Time-shrinking and categorical temporal ratio perception: Evidence for a 1:1 temporal category

Hoopen, Gert ten
Department of Psychology, Leiden University

Sasaki, Takayuki
Psychoacoustic Laboratory, Miyagi Gakuin Women's University

Nakajima, Yoshitaka
Department of Acoustic Design, Kyushu University

Remijn, Gerard Bastiaan
Department of Acoustic Design, Kyushu University

他

出版情報：Music Perception. 24 (1), pp.1-22, 2006-09-01. University of California Press Journals
バージョン：

Time-Shrinking and Categorical Temporal Ratio Perception: Evidence for a 1:1 Temporal Category

Gert ten Hoopen
Department of Psychology, Leiden University

Takayuki Sasaki
Psychoacoustic Laboratory, Miyagi Gakuin Women's University

Yoshitaka Nakajima
Department of Acoustic Design, Kyushu University

Ger Remijn
Department of Acoustic Design, Kyushu University

Bob Massier
Department of Psychology, Leiden University

Koenraad S. Rhebergen
Department of Psychology, Leiden University

Willem Holleman
Department of Psychology, Leiden University

In a previous study, we presented psychophysical evidence that time-shrinking (TS), an illusion of time perception that empty durations preceded by shorter ones can be conspicuously underestimated, gives rise to categorical perception on the temporal dimension (Sasaki, Nakajima, & ten Hoopen, 1998). In the present study, we first survey studies of categorical rhythm perception and then describe four experiments that provide further evidence that TS causes categorical perception on the temporal dimension. In the first experiment, participants judged the similarity between pairs of /t1/t2/ patterns (slashes denote short sound markers delimiting the empty time intervals t1 and t2). A cluster analysis and a scaling analysis showed that patterns liable to TS piled up in a 1:1 category. The second and third experiments are improved replications in which the sum of t1 and t2 in the /t1/t2/ patterns is kept constant at 320 ms. The results showed that the 12 patterns /115/205/, /120/200/, . . . , /165/155/, /170/150/ formed a 1:1 category. The fourth experiment utilizes a cross-modality matching procedure to establish the subjective temporal ratio of the /t1/t2/ patterns and a 1:1 category was established containing the 11 patterns /120/200/, /125/195/, . . . , /165/155/, /170/150/. On basis of these converging results we estimate a domain of perceived 1:1 ratios as a function of total pattern duration (t1 + t2) between 160 and 480 ms. We discuss the implications of this study for rhythm perception and production.

Received November 14, 2002, accepted January 9, 2006

Key words: time-shrinking illusion, categorical rhythm perception, domain of perceived 1:1 ratios, cross-modal matching, multidimensional scaling

Categorical perception, the condensation and accumulation of similarities and differences between objects and events, has been amply investigated in speech perception (e.g., Repp and Liberman, 1987; Rosen and Howell, 1987). In music perception, studies on categorical perception have been mainly done in relation to the perception of musical pitch (e.g., Burns, 1999). The first scholar who pointed out that categorical perception might also play a role in time and rhythm perception was Fraisse (1978, 1982). From several temporal production and reproduction studies Fraisse found that durations in a temporal pattern were often assimilated or contrasted. Durations that do not differ too much can be assimilated: "If two durations belong to the same category, there is a tendency to equalize these durations. We prefer to say that there is assimilation since the equalization is not absolute" (Fraisse, 1982, p. 167). On the other hand, when durations in a temporal pattern clearly differ, they can be contrasted, that is, their difference is perceptually boosted.

We give a few examples from an experiment reported by Fraisse (1978, pp. 241–242). He presented temporal patterns comprising three consecutive durations (/t1/t2/t3/) which had to be reproduced. The pattern

/300/450/450/ ms for instance was reproduced as /270/560/530/ ms, thus the model ratio of 1.5 between t1 and t2 was boosted to 2.07. The pattern /360/450/450/ was reproduced as /390/450/460/ ms, thus the ratio between t1 and t2 of 1.25 was reduced to 1.15. Such contrast (or distinction) and assimilation effects were also established between t2 and t3. To quote Fraisse (1978): "There is a real dynamic organization of patterns and the relative duration of one of the intervals is modified by the duration of the others" (p. 240), and "These constant errors conform to the laws of assimilation and distinction" (p. 241). Fraisse's interpretation that temporal patterns are simplified for reasons of perceptual economy with tendencies toward 1:1 and 2:1, and 1:2, is a precursor of studies on categorical rhythm perception (Clarke, 1987, 2000; Desain & Honing, 2003; Papadelis & Papanikolaou, 2004; Parncutt, 1994; Schulze, 1989; Windsor, 1993), and rhythm categorization (Large, 2000).

Clarke (1987) presented musical participants short musical sequences of two measures. The first measure provided the context which was either a triple meter or a duple meter. The second measure contained three notes. The durations of the first two notes were varied between a 2:1 (640: 320 ms) ratio and a 1:1 (480: 480 ms) ratio in nine equal steps. The 18 sequences, nine in the triple context and nine in the duple context, had to be identified as 2:1 or 1:1 types, and the average identification curve showed a relatively steep slope between the fifth sequence (560 ms : 400 ms) and the seventh sequence (520 ms : 440 ms). The participants were also confronted with all pairs of sequences of which pair members were two steps apart, and were required to judge whether the two sequences were the same or different. The peak of the discrimination function coincided nicely with the steepest slope of the identification function, and Clarke concluded that this was strong evidence for a categorical distinction. Clarke also separated the data with respect to metrical context, and it appeared that there were two distinct category boundaries, one for the triple, and another for the duple meter.

Schulze (1989) replicated Clarke's seminal experiment. One of Schulze's criticisms was that Clarke's participants had only two response categories (2:1 and 1:1) at their disposal in the identification task, which prevents a conclusion about the origin of the boundary: sensitivity or response criterion. His second criticism was that Clarke did not vary the tempo of the rhythmic sequences, implying that the participant could in principle have discriminated the sequences by comparing two corresponding note durations in the two sequences. Hence, Schulze chose for graded responses in his experiment, and varied the tempo. Like Clarke, he varied the meter (triple and duple), but there were no contexts.

Schulze had two participants who were musically very well trained. There were four experimental conditions, indicated by the arrows numbered 1 through 4 in Figure 1. In each of these conditions, there were eight graded temporal patterns of three interval durations,

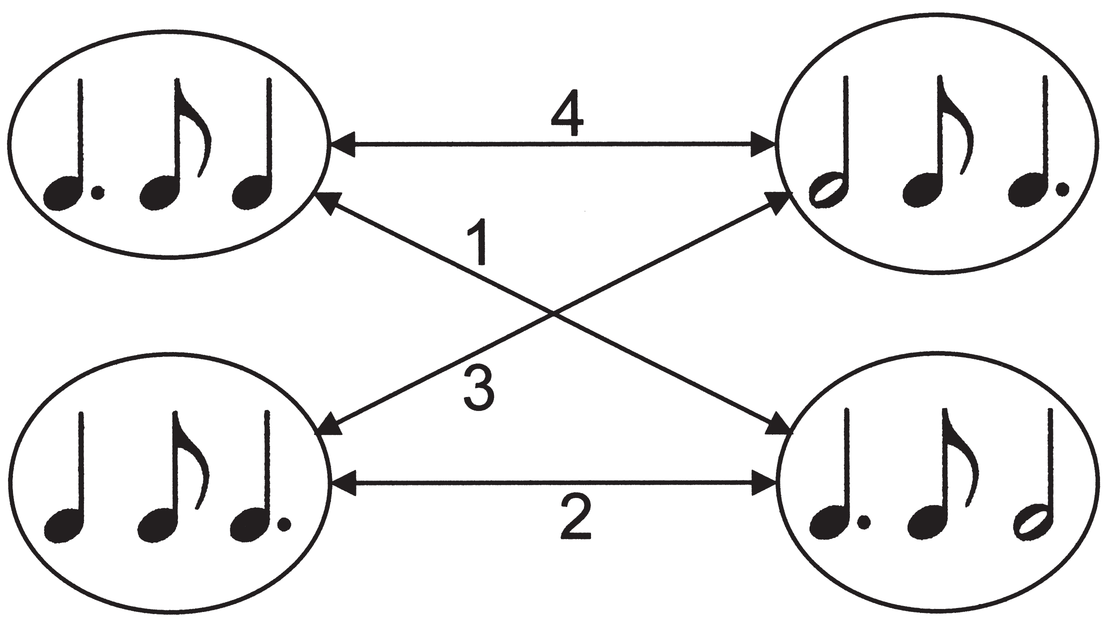

FIG. 1. End point rhythms used by Schulze (1989). See text for further explanation.

where the 2nd and 7th pattern represented the duple and triple rhythms drawn in Figure 1, and the 1st and 8th patterns were even more extreme. For clarity, we give one example when the total duration was 1200 ms for condition 1: /420/140/640/, /450/150/600/ (ideal duple), /480/160/560/, /510/170/520/, /540/180/480/, /570/190/440/, /600/200/400/ (ideal triple), and /630/210/360/. The total duration could also be 1080 ms and 1320 ms, in which the patterns were scaled proportionally. In each of the four conditions, the participants were extensively familiarized with the 24 rhythms (8 patterns $\times$ 3 tempi). In the experimental sessions, they had to identify the 24 randomly presented rhythms, thus the task was one-to-one mapping of responses to temporal patterns. Schulze did not require his participants to perform a discrimination task, but calculated an index of discriminability from the stimulus-response confusion matrices. The results did only partly support categorical rhythm perception. Only in two of the four conditions, Schulze found nonmonotonic discrimination functions.

Large (2000) examined the effect of metrical context on rhythm categorization using a modified method of limits. Two musicians listened to two kinds of rhythmic sequences. Ascending sequences consisted of rhythm cycles changing from a 1:1 ratio to a 2:1 ratio gradually, while descending sequences contained rhythm cycles changing from 2:1 to 1:1. The participants had to judge whether the rhythm was duple or not for the ascending sequences and whether the rhythm was triple or not for the descending sequences. The results showed that metrical context influenced the categorization of rhythm patterns. The interesting result of Large's study was that he found three categories instead of two, a duple, a triple, and a third category of 'neither duple nor triple.'

Recently, Desain and Honing (2003) presented an extensive study on categorical rhythm perception. Because they applied mathematical concepts not often used in studies of music perception, we shall describe their study in a more detail. In a thorough theoretical introduction the authors explained a couple of concepts in order to enable a clear definition of rhythmic categories. The concept of "performance space" stands for the space of all possible performances of N time intervals, that is, each point in this N-dimensional space represents a different temporal pattern. In the case of patterns with three time intervals (/t1/t2/t3/), used by the authors as stimulus material, one obviously has a 3-dimensional performance space, and because Desain and Honing chose a total duration of the three intervals of 1 second, performance space is a slice which can be graphically drawn as a "ternary plot." This ternary plot is an equilateral triangle, the three sides each representing the total duration of 1 s, and the values of the three time intervals (t1, t2, and t3) are each represented on a different side. By connecting these three t-values in parallel to the sides, each temporal pattern gets a unique location in the ternary plot (see Figure 2).

