Perception & Psychophysics
1992, 51 (5), 504-507

Notes and Comment

Time-shrinking: A discontinuity in the perception of auditory temporal patterns

YOSHITAKA NAKAJIMA
Kyushu Institute of Design, Fukuoka, Japan

GERT TEN HOOPEN and GASTON HILKHUYSEN
Leiden University, Leiden, The Netherlands

and

TAKAYUKI SASAKI
Miyagi Gakuin Women's College, Sendai, Japan

Recent research at our laboratories in the field of human auditory time perception revealed that the duration of short empty time intervals (< ~200 msec) is considerably underestimated if they are immediately preceded by shorter time intervals. Within a certain range, the amount of subjective time shrinking is a monotonous function of the preceding time interval: the shorter it is, the more it shrinks its successor. In the present study, the preceding interval was kept constant at 50 msec, and the following interval, for which the duration had to be judged, varied from 40 to 280 msec. The results showed that at up to 100 msec, the perceived duration increased to a much lesser extent than did the objective duration. Beyond 120 msec, the perceived duration quickly increased and reached a veridical value at 160 msec. Such a sudden change of perceived duration in a temporal pattern in which the objective duration varies gradually indicates a typical example of categorical perception. We suggest that such a categorization of the time dimension might be a clue for processes of speech and music perception.

Investigation of the temporal context in which short auditory time intervals are perceived is important to clarify the psychological and physiological basis of the perception of speech, music, and other environmental sounds (e.g., Bregman, 1990; Deutsch, 1986; Handel, 1989). In a series of studies on the duration perception of single empty time intervals embedded in temporal patterns, we recently found a new illusion of auditory time perception (Nakajima, ten Hoopen, & van der Wilk, 1991). For the present report, it suffices to summarize the basic finding.

We gratefully acknowledge the Canon Foundation in Europe for providing Visiting Research Fellowships to Y.N. and G.t.H. in 1991. We would like to thank Rob Hartsuiker, Astrid Ihle, Juliette van der Schrier, and Claudia van de Wal for their services as subjects. We would also like to thank Valter Ciocca, Rob Hartsuiker, Kengo Ohgushi, Ryunen Teranishi, Joos Vos, and René van der Wilk for their comments on previous drafts. Correspondence should be addressed either to Yoshitaka Nakajima, Kyushu Institute of Design, Department of Acoustic Design, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815, Japan, or to Gert ten Hoopen, FSW/RUL, Experimental Psychology, P.O. Box 9555, 2300 RB Leiden, The Netherlands.

When listeners adjusted the duration of a comparison empty time interval (C) to that of a standard empty time interval (S) that had a duration of 120 msec, they made almost veridical matches.[1] However, if S was preceded by a shorter empty time interval (P), as shown in Figure 1A, the listeners considerably underestimated the duration of S. When the duration of P was longer than that of S, the perceived duration of the latter was almost the same as its objective duration. Table 1 displays a typical example.

We call "time-shrinking" the phenomenon in which the perceived duration of a time interval S is shortened by the presence of an immediately preceding shorter interval (P). Table 1 shows that the shorter P was, the more S shrunk within a certain range.[2]

A convergent operation to investigate the shrinking mechanism is to keep P constant and to vary S instead. A clue to what happens in that case stems from one of our experiments (ten Hoopen, Vis, Hilkhuysen, & Nakajima, 1989) in which the positions of S and C in the adjustment task were exchanged (see Figure 1B). The argument was that if P shrinks the following interval, it should not matter whether the latter interval functions as standard or comparison, and the prediction was that C should be relatively underestimated in some conditions—that is, C should be adjusted to a longer objective duration than that of S. This test of the time-shrinking illusion was successful.

A very peculiar aspect of the perceived duration of the adjustable interval C was reported by the subjects after this experiment. If, for example, P was 45 msec, S was 120 msec, and the initial C value was 40 msec, the listeners of course increased C in successive approaches to make it equal to S. (The adjustments were accomplished by pressing a mouse button: The longer the mouse button was

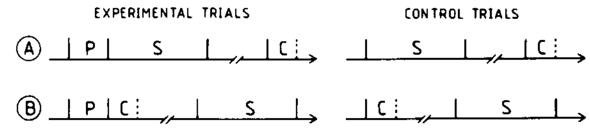

Figure 1. (A) Auditory stimulus patterns in which a comparison empty time interval (*C*) had to be adjusted to a standard empty time interval (*S*). In the experimental condition, *S* was immediately preceded by another time interval (*P*). In the control condition, *S* was presented without *P*. (The patterns in Figure 1A were used in previous studies, as well as in the present one.) (B) Auditory stimulus patterns in which a comparison empty time interval (*C*) had to be adjusted to a standard empty time interval (*S*). In the experimental condition, *C* was preceded by another empty time interval (*P*). In the control condition, *C* was presented without *P*. (The patterns in Figure 1B were only used in previous studies.)

