamura, 1978), the duration of the vowel "i" in the third mora was changed. The sampling frequency in this analysis-synthesis stage was 12 kHz and the synthesis parameters were updated after every 2.5-ms frame. Therefore, the minimum change in duration was 2.5 ms. The duration modification was achieved by deleting or inserting the parameters frame-by-frame. Figure 1 shows (a) sound spectrograms of the original material, (b) its synthesized version with no durational modification, (c) its synthesized version with 25-ms shortening, and (d) its synthesized version with 25-ms lengthening.

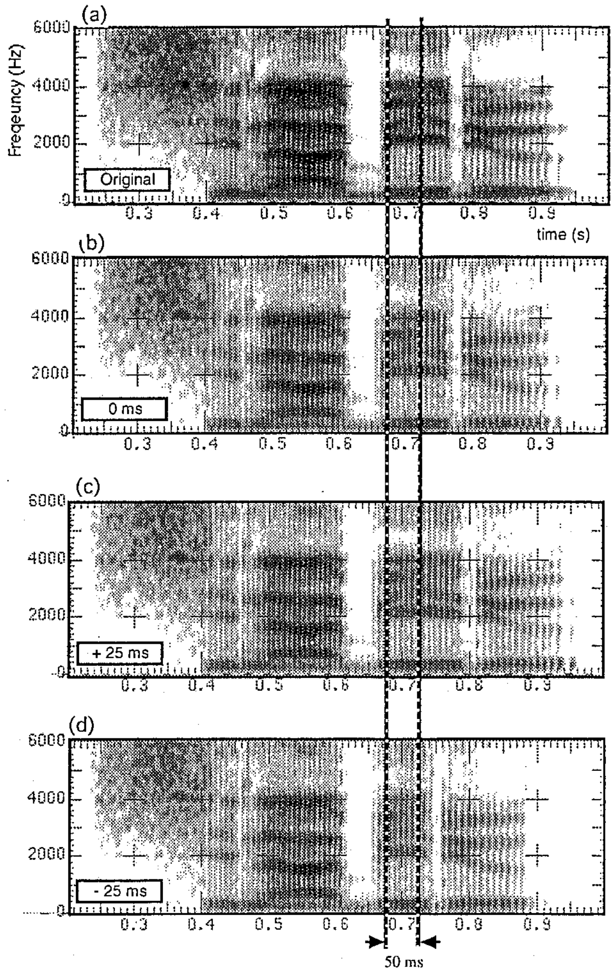

Figure 1. *Spectrograms of original and synthesized speech stimuli. (a) An original naturally spoken Japanese word, "shi.na.bi.ru"; (b) A resynthesized stimulus with the original temporal structure; (c) A resynthesized stimulus with 25-ms lengthening of vowel "i" in the third mora; (d) A resynthesized stimulus with 25-ms shortening of a vowel "i" in the third mora. The two vertical bars indicate the temporal location of the starting point and ending point of acoustic replacement.*

The amount of duration modification was extended up to 75 ms in the lengthening direction. In the shortening direction, it was maintained to -

25 ms because further shortening was likely to change the phonemic quality as well as the duration. Such a change in the phonemic quality had to be avoided because it might have resulted in accommodating a rather different task of comparing the durations of two different words. It is worth noting that the perceptual change would not occur in the third mora but in the fourth mora because the replacement perceptually and/or physically masked the transient between the third and fourth morae.

The replacing length was fixed at 50 ms. Under the noise-replaced condition, the initial 50-ms portion of the vowel /i/ was replaced with a burst of low-pass noise (fc = 5.5 kHz) whose power was 10 dB stronger than the power of the target portion. The duration, bandwidth and power of the noise burst were expected to be sufficient for inducing the phonemic restoration. (Warren, 1970; Bashford & Warren, 1987) Figure 2 shows the power spectrum of the to-be-replaced target portion of "i" and that of the replacing band noise. Preliminary listening to the noise-replaced stimuli assured us that phonemic restoration occurred. For the gap-interrupted condition, the same portion was replaced with a gap (silence). Although this gap-interrupted stimulus could still be recognized as "shi.na.bi.ru", the listeners reported that the interruption was quite unnatural and artificial.

Listeners

Ten listeners with normal hearing ability participated in the experiment. They were undergraduates of Doshisha University and were paid for their participation.

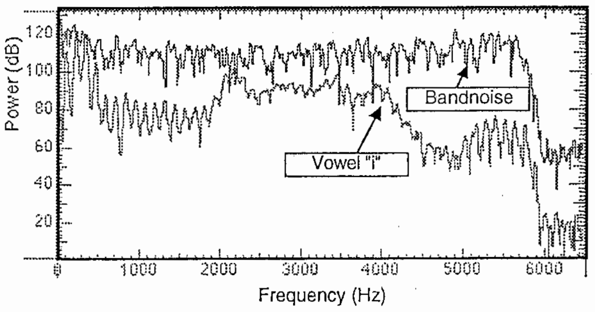

Figure 2. Spectral slices of a replacing noise and a to-be-replaced vowel "i" segment.

Procedure

The PSE was measured for each standard stimulus by the method of adjustment. Each listener listened to a successive presentation of the standard and comparison stimuli, or of the comparison and standard stimuli. The interstimulus interval was 250 ms. A personal computer (Macintosh IIci) controlled the sequence in the experiment. After each presentation of the stimuli, the listeners could adjust the duration of the comparison stimulus by pushing virtual buttons on the CRT display. There were six adjustment steps in each of the directions, i.e., shortening or lengthening. The minimum step was 2.5 ms and the maximum was 15 ms. When the listeners felt satisfied with the result of an adjustment, they were required to push another button, which was a signal to record a PSE value as well as to move on to the next condition. The listeners were required to adjust the duration of comparison stimulus so as to be perceived to be equal to that of the standard stimulus, which was intact, noise-replaced, or gap-replaced. To avoid confusion based on the reversal of the presentation order of the standard and the comparison stimuli, a visual display indicating the presentation order was presented throughout the session. Each listener was required to make four adjustments to each combination of the replacement, standard duration, series, and time order. The order of the conditions was counter-balanced. The number of total adjustments was 240 (= 3 × 5 × 2 × 2 × 4). The experiment was divided into four one-day sessions. At the beginning of each session, the listeners made a practice adjustment with each of three types of acoustically interrupted stimuli at the reference level (0 ms) of the standard duration. Each session took about two hours, including instructions and practice trials. The tests were done in a sound treated room at ATR Human Information Processing Research Laboratories.

The stimuli were upsampled to 44.1 kHz

and fed diotically by a DSP board (digidesign Sound Accelerator board) and headphones (STAX SR-Λ Pro driven by STAX SRM-1 MkII). The average sound pressure level for the presentation was 74 dB for the stimulus without noise (measured with A-weighting by a B&K 2231 Precision Sound Level Meter through a B&K 4134 condenser microphone mounted on a B&K 4153 artificial ear). The background noise level was 21 dB (measured at the location of the listener with a B&K 2231 Precision Sound Level Meter and a B&K 4155 condenser microphone with A-weighting).