Desain and Honing used 66 temporal patterns that systematically sampled the performance space. To avoid a metrical subdivision by the sampling itself, they chose 1/19 s as the temporal grid unit (a prime value), and to stay in a realistic domain of musical durations, the shortest inter onset interval (IOI) of the temporal patterns was 3/19 s (158 ms). Consequently, the IOI varied in 10 steps of 52.6 ms between 158 ms and 684 ms, and the three most extreme patterns in the performance space were /158/158/684/, /158/684/158/, and /684/158/158/ ms (see Figure 2).

In their experiment 1, 29 musically well trained Dutch and Japanese participants were confronted with an identification task, in which they had to notate a musical score for each of the 66 stimuli, yielding the so-called score space. Desain and Honing defined categorization as mapping the performance space into a score space. The musical notations of the 29 participants were converted to integer ratios (e.g., 1:1:2, or 4:3:1), and the distribution of the 29 responses was calculated for each of the 66 patterns. From these distributions, a categorization or so-called time-clumping map was derived, depicting the maximal response proportions. Note that the performance space is discontinuous: the space contains the 66 objective temporal patterns that were presented. The score space, visualized as a time-clumping map, was made continuous by interpolating between the obtained maximum responses at each of the 66 patterns. Though this mapping procedure (from performance to score space) appears quite complicated, the result is very clear: the score space, also cast in a ternary plot, looks like a topographical map representing areas of different soil, in which each "area of soil" stands for a different category of subjective temporal integer ratios.

The time-clumping map shows several interesting aspects. Firstly, the category areas differ in size. The relatively simple ratios 1-1-1, 1-2-1, 2-1-1, and 1-1-2 consume more space in the ternary plot than complex ratios such as 2-3-1. Secondly, the areas are distributed unevenly around the mechanical rendition (the physical temporal pattern in performance space), mostly in the direction of a longer third interval. The time-clumping map also shows the amount of freedom that musicians have in applying expressive timing: the amount is constrained by the sizes of the category areas. Going beyond the boundaries will result in a different rhythm

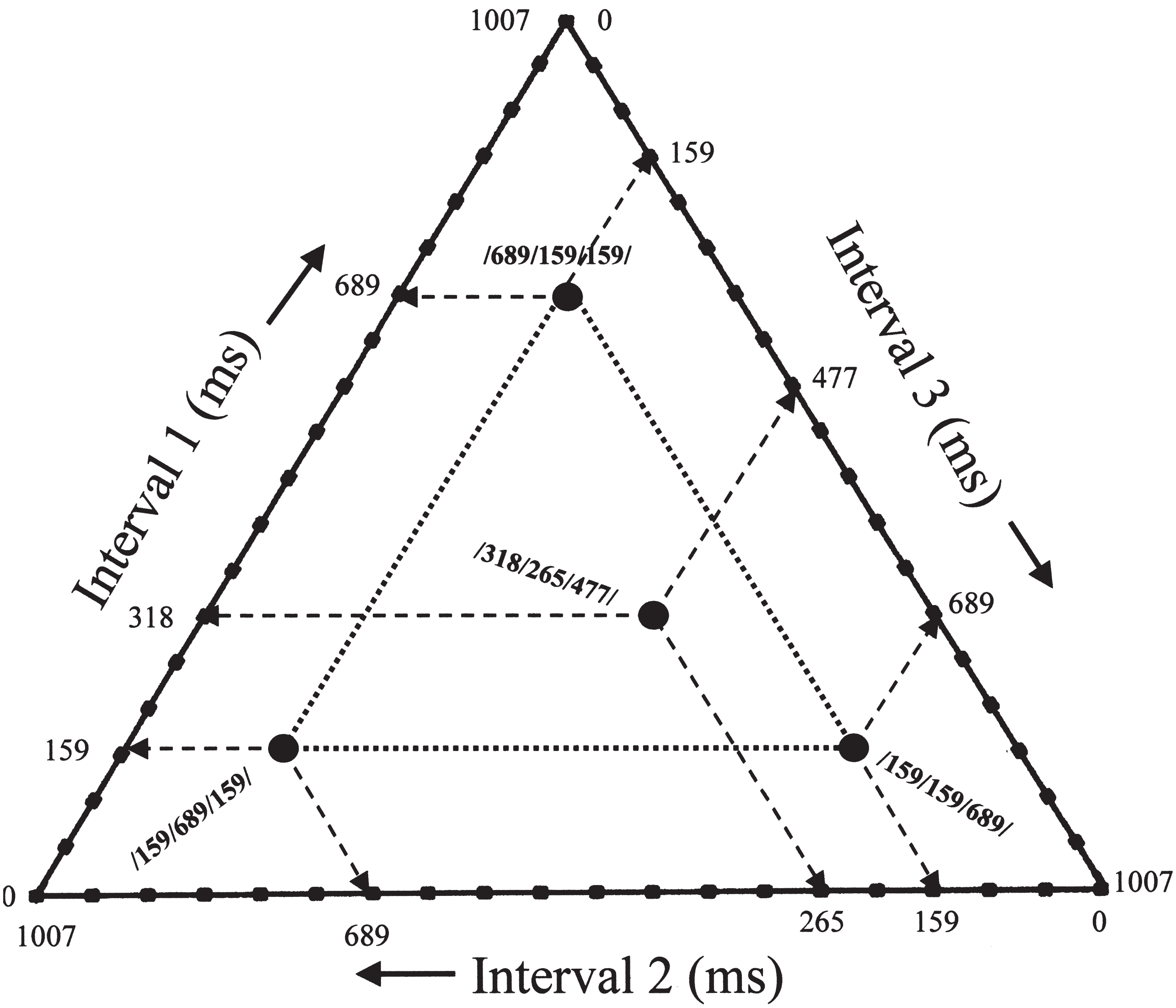

Fig. redu. 9%

FIG. 2. Ternary plot representation of the "performance space" by Desain and Honing (2003). Of the 66 patterns applied, the three most extreme ones, and one example in between are depicted.

Like Clarke (1987) did, Desain and Honing investigated whether varying the metrical context affected the layout of the score space. There were three conditions: no meter, duple meter, and triple meter. The time-clumping map analyses showed that the "duple meter" map and the "no meter" map did not differ very much, and this was explained by the fact that the listener often imposes a duple meter when lacking context. There was a considerable difference between the "duple" and "triple" maps. Whereas there were clear 1-1-2 and 1-2-1 category areas in the duple map, they did not show up at all in the triple map.

Another recent study of categorical perception in the rhythm/meter domain was reported by Papadelis and Papanikolaou (2004). The authors discussed Clarke's (1987, 2000) distinction between categorical information (best-fitting mental rhythmic schemata) and non-categorical information (slight but noticeable temporal deviations perceived as musical expression) but aimed one step further. The authors also sought after the internal structure of rhythmic categories to obtain not only a perceptual mapping of rhythms between, but a mapping within categories as well. To attain this goal, they adopted the goodness rating task (a technique used in phoneme perception studies).

In their experiment, the authors presented stimulus patterns consisting of three cycles of five empty durations delimited by five short percussive sounds: two equal longer durations (called "a") followed by three equal shorter durations (called "b"). In diagram: / a / a /b/b/b/ a / a /b/b/b/ a / a /b/b/b/. Thus, there were different rhythmic strata because the listener could extract different periodic levels. The three cycles were presented at three tempi, of which we discuss the medium tempo (112 MM) in which the duration of interval "a" was 535 ms. The b:a temporal ratio was varied in small steps at the level of the just noticeable difference for duration such that there were 16 patterns, their b:a ratio varying from 0.26 to 0.62, and the prototypical integer ratios of 1:3 (0.33) and 1:2 (0.50) rhythms were included.

In addition to performing the classic tasks of identification and discrimination, the participants had to rate the category goodness of each temporal pattern on a six-point scale. Figure 3 shows the identification functions (solid lines) and the goodness ratings (dotted lines) for the 3-beat rhythm and the 7-beat rhythm. Two important facts emerge. Firstly, the goodness rating curves overlap less than the identification curves. Secondly, the goodness rating curves give a more detailed picture of the structure within the categories, portraying which patterns are the most typical ones for the categories and how the other patterns are graded as regards their typicality. The authors posit the existence of a "perceptual magnet effect" within rhythmic categories in the same vein as has been proposed for categorical perception of speech sounds (Kuhl, 1991; Iverson and Kuhl, 1995).

In the studies discussed above, the gradual transition between patterns was realized by varying the duration of time intervals. Windsor (1993) investigated whether categorical distinction could be found between duple