Table 1
A Typical Example of the Time-Shrinking Illusion: Mean Points of Subjective Equality (PSE_{exper}) of a Standard Time Interval (S) of 120 msec as a Function of the Duration of the Preceding Time Interval (P)*

	Duration of Preceding Time Interval						
	45	70	95	120	145	170	195
PSE_{exper}	66	76	91	105	122	121	120
PSE_{contr}	113	113	113	113	113	113	113
$PSE_{exper} - PSE_{contr}$	−47	−37	−22	−8	9	8	7

Note—All values in milliseconds. The mean PSE for the control condition (PSE_{contr}) in which S was presented alone was 113 msec. *From Nakajima et al.'s (1991) Experiment 2.

pressed, the more the variable sound marker of C was moved.) During their successive approximations to match S, the subjects reported that they perceived an abrupt change of C's subjective duration, although there was not a large amount of mouse-button pressing.

To examine these introspective reports, we devised stimulus patterns in which P was fixed at 50 msec and the following interval S increased from 50 to 160 msec and decreased back to 50 msec in steps of 10 msec. The patterns were presented in a quick succession (1.4 sec between patterns), as shown in Figure 2. This series evoked very similar phenomenological reports from all listeners (30) to whom it was presented. They reported that Patterns 1 to 5 (±1), and Patterns 20 (±1) to 24 (Figure 2) yielded very similar rhythmic impressions even though the ratio of the objective intervals gradually changed from 1:1 to about 1:2, or vice versa. Furthermore, all listeners reported a sudden change in the subjective ratio from Patterns 5-7 and from Patterns 19-21.

This informal listening test suggested an interesting characteristic of the time-shrinking phenomenon. The preceding interval P kept reducing the perceived duration of S as the objective duration of the latter was increased from 50 to ~100 msec. It was only when its duration became longer than ~100 msec that S was perceived as having a clearly longer duration than P.

We decided to submit these informal observations to a formal experiment. Eight subjects with normal hearing (the 4 male authors—2 Japanese and 2 Dutch—and 4 Dutch students—1 male and 3 female) matched the duration of the adjustable comparison interval C to that of S. Within a trial, S could have one of the several durations that are specified in Table 2, and it could be presented either with a preceding P (experimental condition) or without P (control condition). The interval C was presented 5 sec after the onset of S, and the initial value of C was either sufficiently short (ascending situation) or long (descending situation) compared with S. A Commodore Amiga 500 computer was used to generate the stimuli, to control the experiment, and to collect the data.[3]

The duration of C could be changed by means of mouse-button pressing in "Shorten" and "Lengthen" panes. By clicking in a "Presentation" pane, one could listen to the result of the most recent adjustment 2 sec later. The subjects could listen to each stimulus pattern as many times as they wished. When they were satisfied with the adjustment, a "Finish" pane could be clicked at, thereby recording the final duration of C, that is, the point of subjective equality (PSE) of S, into the computer's memory. Each subject did 96 trials [24 (12 control + 12 experimental) × 2 (ascending/descending) × 2 (replications)], which were presented in randomized order. A training session of 30 trials preceded the experiment proper. Figure 3 displays the median PSEs in the control condition (circles); they were veridical, as expected.

More interesting was the trajectory that the median experimental PSEs (squares) followed as a function of the objective duration of S. When the point of objective equality (POE) increased from 40 msec on, the median PSE of S increased to a much smaller extent. Only after the POE of S exceeded 100 msec did the PSE move quickly toward the control value, which was reached at about POE = 160 msec, and it became even slightly greater than the control value after the POE exceeded 160 msec. Inspection of the 32 experimental PSEs at POE = 120 msec revealed a bimodal distribution: About half of the PSEs approached the veridical value, but the other half still centered around a shrunken value (see Table 3). Since it seemed risky to assume normally distributed PSEs, we did not subject the data to a parametric statistical test but instead performed Wilcoxon matched-pairs signed-ranks tests, one at each value of S (see Figure 3).

The results of the experiment correspond to the listeners' claim that Patterns 1 to 5 (±1) and 20 (±1) to 24