Results

As an index for the illusory tendency, a PSE shift was calculated by subtracting the point of objective equality (POE) from the PSE. When this index is positive, it indicates the tendency of overestimation. When this index is negative, it indicates the tendency of underestimation. Figure 3 shows the average PSE shift pooled over the listeners, time order, and adjustment series for each of the acoustic replacement as a function of the standard duration. The shrinkage by noise replacement was observed again. The duration of the gap-replaced speech was far more underestimated than that of the noise-replaced speech. An ANOVA {Replacement × Standard duration × Time order × Adjustment series Listener} was performed with the Listener as a blocking factor. The main effects of the Replacement, Standard duration, and Adjustment series were significant [$F(2,18) = 140.31$, $p < 0.0001$; $F(4, 36) = 3.32$, $p < 0.03$; $F(1, 9) = 13.34$, $p < 0.006$, respectively]. There was a significant interaction between the Standard duration and the Replacement [$F(8, 72) = 5.05$, $p < 0.0001$]. As shown in Figure 3, the profile of the gap-replaced condition was quite different from the other two, which might have led to the interaction. To check this possibility, another ANOVA was performed on a subset of data excluding the gap-replaced condition. The main effects of the Replacement, Standard duration, and Adjustment series were once again significant [$F(1, 9) = 41.59$, $p < 0.0001$; $F(4, 36) = 7.36$, $p < 0.0002$; $F(1, 9) = 9.23$, $p < 0.02$, respectively]. The interaction between the Replacement and the Standard duration was not significant for this subset [$F(4, 36) = 1.74$, $p < 0.17$]. Therefore, the existence of noise replacement shrunk the perceived duration of the restored speech stimulus, and the amount of shrinkage was almost equivalent among the Standard durations. The main effect of the Replacement did not significantly interact with the Adjustment series [$F(2,18) = 0.397$, $p < .687$]. Although the results of Tsuzaki et al. (1994) failed to show a prominent shrinkage by noise replacement in their ascending series, the results of the current experiment show a significant shrinkage by noise replacement both in the ascending and descending series.

The results of Experiment 1 showed that the PSE shift could not be regarded as a monotonic function of the standard duration as it was in the experiment of Tsuzaki et al (1994). For example, under the intact condition, the degree of underestimation increases up to the 25-ms standard duration, and then the direction of

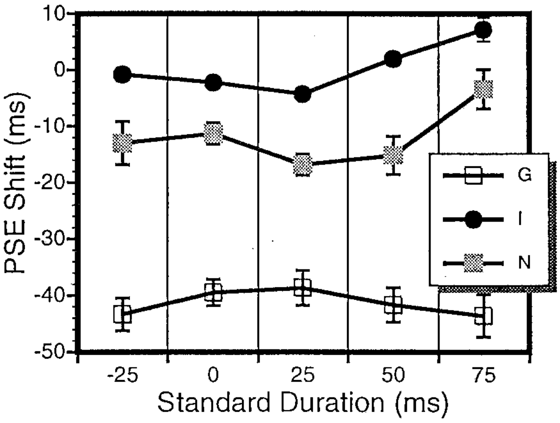

Figure 3. PSE shifts plotted as a function of the standard duration in Experiment 1. The standard duration is expressed in terms of relative value to the original duration. The parameter is the factor of noise replacement; (a) Intact [I], (b) Noise-replaced [N], and (c) Gap-replaced [G]. Each error bar indicates the standard error for each cell with an individual average score as a single sample point.

illusion switches to overestimation. In spite of this non-monotonic change against the standard duration, the relation between the intact condition and the noise-replaced condition remained fairly parallel.

The main effect of the time order was not significant [$F(1, 9) = 1.21$, $p < 0.31$]. However, it interacted with other factors [time order × standard duration, $F(4, 36) = 3.74$, $p < 0.02$; time order × replacement, $F(2, 18) = 3.74$, $p < 0.05$; time order × replacement × series, $F(2, 18) = 16.23$, $p < 0.0001$; time order × replacement × standard duration, $F(8, 72) = 6.18$, $p < 0.0001$]. These interactions reflect the fact that an obvious time order error was observed when the standard duration was 50- or 75-ms in the intact condition while the time order error was moderate in the remaining conditions in Figure 4.

Discussion

The results showed that acoustically replaced speech is perceived as being shorter than intact speech. Reliable shrinkage was observed under both adjustment series, i.e., ascending and descending. However, the amount of shrinkage involved was completely different between the case when replacement was made with a noise burst and the case when it was made with a gap. The results of the noise-replaced condition were closer to those of the intact condition than those of the gap-replaced condition. The perceptual restoration occurred under the noise condition while it did not under the gap condition. In this sense, imposing a noise burst benefited the restoration of the missed duration.

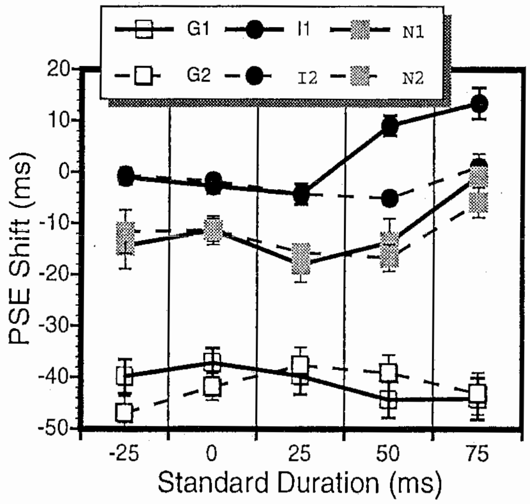

Figure 4. PSE shifts plotted as a function of the standard duration in Experiment 1 to display the time order error. The standard duration is expressed in terms of relative value to the original duration. The parameter is a combination of the factor of noise replacement and the time order of presentation. Each letter indicates a condition of the replacement; (a) Intact [I], (b) Noise-replaced [N], or (c) Gap-replaced [G]. Each number after the letter indicates the order of presentation; (a) Standard -> Comparison [1] or (b) Comparison -> Standard [2].

However, the results also suggested that this restoration of duration may not be complete because a reliable underestimation was observed, compared to the case when no such restoration was necessary, i.e., the case of the intact stimulus. Therefore, it is suggested that the perceived duration is influenced by a process of perceptual grouping, or the perceptual reconstruction of auditory events. Although the factor of time order interacted with the other factors, the shrinkage by replacement was clearly observed under both time order conditions. In the following discussion, we will focus on the two main factors, i.e., replacement and standard duration.