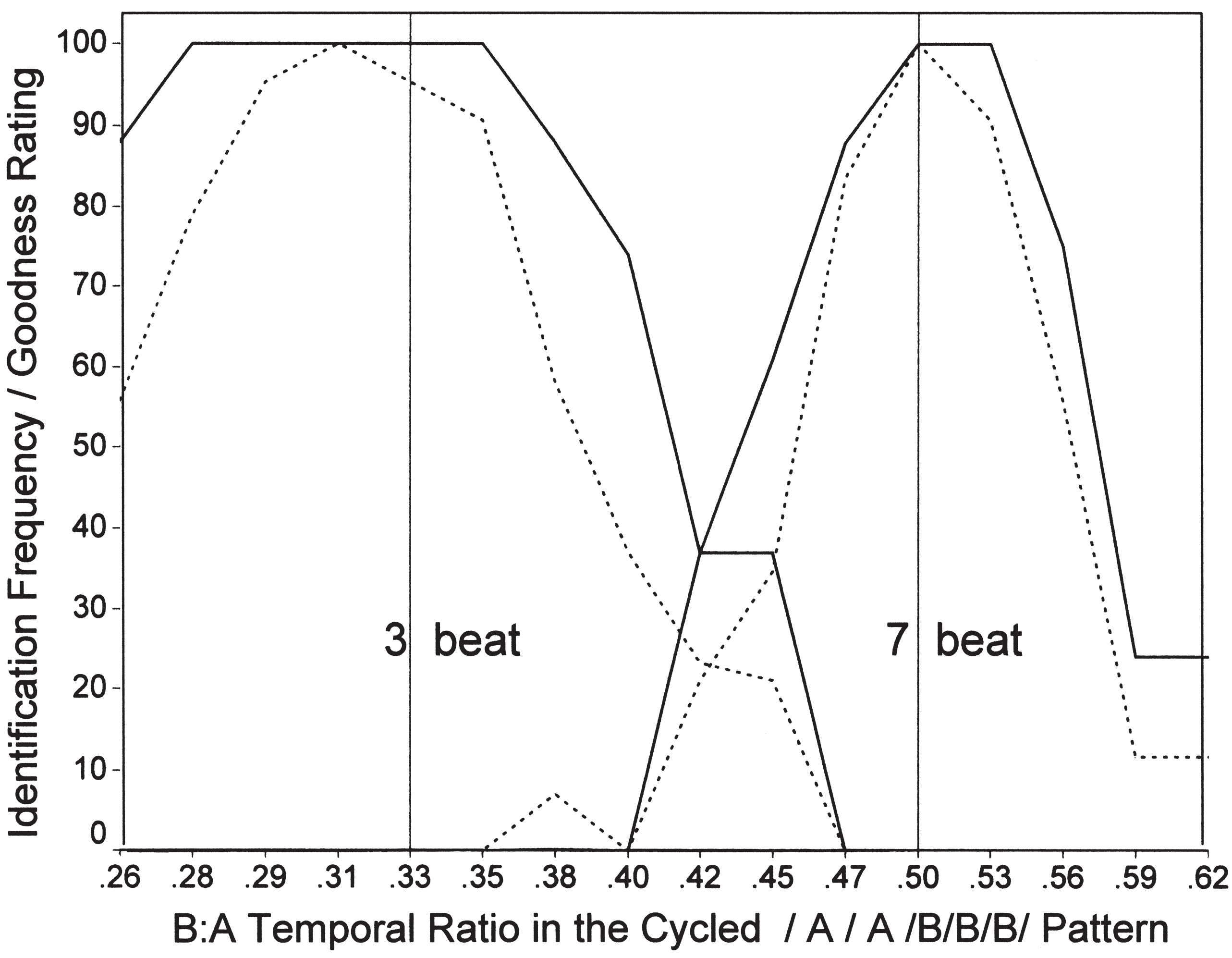

Fig. redu. 2.5%

FIG. 3. Redrawing of a figure provided by Papadelis (personal communication, January, 2005). The solid lines are the identification functions for the rhythmic categories. The dotted lines represent the rhythmic goodness ratings (the ratings between 0 and 6 were rescaled). Reference lines for the integer b:a ratios of 1:3 and 1:2 are inserted.

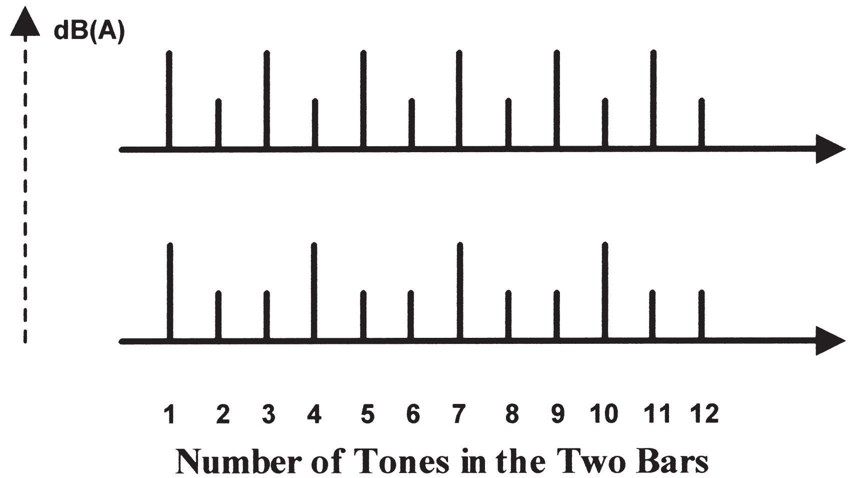

FIG. 4. Endpoint meters used by Windsor (1993). The upper meter is duple and the lower rhythm is triple. See text for further explanation.

and triple meters by manipulating the intensity of the sounds (and hence the dynamic accent) in isochronous tone sequences with inter tone intervals of 250 ms. Figure 4 shows the two endpoint meters, in which the accented tones were 73 dB(A), and the unaccented tones 59 db(A). There were nine intermediate patterns of which the 3rd, 5th, 9th, and 11th tones of the duple sequence were stepwise decreased in intensity, and the 4th and 10th tones stepwise increased in intensity, finally resulting in the triple sequence. In experiment 1, music students had to identify the 11 sequences (s1, s2, to s11) four times, the 44 presentations being randomized, from which the identification curve was established. The participants also had to perform a discrimination task according to the ABX psychophysical procedure. They heard 36 stimuli, each comprising three sequences, and had to indicate whether the last one (the standard X) was the same as either the first or second sequence (the comparisons A and B). Each stimulus consisted of a discrimination test across nonadjacent pairs of sequences (s1/s3, s2/s4, to s9/s11).

It turned out that the identification and discrimination curves were not so easy to interpret. The identification curve was shallower than the ones typically encountered in classic categorical perception studies, and the discrimination function had no clear peaks. Windsor ran a second experiment in which the identification procedure was the same, but the discrimination task comprised all adjacent ABX pairs. The discrimination function now showed two pronounced peaks in contrast to experiment 1.

The discrimination scores were highest around s3 on the one hand, and around s8 on the other hand. Windsor's interpretation was that those peak scores suggest two boundaries, separating three sets of patterns: a middle set of nonmetrical patterns, embraced by a set of duple (s1, s2, s3), and a set of triple patterns (s9, s10, s11). Such an interpretation nicely squares with the observation that the middle part of the identification function was shallow. As Clarke (2000) commented on Windsor's study: "The idea of a non-metrical category within a categorical perception study of meter might seem to undermine the idea of categorical perception, but it is important to note that the central region was not simply an undifferentiated area of variable perception. The absence of meter is, after all, a perfectly legitimate and quite perceptually striking rhythmic effect." (p. 6).

Windsor's study, and Clarke's comment on it, is a very suitable opportunity to now introduce the purpose of the present study. In Nakajima et al. (1992) and Sasaki et al. (1998) we claimed to have found some evidence for categorical perception in the auditory temporal domain as a consequence of an illusion of time perception which we named time-shrinking (TS). Whereas students of categorical perception often suppose and find that two (or more) adjacent perceptual categories result from gradual variation along some physical dimension, we found evidence for just one temporal 1:1 category embraced by regions of continuous temporal perception. Our present purpose is to further this evidence. Another important purpose is to investigate what happens when /t1/t2/ patterns are varied gradually between end point rhythms of 1:1 and 1:2, to our best knowledge not investigated before. Clarke (1987) and Large (2000) varied their /t1/t2/ stimuli between

1:1 and 2:1, and Parncutt (1994) discussed the perceptual difference between nominal 2:1 and 3:1 ratios. Schulze (1989) and Desain and Honing (2003) studied /t1/t2/t3/ patterns.

To introduce the four experiments to be reported, we explain the TS-illusion and describe our studies which showed evidence that TS might cause categorical perception in the temporal domain. Nakajima, ten Hoopen, and van der Wilk (1991) was the first study in which we reported TS: When two silent intervals, marked by three short sounds, neighbor each other and the first interval (t1) is shorter than the second interval (t2), the latter can be underestimated to a considerable degree. This underestimation is at maximum when t2 - t1 is approximately 80 ms and suddenly disappears when t2 - t1 exceeds 100 ms. Furthermore, the illusion appears most strongly when the two-interval pattern is relatively fast. If t1 gets longer than about 200 ms, the amount of underestimation, though still present, is far less (Nakajima et al., 2004). We give two representative examples from previous studies: In a /160/240/ pattern, t2 is underestimated by about 50 ms, and in a /320/400/ pattern, t2 is underestimated by about 20 ms. If one wants to listen to this phenomenon, one can consult audio demonstration #26 at http://www.design.kyushu-u.ac.jp/∼ynhome/ENG/Demo/illusions2nd.html. Eleven /t1/t2/ patterns are presented consecutively, and the first pattern is /160/160/. In each next pattern, t1 is reduced by 10 ms, and t2 increased by 10 ms. As long as the difference between t2 and t1 does not exceed about 100 ms, both neighboring durations are perceived as almost equal, and the heard impression is typically a 1:1 ratio.

We studied the illusion in a variety of ways and for details we refer to Nakajima et al. (1991), ten Hoopen et al. (1993), ten Hoopen et al. (1995), Sasaki et al. (2002), and Nakajima et al. (2004). However, one way the illusion was investigated is pertinent to the present study and should be explained. In Nakajima, ten Hoopen, Hilkhuysen, and Sasaki (1992), t1 was fixed at 50 ms, and t2 varied between 40 ms and 280 ms. The results, gathered by the method of adjustment, showed that when t2 was as long as t1, thus 50 ms, its point of subjective equality (PSE) was 47 ms. When t2 was twice as long, thus 100 ms, its PSE was 62 ms. Thus, when t2 increased from 50 to 100 ms, its PSE increased at a much slower rate from 47 to 62 ms. Apparently, t2 assimilated to a great extent to t1, and because t1 did hardly or not assimilate to t2, TS can be conceived of as unilateral temporal assimilation.

Sasaki, Nakajima, and ten Hoopen (1998) attempted to find more support for the notion that TS causes a perceived 1:1 category, despite of changing physical ratios. Of the three psychophysical experiments reported, we discuss the two most salient ones. In experiment 1, there were four conditions, defined by total pattern duration (t1 + t2), which were 90, 180, 360, and 720 ms. In each condition, the total duration was kept constant, but the temporal ratio between the two time intervals was varied in small steps. Participants were required to adjust a comparison time interval (tc) to the durations of t1 and t2 in the /t1/t2/ patterns (of course in separate sessions) as many times they needed until satisfied by the subjective equality between tc and t1, or tc and t2. These final tc's were the PSEs of t1 and t2. We will only discuss the results of the 180 ms condition, because that was the only total duration common to the experiments. Clear TS was found in the patterns /45/135/, /60/120/, and /75/105/, that is, t2 was underestimated as expected. Furthermore, hardly any assimilation of t1 to t2 was found. Only in the /105/75/ pattern, t1 assimilated by about 15 ms to t2, and we will discuss that unexpected result in the general discussion. The most important information, however, was not so much knowledge of the PSE-values of t1 and t2 per se, but rather of their ratio. It turned out that the ratio between the PSEs of t1 and t2 lay very close to 1:1 (between 0.9 and 1.1) although the physical ratio changed from /60/120/ to /105/75/ (from 0.5 to 1.4).

In Sasaki et al's (1998) experiment 3, just noticeable displacements of an intervening sound marker at several positions within an interval of 180 ms were established by a transformed up-down method in an ABX-paradigm. The results showed that listeners had an easy time discriminating the patterns /30/150/, /40/140/, /50/130/, and the patterns /110/70/, /120/60/, /130/50/, /140/40/, /150/30/, as compared to the difficult discrimination between the intermediate patterns /60/120/, /70/110/, /80/120/, /90/90/, and /100/80/. The authors argued that this was still another estimate of the 1:1 temporal category and noticed that the category was asymmetric with respect to the objective 1:1 ratio (/90/90/).