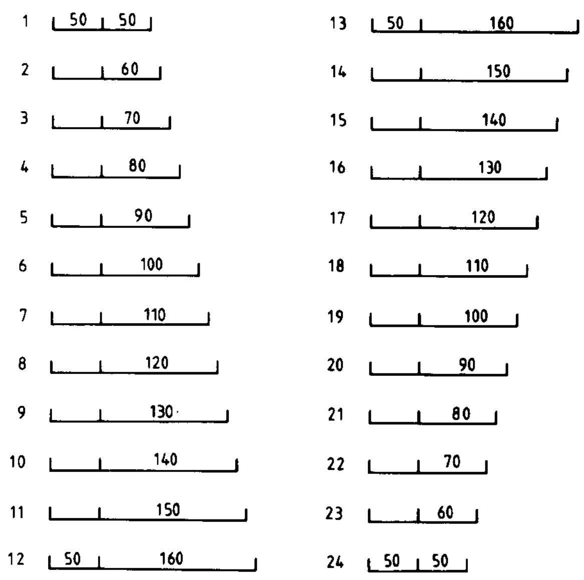

Figure 2. Diagram of the sequence of auditory temporal patterns used in the informal listening test. The first empty time interval remained 50 msec, whereas the second one increased in steps of 10 msec from 50 to 160 msec and then decreased again to its starting value. Up to pattern 5 (±1) and down from pattern 20 (±1), a 1:1 percept was reported despite changing time ratios. (Notice that by visual inspection of the patterns, no such categorical boundary can be detected—the perceived spatial ratio changes gradually.)

Table 2
Durations of the Standard Empty Time Intervals (S) Without or With Preceding Time Interval (P) as Used in the Method of Adjustment

Control Condition S	Experimental Condition P/S
40	50/40
50	50/50
60	50/60
70	50/70
80	50/80
100	50/100
120	50/120
140	50/140
160	50/160
200	50/200
240	50/240
280	50/280

Note—All values in milliseconds.

in the informal listening test were perceived as having a 1:1 ratio. That is, time interval pairs of 50/40 (in milliseconds) through 50/~100 all lump into the same perceptual category, and the perception changes very slowly in this region. From 50/160 on, however, the perception is quite different and changes step by step. (The statistical test showed even higher PSEs than in the control condition at 50/240 and 50/280, which we had not expected from our previous observations.) Between 50/100 and 50/160, both of the two perceptual modes mentioned above seem to appear, as is evidenced by the clear bimodal distribution of PSEs at 50/120 (Table 3).

The phenomenon of categorical perception has been well documented in speech perception, especially for consonants (e.g., Repp, 1984). Our results suggest that duration values can also be clues in differentiating the meanings of spoken words or sentences. If our perceptual system has a mechanism to categorize durations or duration ratios, it can be efficiently used in our linguistic system. A very clear example is the Japanese language, in which speech is controlled by an isochronic temporal framework called "mora timing" (e.g., Port, Dalby, & O'Dell, 1987). It is important when speaking or listening to Japanese to determine whether or not the neighboring durations are practically equal. Our present experimental situation may be an experimental model to investigate this kind of linguistic behavior. In the domain of music, the existence of categorical rhythm production was reported in 1946 by Fraisse; recently, evidence has been presented (e.g., Clarke, 1987; Sloboda, 1985) that categorization exists for rhythm perception as well. We suggest that categorization of time, for which we found evidence in a primitive way, may be one of the hidden mechanisms on which the perception of speech and music is based.[4] Now we are trying to clarify the mechanism of this illusion. We invented several hypotheses to explain the illusion, and one is still alive: the assimilation hypothesis. The shrinking of S happens when P is less than S, but, nevertheless, P and S are perceived as equal or almost equal to each other. Thus, they are more similar to each other than would be expected from their physical durations. We propose that this is a typical example of perceptual assimilation: When several similar objects are situated close to each other, they can be perceived as even more similar. On the other hand, when two objects are considerably different from each other, the assimilation does not take place, but contrast between them appears instead. This indeed happened in our experiment when the difference between P and S was big enough. It can be seen from Figure 3 that S was significantly overestimated at 240 and 280 msec. Assimilation and contrast usually appear in the same kind of context just by changing the difference among several objects. Our experimental data give a typical example on this point when the temporal pattern changes.

A familiar visual example of assimilation is the old brickstone wall. The colors of the brickstones that form the wall appear similar to each other and are perceived differently from how they appear in isolation.

Although assimilation in visual spatial perception seems generally bilateral, the present type of assimilation we found for time perception seems unilateral. The shrinking of the second duration, S, is not paralleled by a sub-