It is worth noting that the listeners were instructed to match the durations of the whole speech stimuli, and that the acoustic replacement, whether it was made with a noise burst or with a gap, was located in the middle of the speech in Figure 1. It is not plausible to assume that the existence of this replacement affects the detection of the beginning point and the ending point of the speech stimulus. Therefore, if a listener tries to be as accurate as possible, his/her optimal strategy would be to measure a temporal interval between these two time markers. In this case, no significant effect of the replacement would emerge.

As the results showed, this was not the case. The perceived duration was affected by the factor of replacement. This suggests that the property of filling is important in the perceptual estimation of a duration.

The main effect of the standard duration was also significant. Compared to the profile of a simple monotonic decreasing function obtained in the experiment of Tsuzaki et al. (1994), the PSE shift changed as a complicated function of the standard duration. The profile of a monotonically decreasing function can be explained by assuming a central tendency (Woodworth, 1938), a general tendency for a listener to use the middle point of the presented stimulus set as a perceptual anchor. Because a naturally spoken sample of a meaningful word was used as a seed of the standard stimulus both in the current experiment and in the experiment of Tsuzaki et al. (1994), one can expect a further effect of the perceptual anchoring based on a certain piece of stored knowledge for the word.

Effects of perceptual anchoring (assimilation) have also been reported for vowel quality (Kuhl, 1991). If we assume that a similar anchoring effect occurs for the durational perception, a stimulus longer than the natural (central) one will tend to be perceived as shorter, and a stimulus shorter than the natural one will tend to be perceived as longer. However, the results of the current experiment showed that the PSE shift turns in the positive direction when the standard duration becomes 50-ms longer than the original sample.

This non-monotonic relation is difficult to explain by assuming a single anchor. It can, however, be explained by assuming multiple anchors based on the constraint of the moraic structure in Japanese words. In Japanese, there is a phonemic contrast mainly based on the durations of vowels. For example, the word "o.jI.sa.n" which means "uncle", is different from the word "o.jI.I.sa.n" which means "grandfather". The former is classified as a four-mora word and is transcribed with four kana characters. The latter is classified as a five-mora word and is transcribed with five kana characters.

In the case of the current stimulus, although the original word was a four-mora word, "shi.na.bI.ru," another anchoring point of a five-mora word, "shi.na.bI.I.ru," can emerge by lengthening the vowel in the third mora. Below a certain limit of lengthening, a stimulus might tend to be heard as a slightly longer four-mora word. Beyond that, however, it might be heard as a slightly shorter five-mora word.

As long as the acoustic replacement is limited to that made with noise, the PSEs for the replaced standard stimuli could be adjusted to be shorter than the intact standard stimuli by a fairly constant amount at any standard duration. Importantly, a noise-replaced stimulus was perceived to be shorter than its intact counterpart, and even the latter itself tended to be overestimated, i.e., at 50- and 75-ms. This indicates that the observed effect of the noise replacement was not a mere enhancement of the illusory tendency originated by a schema-based factor, such as the stored knowledge for the word.

Under the gap-replaced condition, the profile of illusion as a function of the standard duration seemed different from the other two conditions. However, if one replots the points from the gap-interrupted condition by shifting their X-coordinates by 40 ms, one will get Figure 5. In Figure 5, the pattern of PSE shifts under the gap-replaced condition does not appears so different from that under the other two conditions. In addition, the amount of this parallel shift (40 ms) roughly corresponds to the overall amount of underestimation under the gap condition relative to the intact condition. This suggests a model postulating two processing stages. The first stage assembles several segments into a coherent speech, where the existence of a gap results in shrinkage of duration by a certain amount. The second stage utilizes speech specific knowledge and skews images slightly to fit into certain categories. If this model holds, the durational shrinkage observed with acoustically replaced stimuli will no longer require any speech specific knowledge and should be observable with non-speech stimuli. The purpose of the second

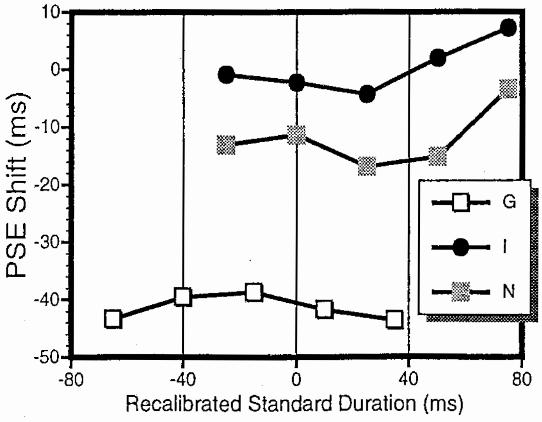

Figure 5. A plot applying a 40-ms parallel shift along the X-axis to points under the gap-replaced condition. The remaining data are the same as in Figure 3.

experiment was to test this prediction.

EXPERIMENT 2

In this experiment, a single burst of a sinusoid was used as the standard stimulus. To make a parallel comparison to the speech stimulus in Experiment 1, the duration of the tone burst was set to be equivalent to the segmental duration of the third mora of the speech stimulus in Experiment 1. An unlikely assumption is a constraint for the duration of the tone burst. In other words, no top-down information exists on how long a tone burst should last. Therefore, if a durational shrinkage by noise replacement were to be observed under this condition, it would be explained in terms of a certain bottom-up, non-speech-specific process.

Method

Design

A three-way factorial design was applied. The first factor was the existence of noise replacement in the standard stimulus, i.e., (1) intact [I], or (2) noise-replaced [N]. To reduce the size of the experiment, the gap-replaced condition was not tested. The second factor was the duration of the standard stimulus. There were three levels for this factor; i.e., 120, 150, and 180 ms. The third factor was the series of adjustments; i.e., ascending or descending. All of these factors were treated as a within-subject comparison.

Stimulus

The standard and comparison stimuli were sinusoidal tone bursts. The frequency was fixed to 240 Hz. This frequency roughly corresponded to that of the second harmonic of the to-be-replaced vowel "i" in Experiment 1, which had the maximum power (see Figure 2). The duration of the standard stimulus was either 120, 150, or 180 ms. The amplitude envelope had an initial 10-ms rising part and a final 10-ms falling part to suppress the perception of clicks. Under the noise-replaced condition, a 50-ms noise burst identical to what was used in Experiment 1 replaced a portion of the sinusoid from the 30 ms point to the 80 ms point of the tone burst. The presentation level of the tone burst was 69.5 dB SPL, and that of the noise burst was 85.5 dB SPL. The comparison stimuli were sinusoidal tone bursts with the same amplitude as the standard. They were never noise-replaced, and their duration was adjustable in 2.5-ms steps.