In the present study, it is our aim to estimate the size and asymmetry of the category by the methods of paired comparisons and cross-modality matching. The data of the first three experiments, in which the method of paired comparisons was applied, will be analyzed by hierarchical cluster analysis (HCA) and multidimensional scaling (MDS). Both methods are often used in a complementary fashion to reveal the underlying structure of the data (e.g., Davison, 1992). In the fourth experiment a cross-modal absolute judgment task (auditory to visual) is utilized to inspect whether its results converge with those of the paired comparison tasks.

Experiment 1

Pairs of /t1/t2/ patterns had to be rated for temporal similarity. The value of t1 was fixed at 160 ms, and t2 varied from 160 to 330 ms. Our prediction was that patterns in which t2 - t1 does not exceed about 80 ms, should fall into a 1:1 temporal ratio category. We expected the HCA and MDS to converge upon the size of the category.

Method

PARTICIPANTS

Four male students of psychology at Leiden University participated. Two students were paid for their services, and the other two obtained curriculum credits. Their ages ranged between 22 and 26, and they had normal hearing according to a screening with a pure tone audiometer.

STIMULI AND DESIGN

There were 18 experimental temporal patterns. Each pattern comprised of three short sound bursts that marked the two empty intervals t1 and t2. The sound bursts were approximate square waves of 1000 Hz that lasted 10 ms and started and stopped at zero-crossing. The sound level of the bursts was 93 dBA when played continuously. The first interval, t1, had a constant duration of 160 ms, and the second interval, t2, varied from 160 ms to 330 ms in steps of 10 ms, intervals measured from onset to onset. We called these experimental patterns "short-long" (/160/160/, /160/170/, /160/180/, . . . , /160/330/). The 18 control patterns were their "long-short" temporal mirror images (/160/160/, /170/160/, /180/160/, . . . , /330/160/). A stimulus pair comprised of two temporal patterns from the same set, that is, either from the set of experimental "short-long" patterns, or from the control set of "long-short" patterns. Within each set, all $18 \times 18 = 324$ possible combinations were used as stimuli. The time that elapsed between the onsets of the two consecutive patterns in a stimulus pair was 2 seconds. The timing of the patterns was calibrated by recording them on tape and measuring the sound sample patterns (by Goldwave v. 4.0). The temporal deviations were negligible.

PROCEDURE AND APPARATUS

A block contained the 324 short-long and the 324 long-short stimulus pairs. The participants did the block six times (in which the 648 pairs were randomized differently), divided over 12 sessions and the first block served as training. Each session, in which a half block was done, started with six warm-ups. The participants were instructed to judge the temporal similarity between the two patterns in the stimulus, which was presented only once, by keying a numerical judgment on the data entry board of the computer. They were allowed to use any range of numbers they preferred, under the condition that more similar patterns should receive lower numbers, more dissimilar patterns higher ones. The participants were required to ignore the total duration of the patterns, and to base their similarity ratings just on their temporal ratio impression.

The sound patterns were generated by a Commodore Amiga 500+ computer, routed via an amplifier (JVC AX11), and presented to the left shell of headphones (AKG-K140). The level of the sound markers was measured by a sound level meter (Brüel & Kjaer, 2203), mounted with an artificial ear (Brüel & Kjaer, 4152). The experiment was completely controlled by a Basic program.

Results and Discussion

The data of the training block were not analyzed. We first established the individual similarity matrices by averaging the five ratings in each cell, and because symmetric entries above and below the main diagonal concern the same temporal pattern pairs (only the presentation order of the two patterns to be compared was the reverse), we averaged once more over the upper and lower triangles of the matrices. This procedure seemed warranted in view of the high correlations between the upper and lower triangles of the similarity matrix. In the short-long matrix the Pearson-correlation between the 153 ratings in the upper triangle and their 153 reverse order ratings in the lower triangle was 0.94 ($p < 0.01$, two-tailed). In the long-short matrix, this correlation was 0.96 ($p < 0.01$, two-tailed). Recall that the rating scale was free and the numeric ranges indeed differed between the four participants. We normalized the ratings (between 0 and 1) and averaged them over the four participants to get the final "short-long" and "long-short" matrices. Thus, the final cell values in these matrices were each based on 40 ratings.

The two similarity matrices were submitted to a HCA. Because we were more interested in the main partitions between clusters than in hierarchy, we ran the HCAs with the least hierarchical algorithm available in SPSS, Ward's method (e.g., Everitt, 1993). Figure 5 shows a cluster of seven patterns (/160/160/ through /160/220/ as indicated by the gray rectangle) at the left side of the main cluster partition of the "short–long" dendrogram. Notice that there is no hierarchy at all within this cluster, contrary to the other main cluster in

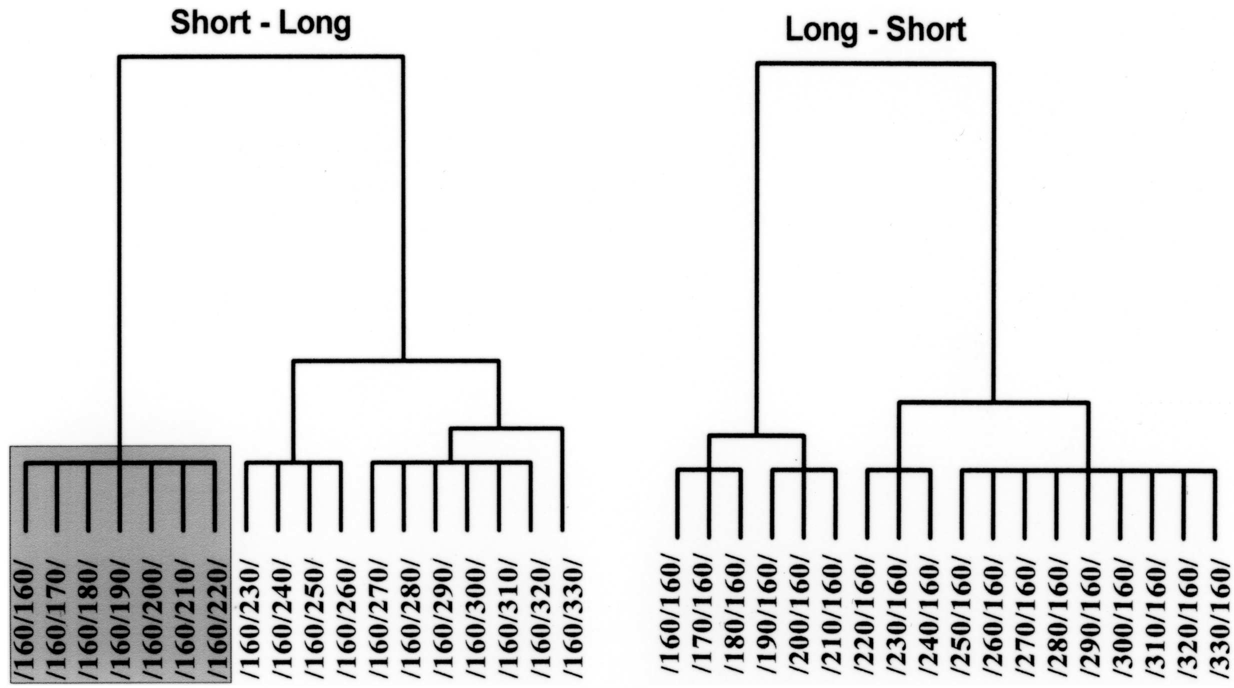

FIG. 5. Experiment 1: Dendrograms for the experimental "short-long" and the control "long-short" conditions as established by hierarchical cluster analyses. The time-shrinking cluster in the "short-long" condition is indicated by the gray-shaded rectangle.

the dendrogram, which is extra support for the perceptual equivalence of the shrunk cluster members. Also note that the control dendrogram of the "long-short" patterns, not liable to TS, differs in structure from the experimental dendrogram. The cluster left to the main partition (/160/160/ to /210/160) is smaller and has still internal hierarchy, indicating that the cluster members are less equal than those in the gray-shaded cluster (cf. Figure 5).

The two similarity matrices were also submitted to a MDS (SPSS, Alscal), which were run for one dimension, on an ordinal level using Euclidian distances, to get scales of temporal ratio similarity of the patterns. The stress values were 0.06 (Rsq = 0.98), and 0.056 (Rsq = 0.99) for the "short-long" and "long-short" conditions, respectively. For the experimental "short-long" condition, we expected that the scale distances between temporal patterns transformed by TS were relatively small as compared to those between their mirror patterns in the control "long-short" condition. We also expected that the distance on the "short-long" scale between shrunk and nonshrunk patterns would be relatively large.

Figure 6 shows that the largest distance on the similarity scale in the "short-long" condition was between patterns /160/230/ and /160/240/. Thus, the eight patterns /160/160/ through /160/230/ appear to fall in a category, indicated by the solid ellipse, and the dotted ellipse is the HCA estimate. Thus, the estimates of the category by the scaling and cluster analyses differed by one pattern only. It is indicated by the shallow lines connecting the upper and lower scale in Figure 6, that the scale distances between the shrunk patterns are much smaller than the distances between their "long-short" mirror patterns.

We performed two tests whether the spacing of the eight short-long patterns /160/160/ through /160/230/ on the upper scale of Figure 6 differed from the spacing of their mirror patterns /160/160/ through /230/160/ on the lower scale of Figure 6. In the first test, the two sets (short-long and long-short) of the seven distances between their eight z-scores were submitted to a paired samples t-test. It turned out that the eight short-long patterns had been heaped up significantly more than their eight mirror patterns ($p < .025$, one-tailed). We subsequently submitted the two sets of 28 similarity ratings in the short-long and long-short conditions, underlying the scale distances, to a Wilcoxon test (note that a t test is not appropriate here because the ratings are ordinal). It turned out that the short-long ratings were significantly smaller (= more similarity) than their long-short counterparts ($p < 0.0001$, one-tailed). Both tests and mere visual inspection clearly show that