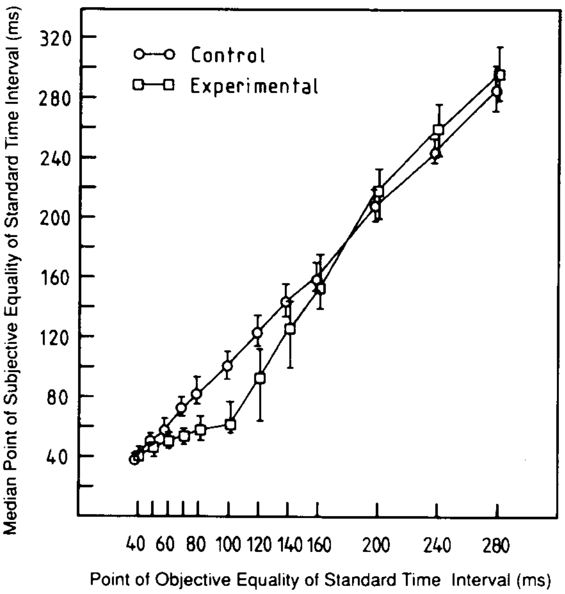

Figure 3. The median point of subjective equality (PSE) as a function of the point of objective equality (POE) for the control conditions (standard empty time interval S presented alone [circles]) and for the experimental conditions (standard S preceded by an interval P of 50 msec [squares]). The vertical bars represent the first and third quartiles (Q1 and Q3). Wilcoxon matched-pairs signed-ranks tests (two-tailed) showed the following results: The experimental PSEs were significantly smaller than the control PSEs at the 1% level when POE = 50--140 msec. Significantly greater PSEs appeared when POE = 40, 240, and 280 msec at the 5%, 1%, and 5% level, respectively.

Table 3
Stem-and-Leaf Display of the 32 PSEs (in msec) in the 50/120 msec Condition

5	0, 1, 3, 5
6	0, 1, 2, 3, 4, 6
7	0, 2, 3, 7
8	0
9	2, 3, 4, 8
10	1, 4, 8
11	0, 3, 3, 5
12	0, 3, 7, 9
13	1, 8

jective increase of the first duration, *P*, as we demonstrated in Nakajima et al. (1991). Unilaterality is very common when the temporal dimension is involved. To mention only one example, auditory forward and backward masking are far from symmetrical. Usually, the influence of preceding events on succeeding events is bigger than the other way around, which was also the case in our illusion.

The assimilation hypothesis, however, does not fully describe the results. There is at least one important thing that needs to be explained in the future: When *P* is physically longer than *S*, the perception of *S* is not affected by *P* as clearly as when *P* is shorter than *S*. Nevertheless, our present data are a clear example of categorical perception of the temporal dimension. In future research, it would be desirable to show the nature of this kind of categorical perception by using different experimental methods such as paired comparison or discrimination.

REFERENCES

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Clarke, E. F. (1987). Categorical rhythm perception: An ecological perspective. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 19-33). Stockholm: Royal Swedish Academy of Music, No. 55.

Deutsch, D. (1986). Auditory pattern recognition. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 32/1-49). New York: Wiley.

Fraisse, P. (1946). Contribution à l'étude du rythme en tant que forme temporelle. *Journal de Psychologie*, **39**, 283-304.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events.* Cambridge, MA: MIT Press.

Nakajima, Y., ten Hoopen, G., & van der Wilk, R. G. H. (1991). A new illusion of time perception. *Music Perception*, **8**, 431-448.

Port, R. F., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America*, **81**, 1574-1585.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 10, pp. 243-335). New York: Academic Press.

Sloboda, J. A. (1985). *The musical mind: The cognitive psychology of music.* Oxford: Oxford University Press.

ten Hoopen, G., Vis, G., Hilkhuysen, G., & Nakajima, Y. (1989). A new illusion in time perception II. *Proceedings of the First International Conference on Music Perception and Cognition* (pp. 237-240). Kyoto, Japan.

NOTES

1. An empty time interval was an interval bounded by two sound markers of 10 msec each and measured from onset to onset. The sound markers consisted of complex tones with frequency components of 1, 3, and 5 kHz and started and stopped at zero-crossing points.

2. The equation through the PSEs of *S* for *P* values of 45, 70, 95, and 120 msec was PSE = 41 + 0.53 · *P* msec ($r^2 > .99$).

3. The sound output of the Amiga 500 was fed into an amplifier (JVC, AX211) and presented monaurally through headphones (JVC, H-610). Listening level was calibrated to about 95 dBA (with a continuous tone of the same waveform and amplitude) by a Brüel & Kjaer precision sound level meter (type 2203) connected to an artificial ear (type 4152) mounted with a condensor microphone (type 4144). The timing of the stimulus patterns was meticulously calibrated by a digital storage oscilloscope (Gould Advance), as well as by sound-analysis software (Aegis AudioMaster and Sunrize Studiomagic).

4. We could not find any reference to the phenomenon we found, either in journals or in such encompassing reviews as the ones by P. Fraisse (*Annual Review of Psychology*, 1984, Vol. 35, pp. 1-36) and by M. R. Jones & M. Boltz (*Psychological Review*, 1989, Vol. 96, pp. 459-491). Also, we could not find any reference to the present phenomenon in the books edited by J. Gibbon and L. G. Allan (*Timing and time perception, Annals of the New York Academy of Sciences*, Vol. 423, 1984) and by J. A. Michon & J. L. Jackson (*Time, mind, and behavior*, Springer-Verlag, Berlin, 1985).

(Manuscript received November 4, 1991;
accepted for publication December 9, 1991.)