Procedure

As in Experiment 1, the PSEs of the six (2 X 3) standard stimuli were measured by the method of adjustment. The listeners were required to adjust the physical duration of the comparison stimulus until the subjective durations for both stimuli were perceived to be equivalent. Some of the listeners in Experiment 1 also reported that they were confused and distracted when the comparison stimulus preceded the standard stimulus. Therefore, the presentation order of the standard and the comparison stimuli was fixed in this experiment so that the standard stimulus preceded the comparison stimulus with an interstimulus interval of 705 ms. We assume that fixing the presentation order would not result in any significant difference for the purpose of the current experiment, although some interaction was observed between the time order and other factors in the results of Experiment 1. This is because time order errors were mainly observed under the intact condition in Experiment 1, and there were no substantial differences in the effects of the replacement.

Each listener was required to make eight repetitions of adjustments for each of the six standard stimuli. In the same session, they were also tested in a similar task of durational adjustment with different stimuli. Each listener was tested in two sessions on different days. One session lasted about an hour.

The same experimental equipment as in Experiment 1 was used to present the stimuli and to control the experimental sequences. The tests were done in a sound treated booth at ATR Human Information Processing Research Laboratories.

Listeners

Seventeen listeners with normal hearing ability participated in the experiment. They were undergraduates of Doshisha University and were paid for their participation.

Results

The average PSE shifts pooled over the listeners and the adjustment series were calculated as a function of the standard duration for the intact condition and the noise-replaced condition in Figure 6. The values for the noise-replaced stimuli were lower than those for the intact stimuli for all of the standard durations, and this indicated that the duration of the noise-replaced stimuli was perceived to be shorter than that of the intact counterpart. There was no clear effect of the standard duration on the PSE shift under the intact condition, while the PSE shift decreased monotonously as a function of the standard duration under the noise-replaced condition.

A four-way (Replacement × Standard duration × Adjustment series × Listeners) ANOVA was performed on the data with the Listeners as a blocking factor. Both main factors, i.e., the Replacement and the Standard duration, were significant [$F(1, 16) = 109.78, p<.0001$; $F(2, 32) = 24.82, p < .0001$, respectively]. They also significantly interacted with each other [$F(2,32) = 24.18, p < .0001$]. The main effect of the Adjustment series was also significant [$F(1,16) = 22.14, p < .0001$]. However, it did not interact with the Replacement [$F(1,16) = 0.2512, p < 0.6230$]. Therefore, this factor will be pooled in the following discussion.

Discussion

Shrinkage of the perceived duration by noise replacement was clearly observed with a single tone burst. Although there was a significant interaction between the Replacement and the Standard duration, the effect of noise replacement was significant even for the shortest standard duration, i.e., 120 ms, where the difference between the noise-replaced condition and the intact condition was minimum. A post hoc ANOVA for a subset of the 120-ms condition showed a significant effect of the factor of Replacement [$F(1, 16) = 28.4330, p <.0001$].

The fact that there was no systematic effect of the standard duration under the intact condition is consistent with the assumption that there is no *a priori* knowledge about how long a pure tone burst should sound. A *post hoc* test using a data subset under the intact condition showed that the effect of the standard duration was not statistically significant [$F(2, 32) = 0.5898, p < 0.5604$].

However, there might be another source of the top-down information, or contextual information, i.e., a central tendency driven by the stimulus set. Because a significant effect of the standard duration was observed under the noise-replaced condition [$F(2, 32) = 34.9152, p <.0001$], and because the PSE shift decreased monotonously as a function of the standard duration, it is probable to postulate a potential bias by the central tendency. Under the intact condition, the listeners would succeed in suppressing that bias because the stimulus was sufficiently simple and clear. Under the noise-replaced condition, on the other hand, the addition of noise reduced the reliability of the input signal and increased the contribution by the contextual information. From this, the effect of the standard duration emerged.

The purpose of using non-speech stimulus in this experiment was to exclude effects of the top-down information. Although it was not possible to exclude them completely, as shown in the results of the noise-replaced condition, the observed effect of the noise-replacement could not be fully accounted for solely by the top-down factor. If the noise replacement were merely to function as a factor facilitating the central tendency, the two lines in Figure 6 should have crossed each other near the middle point, and the overall shrinkage under the noise-replaced condition should not have happened.

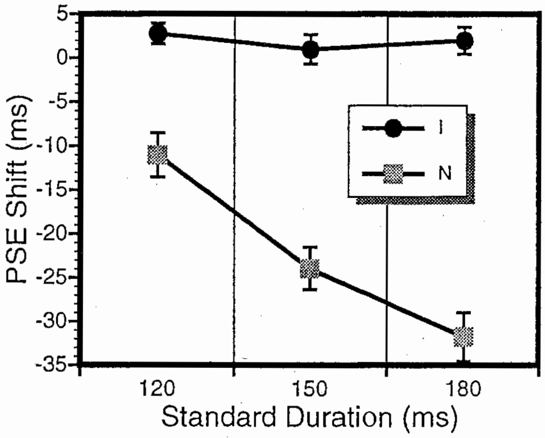

Figure 6. PSE shifts plotted as a function of the standard duration in Experiment 2. The parameter is the factor of noise replacement; (a) Intact [I] and (b) Noise-replaced [N].

General Discussion

Whether the target stimulus was a spoken word or a tone burst, a noise replacement performed on its portion resulted in a significant shrinkage of the perceived duration. At first glance, this appears to contradict the observation by Warren et al. (1994). They reported that the apparent duration of sound extends under the circumstances of auditory induction. However, the target of the durational estimation in the current experiments differed from that in the experiments of Warren et al. (1994). In the current experiments, the target interval contained a noise segment (an inducer), while the target segment was flanked by inducers in the experiments of Warren et al. (1994). Because of this difference in the stimulus configuration, it is not appropriate to directly compare the results of the two studies.

If one assumes that a mental calculation of the duration could be achieved merely by subtracting the time stamp of a physical starting point from that of a physical ending point, it would be difficult to explain the shrinkage by noise replacement. The observed results suggest that the content of the target interval affects the PSE shift. In literature on time perception, the difference between the <u>empty</u> interval and the <u>filled</u> interval has repeatedly been argued. For example, it has been stated that a time interval marked by two short tones does not perceptually match a physically equivalent interval marked by the starting and the ending of a tone (Goldfarb & Goldstone, 1963; Goldstone & Godfarb, 1963). Furthermore, a <u>divided</u> interval, a time interval marked by the first and the third short tones of a sequence of three short tones, is overestimated compared to its physical counterpart marked by just two tones (Jones & MacLean, 1966; Nakajima, 1979). Therefore, the durational shrinkage observed in the current experiments is not a unique demonstration that contradicts a simple model assuming a subtraction between the start and the end.