Fig. redu. 9%

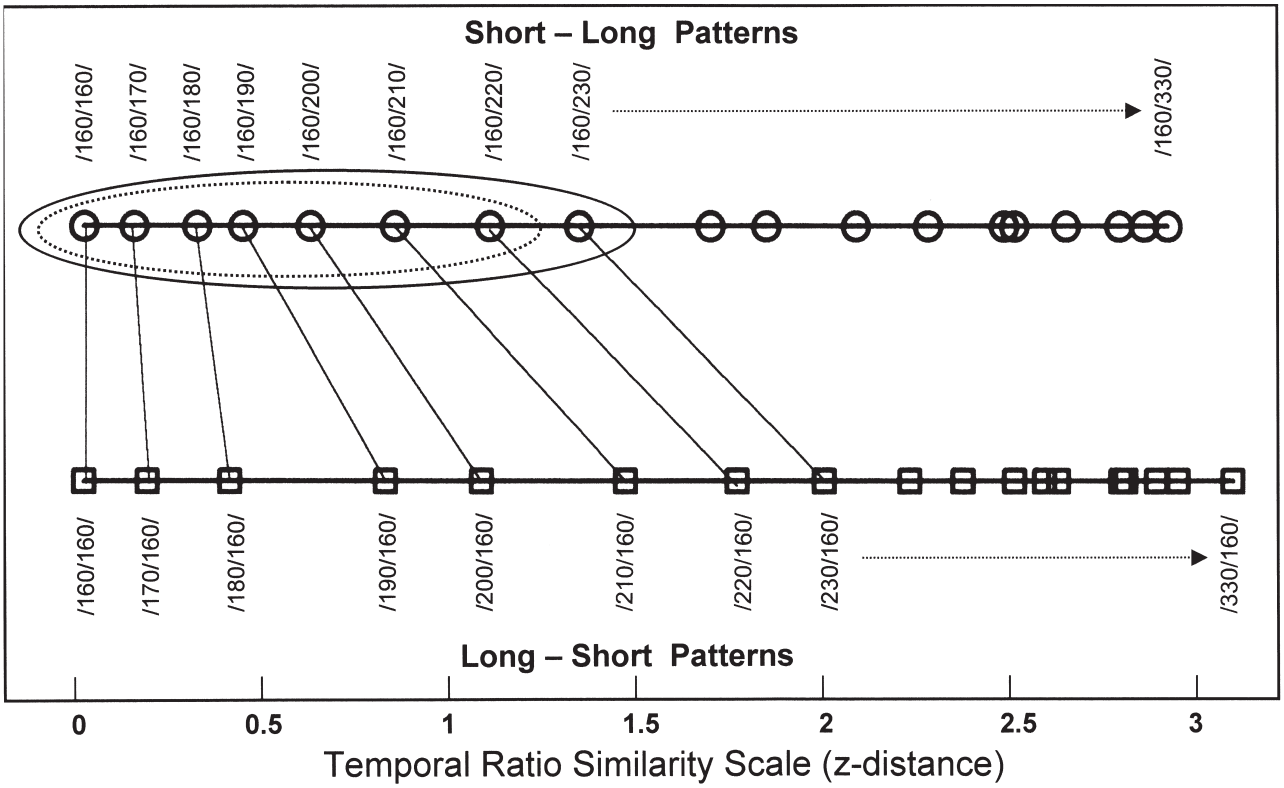

FIG. 6. Experiment 1: Temporal ratio similarity scales for the experimental "short-long" condition and the control "long-short" conditions. The inserted solid ellipse indicates the time-shrinking category estimate of the scaling analysis. The inserted dotted ellipse indicates the estimate of the cluster analysis. The connecting lines between mirror patterns in both conditions illustrate the piling up of the shrunk "short-long" patterns.

patterns /160/160/ through /160/230/ constitute a category.

One might wonder whether the average pattern of results displayed in Figures 5 and 6 is representative for all four subjects. We analyzed the individual data in the same way, thus by HCAs and MDSs. Figure 7 shows the individual similarity scales as yielded by the MDS. The solid ellipses enclose the shrunk patterns separated by the largest distances from the nonshrunk patterns. The inserted dotted ellipses represent the shrunk clusters as estimated by the HCA (dendrograms not shown here). Notice that this latter estimate differs only one pattern for three participants and was the same for participant M.S. We submitted the individual similarity ratings in the short-long categories and their long-short counterparts to Wilcoxon-tests, in the same vein as done with the overall data, and for each participant the difference was significant ($p < 0.002$ [B.M.], $p < 0.02$ [G.R.], $p < 0.0001$ [M.B.], $p < 0.0001$ [M.S.], all one-tailed).

The present experiment was set up to find further evidence for a 1:1 TS category, and we only predicted that patterns at the left side of the upper scale would heap up more than their mirrors on the lower scale (cf. Figure 6). We made no specific predictions about the patterns at the right end of the scales. It is interesting to observe that those patterns piled up disproportionately, and paradoxically even more than those in the time-shrinking category. Also curious is that this happened to a stronger extent in the "long-short" than in the "short-long" condition. We will attempt to give an explanation for these facts in the general discussion.

Experiment 2

It is conceivable that, even though the participants in Experiment 1 were instructed to focus only on the ratio aspect of the temporal patterns, and were required to neglect the total duration of the patterns as well as possible, pattern duration might have affected their similarity ratings. To exclude such a possibility, we designed an experiment in which the ratio between t1 and t2 systematically varied in small steps but the pattern durations (t1 + t2) remained constant.

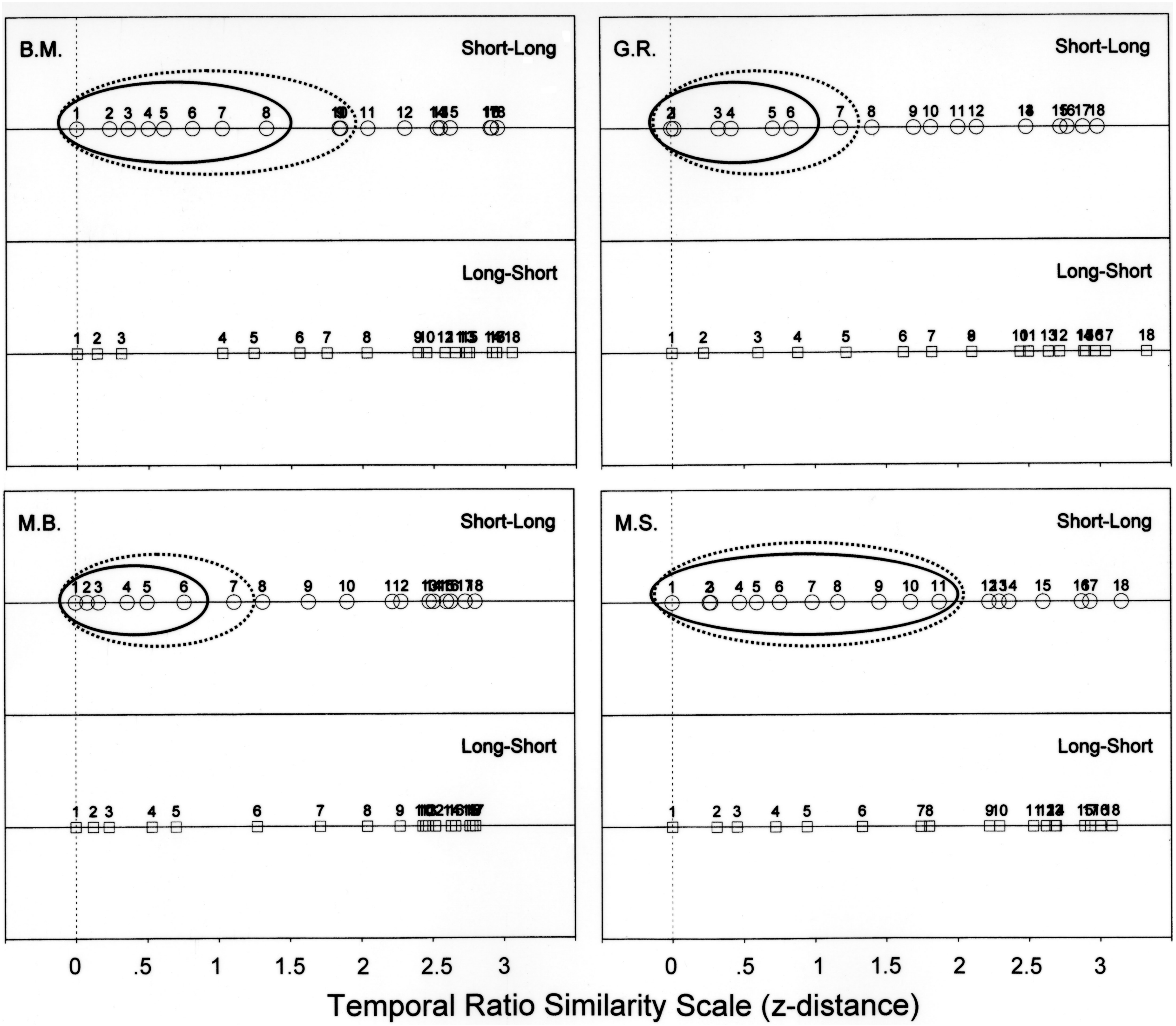

Fig. redu. 10%

FIG. 7. Experiment 1: Temporal ratio similarity scales for the experimental "short-long" condition and the control "long-short" conditions for each of the four participants. The inserted solid ellipses indicate the time-shrinking category estimates of the scaling analysis. The inserted dotted ellipses indicate the estimates of the cluster analysis.

Method

PARTICIPANTS

Five students of music at Miyagi Gakuin Women's University served in the experiment and were paid for their services. Their ages ranged between 20 and 23 and they had normal hearing.

STIMULI

A stimulus contained two patterns, which each consisted of three sounds that marked two empty time intervals. The sound markers were sine waves of 3000 Hz, had a duration of 7 ms including 3-ms rise and fall times, and had an intensity of 70 dBSPL. All /t1/t2/ patterns had the same total duration of 320 ms onset to onset. The value of t1 varied between 20 and 300 ms in steps of 5 ms and consequently, the value of t2 varied complementary between 300 and 20 ms in steps of 5 ms. This resulted in 57 patterns (/20/300/, /25/295/, . . . , to /295/25/, /300/20/). All 57 patterns were paired, thus there were $57 \times 57 = 3249$ pattern pairs.

PROCEDURE AND APPARATUS

There were 66 sessions, of which the first one was a training session containing a random selection of 52 pattern pairs from the 3,249 possible pairs. For the

65 experimental sessions, the 3,249 pattern pairs were randomized. Each session started with 2 warm up trials except for the first one which started with 3 warm-ups. Consequently, all sessions contained 52 trials, and took about 10 minutes each.

The participants were instructed to judge the similarity between the two patterns in each trial by a 5-point rating scale from 1(completely equal) to 5 (completely different). They had to write their rating on response sheets. The pattern pairs were presented every 10 seconds. The stimuli were generated by a computer (Dell, Latitude CP M233ST) and presented to the left shell of the headphones (Stax, SR-lambda professional).

Results and Discussion

We constructed a 57 × 57 similarity matrix averaged over the five participants. First, we ran the same type of HCA as in Experiment 1. Figure 8 shows the resulting dendrogram. The gray-shaded rectangle, the cluster with no internal hierarchy at all, contains time-shrunk patterns from /115/205/ through /155/165/ plus the patterns /160/160/ through /185/135/.

The 57 × 57 similarity matrix was also submitted to a one- and a two-dimensional MDS (SPSS, Alscal), Euclidian model and ordinal level of measurement. The one-dimensional solution could hardly be read due to much visual clutter (compared to the 18 patterns in Experiment 1, we had 57 patterns now). The two-dimensional representation (stress = 0.11, Rsq = 0.95) showed the typical 'horseshoe'-pattern (see Figure 9). Kruskal and Wish (1978, Appendix B) said of such horseshoes that: "it is apparent that the two-dimensional configuration has a special character: it consists of a nearly one-dimensional configuration which has been bent around into a horseshoe shape. Only one curvilinear dimension is sufficient to give a reasonable description" (p. 89). More recently, van der Kloot (1997) explained that one gets such horseshoes when the underlying stimuli are essentially one-dimensional but the scaling program is forced to calculate a representation in two dimensions, and it accomplishes this by putting the squared coordinates of the first dimension on the second one, yielding the parabolic or horseshoe curvature.

Given these expert advices, we decided to display the two-dimensional representation, and interpreted it in a one-dimensional manner along the bent scale. First, we inserted the HCA-estimate of the 1:1 category (cf. Figure 8) as the dotted ellipse in Figure 9. Within this dotted ellipse we established the group of most crowded patterns by the following procedure: In the top of the horseshoe we started at pattern /160/160/ (the objective 1:1 pattern) and searched for relatively big distances between patterns at either side. At the right side this was the distance between /115/205/ and /110/210/, which was twice as big as the distance between /120/205/ and /115/205/. On base of this criterion we decided that pattern /115/205/ just fit the

Fig. redu. 7%

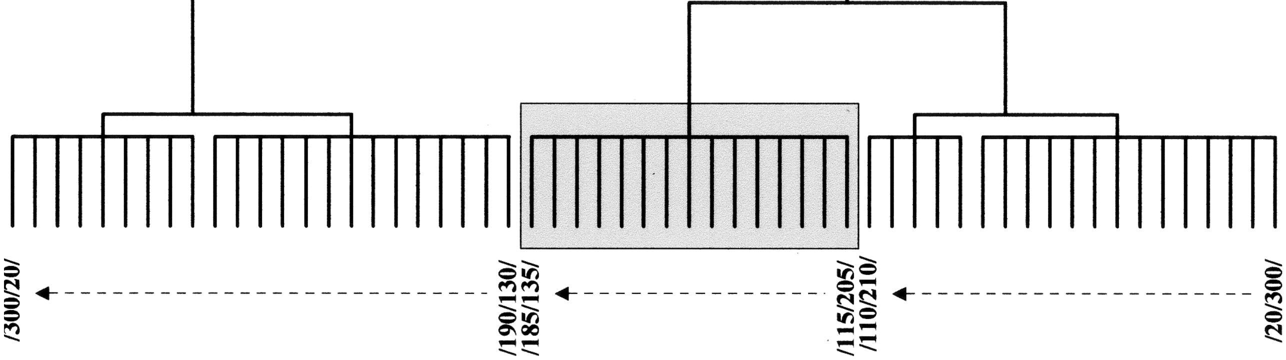

FIG. 8. Experiment 2: Dendrogram for the patterns /20/300/, /25/295/, . . . , to /295/25/, /300/20/, as established by a hierarchical cluster analysis. The estimate of the time-shrinking cluster is indicated by the gray-shaded rectangle.

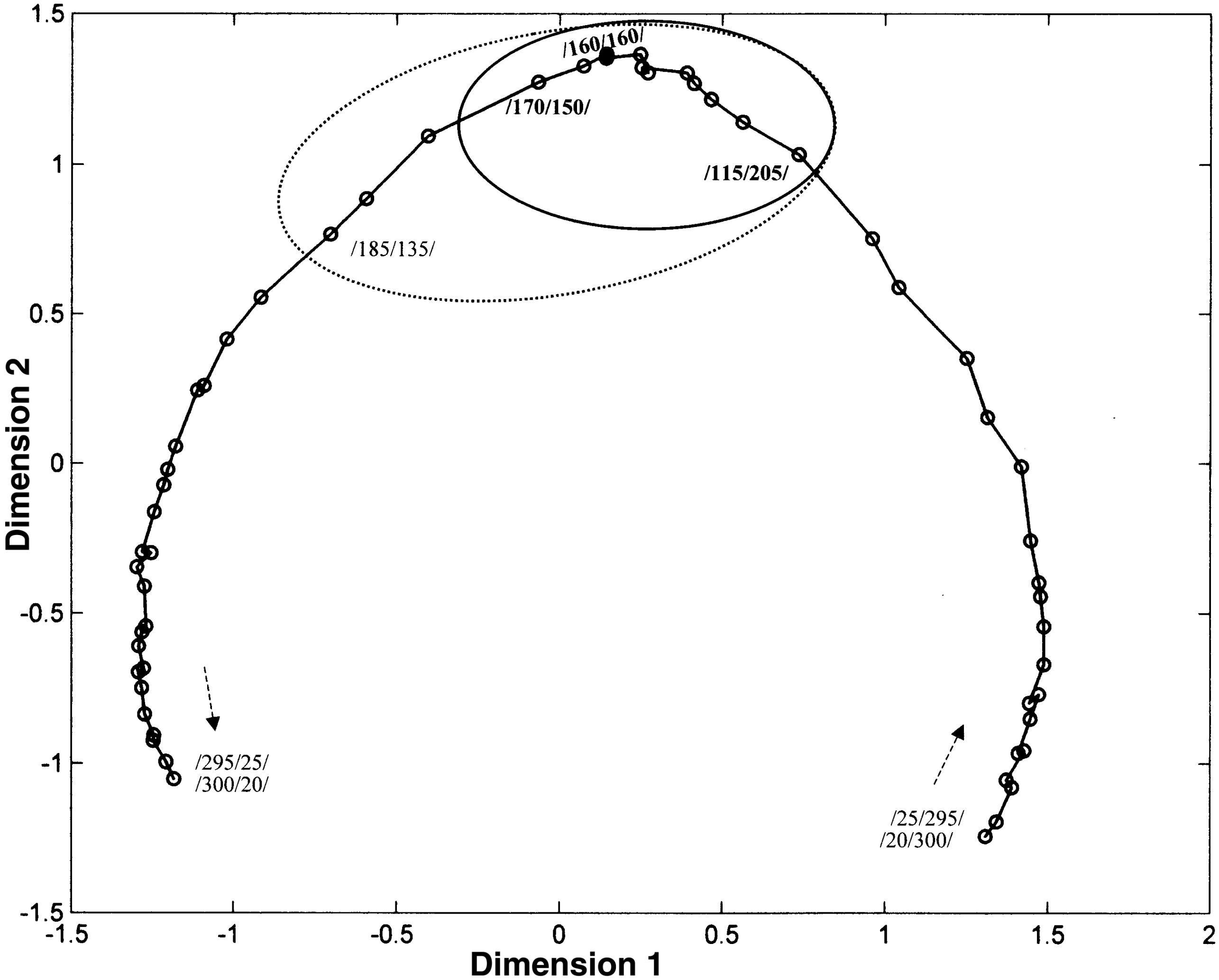

Fig. redu. 28%

FIG. 9. Experiment 2: Temporal ratio similarity scale in horseshoe-form for the patterns /20/300/, /25/295/, . . . , to /295/25/, /300/20/. The estimate of the time-shrinking category on the horseshoe scale is indicated by the solid ellipse, and the estimate by the cluster analysis is inserted as dotted ellipse.

category. On the left side of /160/160/ the distance between /175/145/ and /170/150/ was almost three times as big as the preceding one between /165/155/ and /170/150/, and we decided that the latter pattern just fit the category. The size of the category (/115/205/ to /170/150/) is indicated by the solid ellipse in Figure 9. The category contains typically shrunk patterns (/120/200/ through /155/165/) plus the patterns /160/160/, /165/155/, and /170/150/.

Experiment 3

This experiment had the same purpose as Experiment 2, but was conducted in the Netherlands with Dutch participants.

Method

PARTICIPANTS

Eight female students of Leiden University served in the experiment and were paid for their services. Their ages ranged between 18 and 23 and they had normal hearing. They had never received any music training.

STIMULI AND DESIGN

The stimulus patterns were the same as those in Experiment 2 except for the sound markers, which were the same as those of Experiment 1. The 3249 stimulus pairs were divided over four blocks because of a memory capacity limit of the computer. In block 1, patterns /20/300/ to /160/160/ were paired with patterns from

/20/300/ to /160/160/. In block 2 patterns /160/160/ to /300/20/ were paired with patterns /20/300/ to /160/160/. In block 3 patterns /20/300/ to /160/160/ were paired with patterns /160/160/ to /300/20/. In block 4 patterns /160/160/ to /300/20/ paired with patterns /160/160/ to /300/20/. Thus, all blocks contained $29 \times 29 = 841$ pattern pairs.

PROCEDURE AND APPARATUS

The first session was devoted to audiometric screening, ample instruction and training with one of the four blocks, more specifically, the block that would be the final experimental block for the participant. After that, the participants were required to pass through the four blocks in another four sessions, done on different days. The sessions lasted approximately two hours, including two obligatory breaks. The order of blocks was permuted according to a Latin square. Within each block, the 841 pattern pairs were randomized. The participants were instructed to press their similarity rating (1–5) on the data entry keyboard of the computer. Two seconds after they had keyed their judgment, the next trial was presented. The equipment equaled that of Experiment 1.

Results and Discussion

The eight most extreme /t1/t2/ patterns, /20/300/ through /35/285/, and /300/20/ through /285/35/ were discarded from the analysis. Several participants stated about these patterns that they had difficulties in hearing the first, or last, two sound markers separately. Recall that the duration of the sound markers was 10 ms, which means for instance for pattern /20/300/ that the first silent empty interval is only 10 ms. Several times it occurred that patterns like for example /20/300/ and /300/20/, physically very different, were nevertheless judged equal because closely adjacent markers fused and hence two equal single intervals were heard. A physical reason might be that the present marker duration was 10 ms as compared to 7 ms in Experiment 2 and that the headphones were of slightly lower quality than those in Experiment 2. We believe, however, the main reason to be the fact that the participants in the present experiment were not musically trained at all, whereas those in Experiment 2 were students of music. To get rid of the contaminated judgments, we restricted the range of patterns to be analyzed to 49, that is, from /40/280/ through /280/40/ ms.