However, the current findings revealed several new aspects. First, it is worth noting that the direction of illusion in the current findings was opposite to that observed in the <u>divided</u> interval. One could regard the stimulus configuration in the current experiments as a type of analog of the <u>divided</u> interval because there is an extra sound, a noise burst, in the midst of the interval marked by the starting and the ending of an acoustic event, i.e., a spoken word in Experiment 1 or a pure tone in Experiment 2. The observed fact, however, indicates that the situation is completely different between the two cases. In the case of the <u>divided</u> interval, the original interval is an empty interval, and an extra sound, which functions as a divider, is likely to be grouped into a stream made by the original two sounds. In other words, the original empty interval is heard as a stream comprising two clicks, while the divided interval is heard as a stream comprising three clicks.

On the other hand, in the case of the interrupted duration, the original interval is a filled interval, and an extra sound is likely to be segregated from the original stream. While a listener can hear the extra noise aside from the target stream, the perceived temporal structure of the target stream may not significantly differ in terms of the number of sub-events, despite whether an acoustic replacement exists or not. Therefore, the distinction between the filled interval and the empty interval is important not only because their perceived durations differ but also because our perceptual system is likely to apply different groupings based on their perceptual qualities.

Second, the phenomenon of durational shrinkage by acoustic replacement requires additional mental devices for the computational model of duration perception. Models based on the neural counting process proposed by Creelman (1962) assume a pulse generator in the neural system, whose output is gated by an envelope for the input signal. The total number of pulses accumulated (counted) in this process is hypothesized to correspond to the perceived duration. Creelman assumed a probabilistic process, i.e., a Poisson process, for the pulse generation where the total number of counted pulses as well as its variance changed in proportion to the duration.

This assumption holds by postulating that the shape of the gating function is rectangular. In other words, the throughput of the pulses feeding into a counting process is assumed to be constant. It is difficult to predict the shrinkage simply by applying this model to the current experimental situation. Some modification is necessary. Divenyi and Danner (1977) proposed one of such modifications. They introduced a threshold device to trigger the pulse generator . They also empirically revealed the effects of the marker quality (intensity, frequency, or bandwidth) on the discrimination of time intervals and introduced the concept of latency variance in the triggering process. Their model was proposed to account for the discrimination of time intervals marked by two sounds, i.e., empty intervals, and they did not explicitly deal with filled intervals or intervened intervals. Therefore, we cannot apply their model either.

We propose additional devices in this counting process. One device estimates the strength of the perceptual evidence for the target stream, and a second device controls the opening degree of the gate as a function of the strength of the evidence. Therefore, the throughput of the pulses would decrease if the strength of the evidence were to decrease. It is not a very anomalous assumption that a decrease in the evidence would occur for a noise-replaced sound. This drop in the throughput of the pulses would result in a decrease in the accumulated pulses, which in turn would lead to a shrinkage of the perceived duration. It should be noted that the evidence estimator is not simply an ad hoc device constructed solely to account for the observed phenomenon of shrinkage. It would additionally work as an essential module for perceiving sounds in general. Otherwise, one would be unable to define even when the sound had a starting and ending. One may even argue that the mechanism for gate control by the evidence is not essential but rather arbitrary, and that it is far more parsimonious to assume the gate to be controlled just by the on-off of the event than to assume that the gate is controlled by the degree of estimated evidence. The former, however, implicitly requires another complicated mechanism to extract the on and/or off based on sensory inputs. Therefore, the assumption of the direct gate control by the estimated evidence is not disadvantaged in terms of parsimony.

It might also be argued that the satisfaction of the phonemic restoration, or the auditory induction, is not consistent with the concept of a drop in the amount of evidence. The background of this argument is the view hypothesizing that the perceptual restoration should not occur when there is any loss in the perceptual evidence. At first glance, the constraint of masking potential appears to hold this hypothesis. A sound is illusorily perceived to be continuous when there is no evidence for the disappearance of the sound. When the inducer has sufficient power, the existence of the target stimulus will make no effective difference in the auditory power spectra. Under this framework, there would be no drop in the amount of evidence.

However, we hypothesize that the level of evidence for perceptual continuity is slightly different from that for the gate control. It is important to note that the masking potential is discussed in the framework of the power spectrum model of audition, and the auditory power spectrum is not a unique source of information for human auditory processing. For example, even if an interrupting sound has sufficient power to mask an interrupted sound, as in the current experiments, fine structures in peripheral auditory representations are inevitably disrupted. As Patterson (1994a; 1994b) demonstrated, a fine structure conveyed by a temporal coding can be heard and is probably represented independently of information based on auditory power spectra. Although it might be a reasonable strategy for the auditory system not to take this disruption as evidence for the cessation of the target stream, it does not necessarily mean that there is no loss of evidence for the target stream.

The block diagram in Figure 7 summarizes the proposed <u>pulse gating</u> model of duration estimation. First, a peripheral image of an auditory mixture is constructed. Then, some indication of the arrival of a new event in the mixture image initiates building of a peripheral running image corresponding to the event, and tracking of the event starts. The amount of evidence is estimated based on the features in the event image until it becomes smaller than a certain threshold and the tracking process is terminated. This amount of evidence controls the aperture of the pulse gate, through which pulses are fed into the accumulator. When a noise portion arrives while tracking a speech or a tone burst, a drop in the evidence would happen because of the loss of some features, e.g., periodicity, and the aperture would become narrow. This would make the throughput of pulses lower and the amount of pulses accumulated would decrease. This, in turn, would lead to a shrinkage of duration. In the case of gap-replacement, the aperture would become much narrower, leading to further shrinkage.

One important restriction of this model is that it only works on durations of continuous sounds. A time interval marked by two sounds is beyond the scope of this model, while the original neural counter model by Creelman (1962) can be applied to an <u>empty</u> interval. We do not insist that the output from this model is a unique source of information in the time perception of auditory events. On the contrary, our experimental data show that the PSE shift is also affected by the standard duration. This suggests that there are other sources of information.

In an experiment like Experiment 1, such a source might be language specific knowledge. Findings by Warren and his colleagues (Warren & Sherman, 1974; Bashford & Warren, 1979) have also demonstrated that linguistic knowledge affects the perceived phoneme in phonemic restoration. In Experiment 2, it might be a context specific to the current stimulus set, i.e., the central tendency. Although several models have been proposed to describe how top-down or context-dependent knowledge affects perception, difficulty still remains in making a precise prediction about how these factors would work with the factor of noise replacement as well as how much they would affect the perceived duration. Therefore, if one could observe shrinkage by noise replacement only under a certain condition, the possibility would exist for the noise replacement to simply facilitate the contribution of stimulus-specific knowledge or a context-dependent bias.

The two current experiments revealed that the perceived duration shrunk at various standard durations by noise replacement applied either to a spoken word or to a tone burst. This demonstrates that neither factor, i.e., stimulus-specific knowledge nor stimulus context, is sufficient for explaining the observed shrinkage.

In the pulse gating model, noise-induced shrinkage is explained by a drop in the perceptual evidence induced by noise replacement. However, the possibility still remains that the observed phenomenon can be explained without assuming the pulse gating process. Noise replacement was applied from a point 30-ms after the start of the third mora (in Experiment 1), or from a point 30-ms after the start of the target tone itself (in Experiment 2). This temporal disparity of 30 ms is on the border of backward masking (Rasch, 1978; Zwislocki, 1978, for review). Therefore, we cannot completely exclude the possibility of the noise masking the starting point of the target segment. Although a systematic shrinking of the duration cannot be predicted by simply assuming the masking effect to result in obscuring the starting point, the starting point of the target could perceptually be fused with the starting point of the noise. This would place the time marker of the starting point slightly behind. As a result, the durational shrinkage could be predicted.

We, however, are suspicious of this hypothesis because durational shrinkage was also observed with gap replacement. No backward masking by a gap would be expected for the gap-

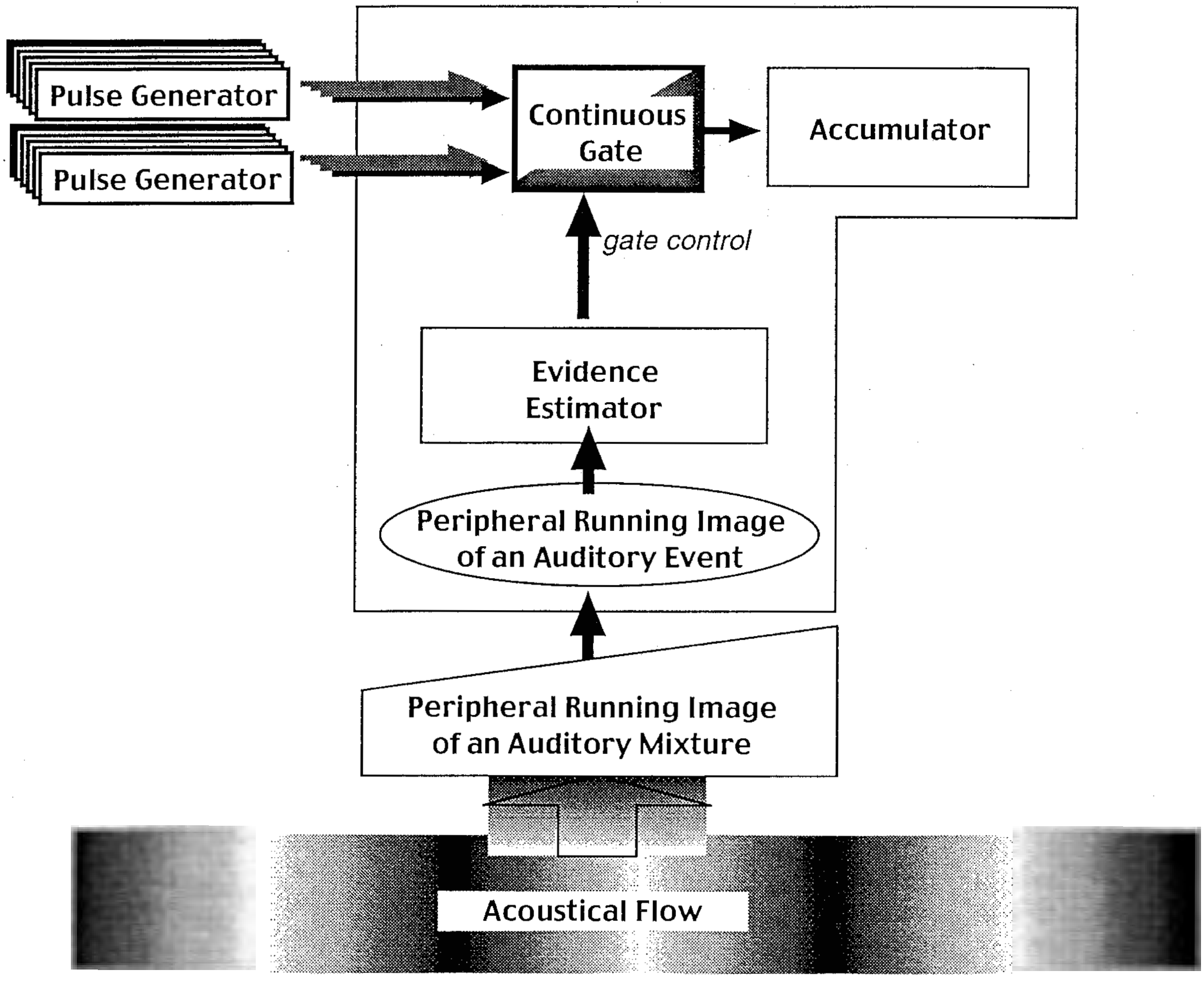

Figure 7. Block diagram of a "pulse gating model" for durational estimation.

replacement. The experimental results showed more prominent shrinkage in the case of gap replacement than in the case of noise replacement.

In our pulse gating model, the mechanism for evidence estimation has not been developed enough to give a quantitative prediction. However, recent progress in the field of computational auditory scene analysis might be able to give a good basis for building a detailed model for evidence estimation (Cooke, 1993; Brown & Cooke, 1994; Todd & Brown, 1996). We believe that the observations in the current experiments and further investigation on durational shrinkage by acoustic replacement will lead to the development of models for computational auditory scene analysis.

Summary

Noise replacement applied on a portion of an auditory event shrunk its perceptual duration. The shrinkage occurred when the target event was a spoken word or a tone burst. Gap replacement resulted in more prominent shrinkage. When the target stimulus was a spoken word, the PSE shift was also influenced by the physical duration. The effect of the standard duration was observed even with the intact stimulus, and it was independent of the factor of noise replacement. This could be explained by the stored knowledge for that word. The shrinkage by the acoustic replacement could be explained by the pulse gating model of durational perception. In this model, the perceived duration is assumed to corresponds to the number of accumulated pulses whose throughput varies depending on the strength of the perceptual evidence.

REFERENCES

Allan, L. G. (1979). The perception of time. Perception and Psychophysics, 26, 340-354.

Bashford, J. A., Jr, & Warren, R. M. (1979). Perceptual synthesis of deleted phonemes. In J. J. Wolf & D. H. Klatt (Eds.), Speech Communication Papers (pp. 423-426). New York: Acoustical Society of America.

Bashford, J. A., & Warren, R. M. (1987). Multiple phonemic restorations follow the rules for auditory induction. Perception and Psychophysics, 42, 114-121.

Bregman, A. S. (1990). Auditory Scene Analysis : The Perceptual Organization of Sound. Massachusetts: MIT Press.

Brown, G. J., & Cooke, M. P. (1994). Computational auditory scene analysis. Computer Speech and Language, 8, 297-336.

Cooke, M. P. (1993). Modeling auditory processing and organization. Cambridge: Cambridge University Press.

Craig, J. C. (1973). A constant error in the perception of brief temporal intervals. Perception and Psychophysics, 13, 99-104.