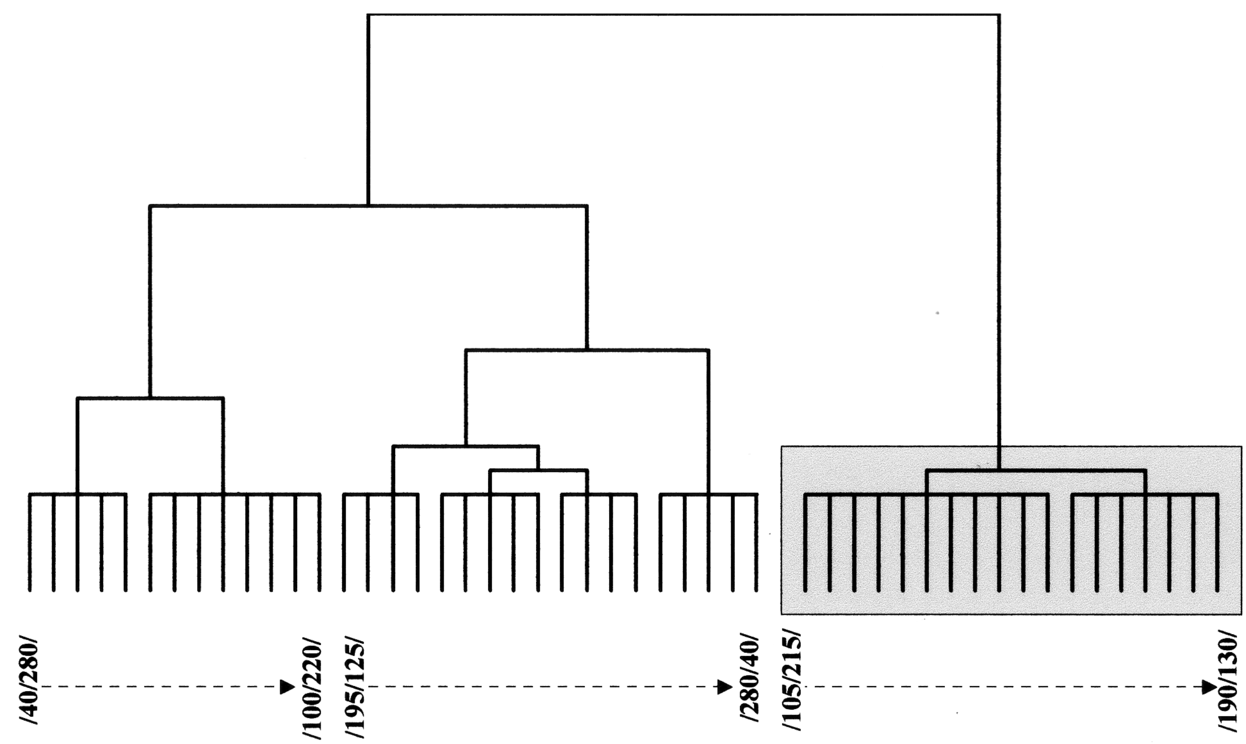

FIG. 10. Experiment 3: Dendrogram for the patterns /40/280/, /45/275/, . . . , to /275/45/, /280/40/, as established by a hierarchical cluster analysis. The estimate of the time-shrinking cluster is indicated by the gray-shaded rectangle.

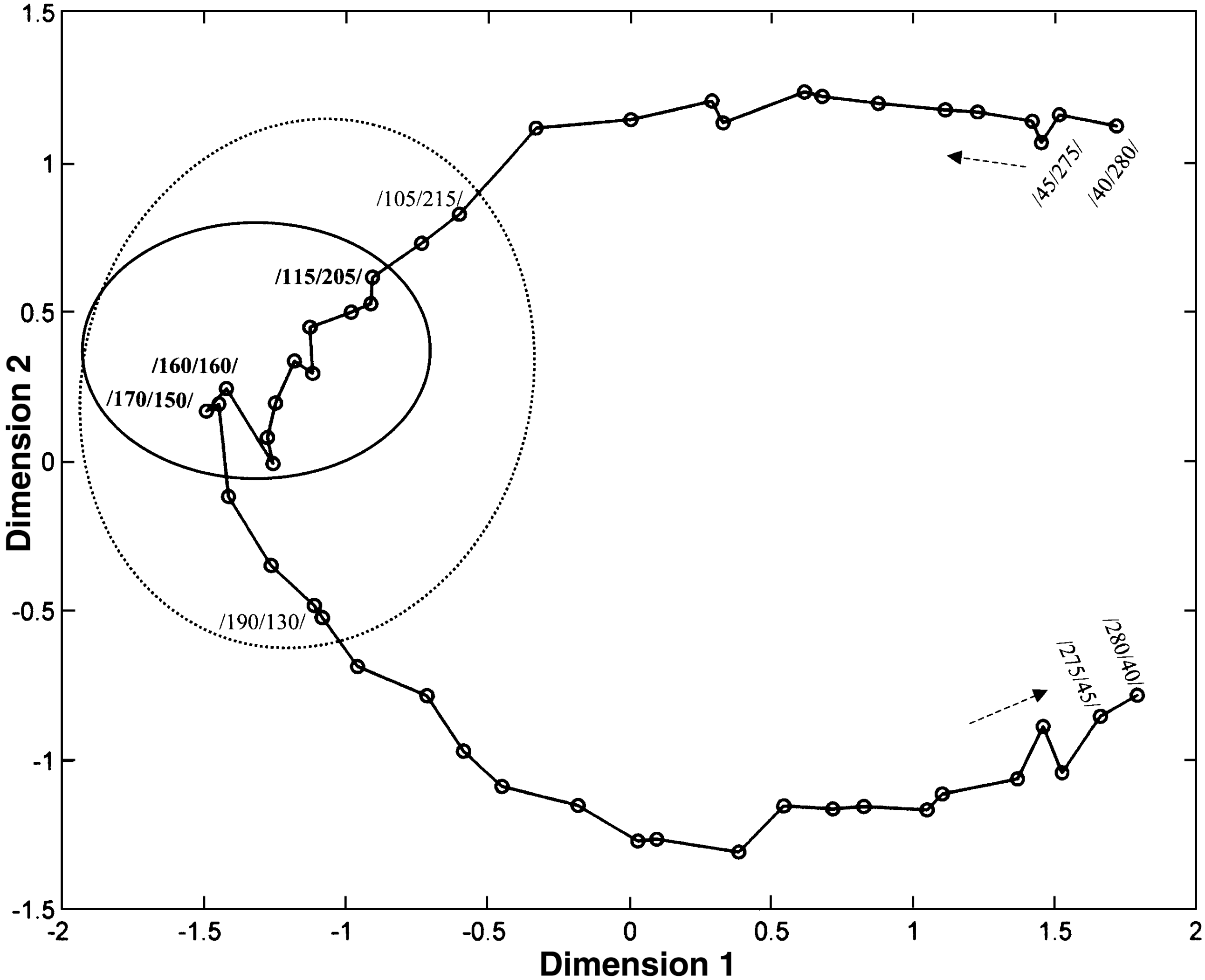

Fig. redu. 25.5%

FIG. 11. Experiment 3: Temporal ratio similarity scale in horseshoe-form for the patterns /40/280/, /45/275/, . . . , to /275/45/, /280/40/. The estimate of the time-shrinking category on the horseshoe scale is indicated by the solid ellipse, and the estimate by the cluster analysis is inserted as dotted ellipse.

A 49 × 49 similarity matrix was formed by adding the similarity ratings of the eight participants in each of the 2401 cells. The similarity matrix was analyzed in the same way as in Experiment 2. The dendrogram is displayed in Figure 10. The tree structure shows that the biggest three clusters comprise respectively patterns /40/280/ to /100/220/, patterns /105/215/ to /190/130/, and patterns /195/125/ to /280/240/. It is the gray-shaded cluster /105/215/ to /190/130/ that contains the shrunk patterns plus the patterns /160/160/ through /190/130/. Notice that this cluster does not deviate much from its counterpart in Experiment 2 (/115/205/ to /185/135/). In both dendrograms (Figures 8 and 10) it can be seen that the gray-shaded clusters not only contain patterns liable to TS but also the bordering patterns /160/160/ to /185/135/ or /190/130/, not liable to TS. The reason that these patterns nevertheless can belong to the 1:1 category will be explained in the general discussion.

Figure 11 shows the representation of the patterns as yielded by the MDS and again a clear horseshoe curvature can be seen (stress = 0.15, Rsq = 0.91). Like in Experiment 2, we first inserted the cluster estimate (cf. Figure 10) as the dotted ellipse and then inspected in the same way as done in Experiment 2 which patterns within this region heaped up most. The solid ellipse surrounds this latter group of patterns (/115/205/ to /170/150/). It is remarkable that, despite cultural differences (Japan and the Netherlands), and the huge difference in music training (well trained vs. not trained), the two solid curved estimates of the 1:1 category do not differ. This strengthens our conjecture that the perceptual process giving rise to the formation of the 1:1 category is very fundamental.

Experiment 4

Experiments 2, and 3 found support for the existence of an asymmetric 1:1 category as a result of TS by submitting paired comparison data to cluster and scaling analyses. The aim of the present experiment was to find further support for such a category by requiring participants to give cross-modal judgments (auditory–visuospatial) of the patterns presented in isolation.

Method

PARTICIPANTS

Eight students of music at Miyagi Gakuin Women's University served in the experiment. Their ages ranged between 18 and 21 and they had normal hearing and visual acuity. They participated on a voluntary basis and were paid 700 Yen per hour.

STIMULI AND DESIGN

The temporal patterns were the same as those of Experiment 2 (/20/300/, /25/295/, /30/290/, . . . , /290/30/, /295/25/, /300/20/ ms). Also the sound markers were the same.

PROCEDURE

There were five sessions, of which the first one was training. Each session comprised of the 57 patterns in random order. The participants listened to the auditory patterns which were generated by a computer (Dell, latitude CP M233ST) and presented to the left shell of Stax, SR-lambda professional headphones. The participants were provided with A4 paper sheets with on each sheet ten horizontal lines of 8 cm long separated vertically by 2 cm. Each horizontal line was marked by two vertical lines of 1 cm, one at each end. The participants were required to draw by pencil an intervening vertical line through the horizontal line scale such that the three visual markers reflected the judged ratio between the two neighboring intervals of the perceived auditory pattern. The participant could press the space bar on the keyboard to listen to the pattern as many times as she wanted. After pressing the spacebar there was a silence of 2 s before auditory presentation. When the participants felt certain of the perceived temporal ratio, they drew the intervening vertical line, and called the next pattern. In the four experimental sessions, three warm-ups preceded the 57 patterns. The intervening positions of the drawn lines were measured by a ruler and the reading precision was 0.5 mm. The sessions, held in a sound attenuated room, lasted about 40 minutes each.

Results and Discussion

Figure 12 shows the box and whisker plots of 32 subjective ratios (8 participants × 4 replications) for each of the 57 objective ratios. It can be seen clearly that there is a category of 1:1 rhythms. The medians of the boxes of the patterns /120/200/ through /170/150/ (the patterns between the vertical dotted lines) impeccably cover up the 0.5 reference line. Thus, even though the objective ratio (t1/ (t1 + t2)) changed widely from 0.375 to 0.531, the percept of these /t1/t2/ patterns remained 1:1. A strong case for the existence of the 1:1 category is that the variability of the judgments within the category is very small, and suddenly increases at the edges of the category and beyond. Notice that this category estimate (/120/200/-/170/150/), yielded by cross-modal judgments, hardly differs from the estimates in Experiments 2 and 3. Note further, as we already described in Sasaki et al. 1998, that the category range lies asymmetrically with respect to the objective 1:1 ratio (in Figure 12 expressed as 160 ms/320 ms = 0.5). Very recently, cross-modality matching (auditory–visuospatial) was also applied by Penel, Hollweg, and Brody (2005) to find evidence for rhythmic categorical perception.