Creelman, C. D. (1962). Human discrimination of auditory duration. Journal of the Acoustical Society of America, 34, 582-93.

Divenyi, P. L., & Danner, W. F. (1977). Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. Perception & Psychophysics, 21, 125-142.

Goldfarb, J. L., & Goldstone, S. (1963). Time judgment: a comparison of filled and unfilled durations. Perceptual and Motor Skills, 16, 243-248.

Goldstone, S., & Godfarb, J. L. (1963). Judgment of filled and unfilled durations: Intersensory factors. Perceptual and Motor Skill, 17, 763-774.

Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. Perception and Psychophysics, 54, 383-394.

Imai, S., & Kitamura, T. (1978). [Speech analysis synthesis system using the log magnitude approximation filter.] Transaction of Institute of Electronics Communication Engineering Japan, [in Japanese], J61-A, 527-534.

James, W. (1890). The principles of psychology. New York: Holt.

Jones, A., & MacLean, M. (1966). Perceived duration as a function of auditory stimulus frequency. Journal of Experimental Psychology, 71, 358-364.

Kato, H., Tsuzaki, M., & Sagisaka, Y. (1992). Acceptability and discrimination threshold for distortion of segmental duration in Japanese words. In J. J. Ohara, T. M. Nearey, B. L. Derwing, M. M. Hodge, & G. E. Wiebe (Eds.), The 1992 International Conference on Spoken Language Processing: ICSLP 92, 1 (pp. 507-510). Banff, Canada: University of Alberta.

Kato, H., Tsuzaki, M., & Sagisaka, Y. Acceptability for temporal modification of single vowel segements in isolated words. Journal of the Acoustical Society of America.

Kuhl, P. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. Perception and Psychophysics, 50, 93-107.

Moore, B. C. J., Glasberg, B. R., Plack, C. J., & Biswas, A. K. (1988). The shape of the ear's temporal window. Journal of the Acoustical Society of America, 83, 1102-116.

Nakajima.Y. (1979). [A psychological investigation of divided time intervals shown by sound bursts]. (in Japanese with an English abstract) The Journal of the Acoustical Society of Japan (J), 35, 145-151.

Patterson, R. D. (1994a). The sound of a sinusoid: Spectral models. Journal of the Acoustical Society of America, 96, 1409-1418.

Patterson, R. D. (1994b). The sound of a sinusoid: Time-interval models. Journal of the Acoustical Society of America, 96, 1419-

1428.

Penner, M. J. (1975). The perception of offset: A problem of decision criteria. Perception and Psychophysics, 17, 587-590.

Pickles, J. O. (1982). An Introduction to the Physiology of Hearing. London: Academic Press.

Rasch, R. A. (1978). The perception of simultaneous notes such as in polyphonic music. Acustica, 40, 21-33.

Sagisaka, Y., Takeda, K., Abe, M., Katagiri, S., Umeda, T., & Kuwabara, H. (1990). A large-scale Japanese speech database. In International Conference on Spoken Language Processing: ICSLP 90, 2 (pp. 1089-1092). Kobe,Japan: The Acoustical Society of Japan.

Samuel, A. G. (1981a). Phonemic restoration: Insights from a new methodology. Journal of Experimental Psychology: General, 110, 474-494.

Samuel, A. G. (1981b). The role of bottom-up confirmation in the phonemic restoration illusion. Journal of Experimental Psychology: Human Perception and Performance, 7, 1124-1131.

Todd, N. (1996). Towards a theory of the principal monaural pathway: Pitch, time and auditory grouping. In W. Ainsworth & S. Greenberg (Eds.), Workshop on the Auditory Basis of Speech Perception, (pp. 216-221). Keele.

Todd, N. P. M., & Brown, G. J. (1996). Visualization of rhythm, time and metre. Artificial Intelligence Review, 10, 253-273.

Tsuzaki, M., Kato, H., & Tanaka, M. (1994). Apparent duration of speech segments interrupted by a noise burst. Journal of the Acoustical Society of Japan (E), 15, 205-206.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. Science, 167, 392-393.

Warren, R. M. (1982). Auditory Perception: A New Synthesis. New York: Pergamon.

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. Psychologic al Bulletin, 96, 371-383.

Warren, R. M., Bashford, J., J.A., Healy, E. W., & Brubaker, B. S. (1994). Auditory induction: Reciprocal changes in alternating sounds. Perception and Psychophysics, 55, 313-322.

Warren, R. M., & Sherman, G. L. (1974). Phonemic restorations based on subsequent context. Perception and Psychophysics, 16, 150-156.

Woodworth (1938). Experimental Psychology. New York: Holt.

Zwislocki, J. J. (1978). Masking: Experimental and theoretical aspects of simultaneous, forward, backward, and central masking. In E. C. Carterette & M. P. Friedman (Eds.), Hearing (pp. 281-336). New York: Academic Press.

Author Notes

The authors thank Ms. Noriko Yoshikawa for her assistance in running the experiments.

A part of this study was presented at the 129th Meeting of the Acoustical Society of America (Washington, DC, 1995).

Appendix A.

ANOVA (I,N,G- 50ms) Experiment 1 --- Speech Stimulus "SHI.NA.BI.RU"

Response: MatchedDur

Summary of Fit

RSquare	0.799197
RSquare Adj	0.732374
Root Mean Square Error	12.86198
Mean of Response	-17.6646
Observations (or Sum Wgts)	2400

Tests wrt Random Effects

Source	SS	MS Num	DF Num	F Ratio	Prob>F
Subject	26272.9	2919.21	9	0.2921	0.9691
Std	12902.8	3225.71	4	3.2282	0.0231
Subject*Std	35972.1	999.224	36	1.6378	0.0692
Intrpt	735739	367869	2	140.3060	<.0001
Subject*Intrpt	47194.3	2621.91	18	2.3954	0.0132
Std*Intrpt	20445.5	2555.68	8	5.0498	<.0001
Subject*Std*Intrpt	36438.6	506.091	72	2.0555	0.0057
UorD	78147.1	78147.1	1	13.3384	0.0053
Subject*UorD	52729.3	5858.81	9	6.8574	0.0001
Std*UorD	2232.41	558.103	4	1.8620	0.1384
Subject*Std*UorD	10790.4	299.733	36	1.5141	0.1411
Intrpt*UorD	499.562	249.781	2	0.3970	0.6781
Subject*Intrpt*UorD	11325.4	629.191	18	4.3303	0.0048
Std*Intrpt*UorD	3527.96	440.995	8	2.8462	0.0083
Subject*Std*Intrpt*UorD	11155.6	154.939	72	1.0847	0.3656
TimeOrder	2677.59	2677.59	1	1.2081	0.3002
Subject*TimeOrder	19947.5	2216.38	9	5.2603	0.0086
Std*TimeOrder	3537.22	884.306	4	3.7414	0.0120
Subject*Std*TimeOrder	8508.82	236.356	36	0.8528	0.6836
Intrpt*TimeOrder	2546.06	1273.03	2	3.7584	0.0433
Subject*Intrpt*TimeOrder	6096.85	338.714	18	1.5089	0.1688
Std*Intrpt*TimeOrder	11577.9	1447.23	8	6.1816	<.0001
Subject*Std*Intrpt*TimeOrder	16856.6	234.119	72	1.6390	0.0188
UorD*TimeOrder	219.01	219.01	1	0.8534	0.3797
Subject*UorD*TimeOrder	2309.58	256.62	9	1.4562	0.2552
Std*UorD*TimeOrder	563.724	140.931	4	0.7582	0.5593
Subject*Std*UorD*TimeOrder	6691.28	185.869	36	1.3012	0.1705
Intrpt*UorD*TimeOrder	4323.58	2161.79	2	16.2292	<.0001
Subject*Intrpt*UorD*TimeOrder	2397.67	133.204	18	0.9325	0.5438
Std*Intrpt*UorD*TimeOrder	1236.96	154.62	8	1.0824	0.3852
Subject*Std*Intrpt*UorD*TimeOrder	10284.8	142.844	72	0.8635	0.7857

ANOVA EXP1 - SPEECH STIMULUS
SUBSET "INTACT NOISE"
Response: MatchedDur
Summary of Fit

RSquare	0.670076
RSquare Adj	0.560376
Root Mean Square Error	12.24904
Mean of Response	-5.80625
Observations (or Sum Wgts)	1600

Tests wrt Random Effects

Source	SS	MS Num	DF Num	F Ratio	Prob>F
Subject	16711.3	1856.81	9	0.2312	0.9840
Std	27212.2	6803.04	4	7.3625	0.0002
Subject*Std	33264.2	924.007	36	1.5948	0.1052
Intrpt	60762.3	60762.3	1	41.5888	0.0001
Subject*Intrpt	13149.2	1461.03	9	1.4985	0.2127
Std*Intrpt	2960.13	740.033	4	1.7429	0.1620
Subject*Std*Intrpt	15285.8	424.606	36	2.3740	0.0238
UorD	53940.1	53940.1	1	9.2280	0.0141
Subject*UorD	52607.2	5845.24	9	9.7849	<.0001
Std*UorD	4473.49	1118.37	4	3.6262	0.0139
Subject*Std*UorD	11103.1	308.419	36	1.5285	0.1530
Intrpt*UorD	451.562	451.562	1	1.4072	0.2659
Subject*Intrpt*UorD	2888.05	320.894	9	5.1288	0.1055
Std*Intrpt*UorD	532.93	133.232	4	1.4128	0.2495
Subject*Std*Intrpt*UorD	3395.04	94.3066	36	0.8367	0.7022
TimeOrder	3164.06	3164.06	1	2.3271	0.1615
Subject*TimeOrder	12237.1	1359.68	9	2.2667	0.1163
Std*TimeOrder	9225.27	2306.32	4	9.3977	<.0001
Subject*Std*TimeOrder	8834.88	245.413	36	0.8053	0.7380
Intrpt*TimeOrder	1958.06	1958.06	1	4.0016	0.0765
Subject*Intrpt*TimeOrder	4403.89	489.321	9	2.9562	0.0388
Std*Intrpt*TimeOrder	1819.87	454.967	4	2.3064	0.0769
Subject*Std*Intrpt*TimeOrder	7101.38	197.261	36	1.7501	0.0488
UorD*TimeOrder	361	361	1	1.4393	0.2609
Subject*UorD*TimeOrder	2257.36	250.818	9	1.3309	0.3017
Std*UorD*TimeOrder	1160.6	290.15	4	1.3177	0.2820
Subject*Std*UorD*TimeOrder	7926.9	220.192	36	1.9536	0.0240
Intrpt*UorD*TimeOrder	1425.06	1425.06	1	17.5990	0.0023
Subject*Intrpt*UorD*TimeOrder	728.766	80.974	9	0.7184	0.6887
Std*Intrpt*UorD*TimeOrder	277.164	69.291	4	0.6148	0.6548
Subject*Std*Intrpt*UorD*TimeOrder	4057.68	112.713	36	0.7512	0.8570

Results of ANOVA for One Tone Burst (EXP.2)

Response: Illusion

Summary of Fit

RSquare	0.725106
RSquare Adj	0.633923
Root Mean Square Error	17.15276
Mean of Response	-10.1746
Observations (or Sum Wgts)	816

Tests wrt Random Effects

Source	SS	MS Num	DF Num	F Ratio	Prob>F
Name	20652.6	1290.79	16	0.1797	0.9994
Std	16351.5	8175.76	2	24.8162	<.0001
Name*Std	10542.5	329.453	32	0.7741	0.7749
Intrpt	119202	119202	1	109.7829	<.0001
Name*Intrpt	17372.8	1085.8	16	1.8294	0.0732
Std*Intrpt	13583	6791.52	2	24.1825	<.0001
Name*Std*Intrpt	8987.01	280.844	32	4.4643	<.0001
UorD	145400	145400	1	22.1362	0.0002
Name*UorD	105095	6568.44	16	12.6230	<.0001
Std*UorD	2265.64	1132.82	2	5.4552	0.0091
Name*Std*UorD	6645.04	207.657	32	3.3009	0.0006
Intrpt*UorD	94.3704	94.3704	1	0.2512	0.6230
Name*Intrpt*UorD	6009.67	375.604	16	5.9706	<.0001
Std*Intrpt*UorD	743.428	371.714	2	5.9088	0.0065
Name*Std*Intrpt*UorD	2013.08	62.9088	32	0.2138	1.0000

ANOVA for -30 ms condition

Response: Illusion

Summary of Fit

RSquare	0.208609
RSquare Adj	0.098878
Root Mean Square Error	21.76631
Mean of Response	-4.15441
Observations (or Sum Wgts)	272

Tests wrt Random Effects

Source	SS	MS Num	DF Num	F Ratio	Prob>F
Name	9416.45	588.528	16	1.2878	0.3095
Intrpt	12994.1	12994.1	1	28.4330	<.0001
Name*Intrpt	7312.13	457.008	16	0.9646	0.4964

ANOVA OTB (Intact condition) for Effect of Standard Duration

Response: Illusion

Summary of Fit

RSquare	0.063238
RSquare Adj	-0.06796
Root Mean Square Error	24.47944
Mean of Response	1.911765
Observations (or Sum Wgts)	408

Tests wrt Random Effects

Source	SS	MS Num	DF Num	F Ratio	Prob>F
Name	8311.95	519.497	16	2.8120	0.0062
Std	217.923	108.961	2	0.5898	0.5604
Name*Std	5911.76	184.743	32	0.3083	0.9999