General Discussion

ROBUSTNESS OF THE 1:1 CATEGORY

The present study supports our conjecture, which was based on phenomenological observation and psychophysical experiments, that TS causes a category of perceived 1:1 ratios. The 1:1 category emerged from Experiment 1, in which the participants judged pattern similarity on a free rating scale, from Experiments 2 and 3, in which a five-point rating scale was used, and from Experiment 4, in which participants made cross-modal judgments. Thus, the way of measurement did not affect the emergence of the category, and hardly its size and asymmetry. Another indication of the robustness of the categorization process is that the same patterns of results were found with participants studying music (Experiments 2 and 4), and participants with no music training at all (Experiment 3).

CASES OF BILATERAL ASSIMILATION

When the total pattern duration was 320 ms, our estimate of the 1:1 category was from patterns /115/205/ to

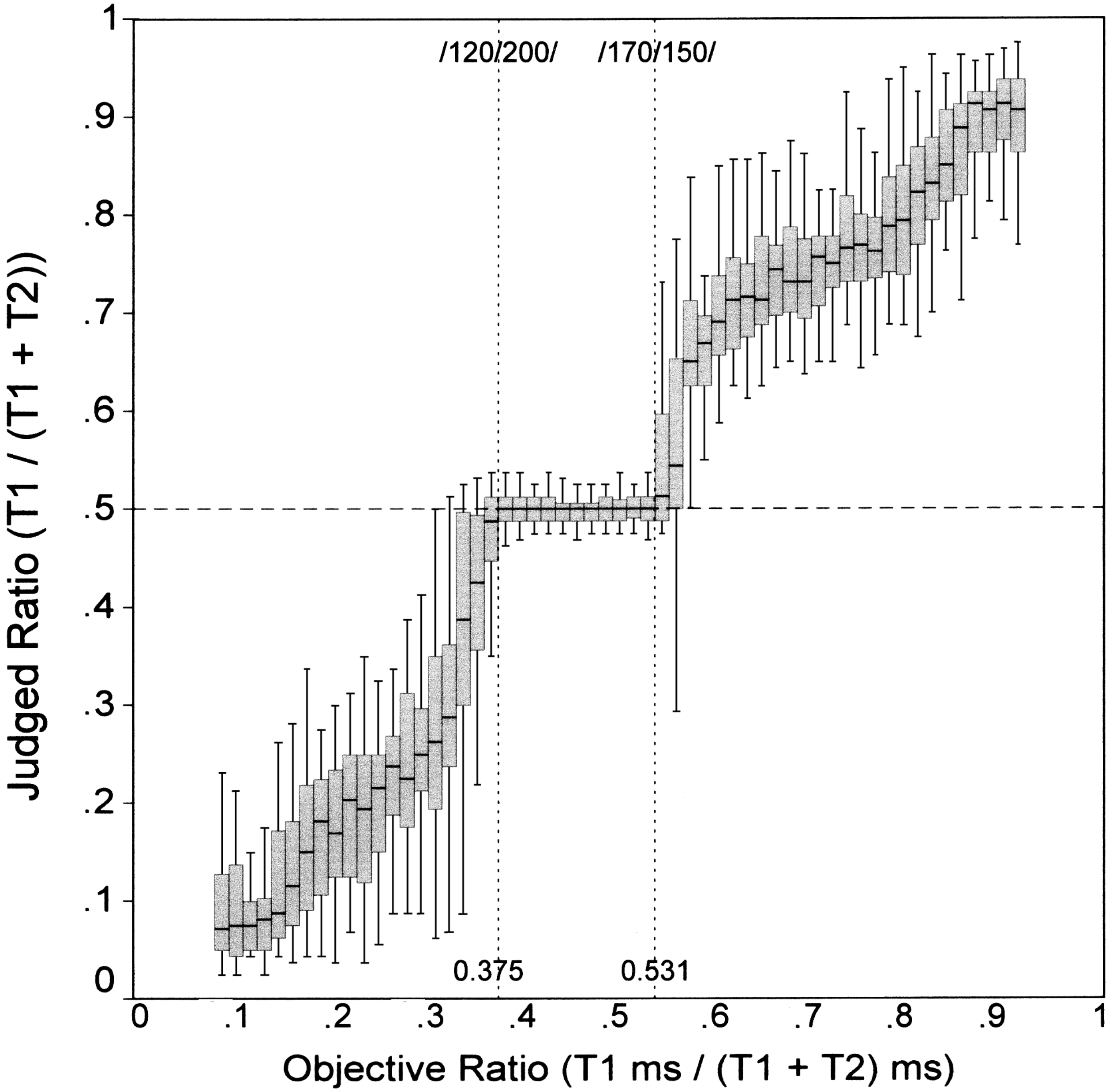

FIG. 12. Experiment 4: Judged ratio (JR = judged t1/(t1 + t2)) as a function of the objective ratio (OR = t1/(t1 + t2)) expressed as box and whisker plots of the 32 absolute judgments for the ratios of the 57 auditory patterns. The inserted vertical dotted lines indicate the asymmetric 1:1 category boundaries.

/170/150/ (Experiments 2 and 3), and from patterns /120/200/ to /170/150/ (Experiment 4). The 1:1 category comprises not only time-shrunk patterns but also contains the patterns /165/155/, /170/150/. One might wonder why these patterns are also member of the 1:1 category. As we mentioned in the introduction, Sasaki et al. (1998) found an unexpected result in their experiment 1 in which PSEs were established of t1 and t2 in /t1/t2/ patterns. In one of their patterns, namely pattern /105/75/, bilateral assimilation occurred: the PSEs of t1 and t2 were 88 and 82 ms respectively. Miyauchi and Nakajima (2005) did a systematic study to establish the temporal conditions under which bilateral assimilation took place. In one of their conditions the total duration (t1 + t2) was 240 ms, and the patterns varied in six steps from /60/180/ to /180/60/. The authors replicated TS: in the patterns /80/160/ and /100/140/, t2 was underestimated. But in addition to this unilateral assimilation, they also found bilateral assimilation. In pattern /140/100/, t1 was underestimated, whereas t2 was overestimated, both by approximately 10 ms, thus perceptually yielding a more regular pattern (/130/110/). Overall, their study showed that bilateral assimilation occurred when (t1 - t2) < 40 ms. It is apparent that such bilateral assimilation worked in the present patterns /165/155/ and /170/150/ as well, thereby reducing the temporal ratio between t1 and t2 and thus molding them to excellent candidates for the 1:1 category.

ANOTHER TEMPORAL RATIO CATEGORY?

It appeared in Experiment 1 (cf. Figure 6) that the /t1/t2/ patterns at the right end of the similarity scales piled up disproportionately, and paradoxically they even piled up more than the time-shrunk patterns at the left end of the "short-long" scale did. For the "short-long" patterns, this right end heaping up happened approximately from pattern /160/280/ on, and for the "long-short" patterns from pattern /260/160/ on. One might be tempted to interpret this piling up as categorical perception too and give the assumed categories the ratio labels 1:2 and 2:1. Such an interpretation, however, is not correct. Nakajima (1987), and Nakajima, Nishimura, and Teranishi (1988) showed that the subjective ratio between two empty consecutive intervals (t1 and t2) is *less extreme* than their corresponding physical ratio. According to Nakajima's processing time hypothesis, the subjective ratio between t1 and t2 is not (t1 : t2) ms, but (t1 + 80) : (t2 + 80) ms. Physical ratios of 160:320 ms and 320:160 ms therefore have subjective ratios of 240:400 and 400:240, that is, 3:5 and 5:3, thus less extreme than 1:2 and 2:1.

The processing time hypothesis might also explain why different ratios are less easily discerned (for example 2:1 from 3:1) than one would expect from their nominal value. Parncutt (1994) stated: "The difference between nominal 2:1 and 3:1 ratios may thus be categorically perceptible only if the difference is exaggerated. Perhaps this is why musicians exaggerate the difference between the two ratios in performance." (p. 47), but he attempted to explain this in terms of strength of metrical accent, even though the processing time hypothesis was already launched in 1987.

We suspect that the heaping up of temporal patterns at the right end of the scales in Figure 6 arose from discriminatory problems. As regards patterns containing more intervals, Friberg and Sundberg (1995), ten Hoopen (1992), and ten Hoopen et al. (1994) reported that up to about 250 ms, the *absolute* difference limen (DL) for temporal discrimination remains constant, but that beyond about 250 ms, the *relative* DL remains constant instead. In other words, Weber's law for temporal discrimination is only applicable beyond 250 ms. Given this hinge between a constant absolute DL and a constant relative DL, and given the fact that the step size of /t2/ (or /t1/) remains 10 ms throughout the whole scale, it is obvious that patterns at the right end of the scales pile up as a result of their bad discriminability (similar tendencies can be seen in Figures 9 and 11: patterns in the short-long and long-short tails of the horseshoe curves increasingly crowd). In an unpublished experiment we established the DLs for patterns /160/160/, /160/210/, /160/260/, /160/310/ by the method of limits, and found DLs of 21, 21, 23, and 28 ms respectively. The DLs for the mirror patterns were 20, 16, 31, and 31 ms respectively. These data show that the discriminability at the right end of the scales (cf. Figure 6) considerably decreases.

In the introduction we mentioned that we never came across categorical perception studies in which the endpoint rhythms were 1:1 and 1:2. It is interesting to ponder what would have happened in such an experiment. Given that the fictitious experimenter would have used a total pattern duration (t1 + t2) of 320 ms, then the 1:1 end point would have been 160:160 ms and the approximate 1:2 endpoint 105:215 ms. In fact, the stepwise varied patterns between these two endpoints were contained in our Experiments 2 and 3 (/160/160/, /155/165/, . . . , /110/210/, /105/215/ ms). On the basis of the results of these experiments (cf. Figures 9 and 11) we can now easily predict what our fictitious experimenter would have found: a boundary between patterns /115/205/ ms and /120/200/ ms. That is, the 1:1 category would have contained 10 patterns, and the 1:2 category only two patterns. We surmise that our fictitious experimenter may have existed but refrained from submitting such unexpected and hard to explain results.

CONSEQUENCES FOR TEMPORAL RATIO PRODUCTION

Though the present study was concerned with categorical perception, it has consequences for production. Experiments 2, 3, and 4, where t1 + t2 was 320 ms, showed that the most conservative estimate of the 1:1 category was from pattern /120/200/ through pattern /170/150/ ms. In the lower boundary pattern /120/200/, the difference between t2 and t1 is 80 ms, of which value we know that TS is at maximum. The upper boundary pattern, /170/150/, is still a member of the 1:1 category because it is regularized by bilateral assimilation, as explained above. We applied these two "rules," (t2 - t1) = 80 ms for the lower boundary, and (t1 - t2) = 20 ms for the upper boundary, to four faster and four slower patterns with total durations of 160, 200, 240, 280, 360, 400,

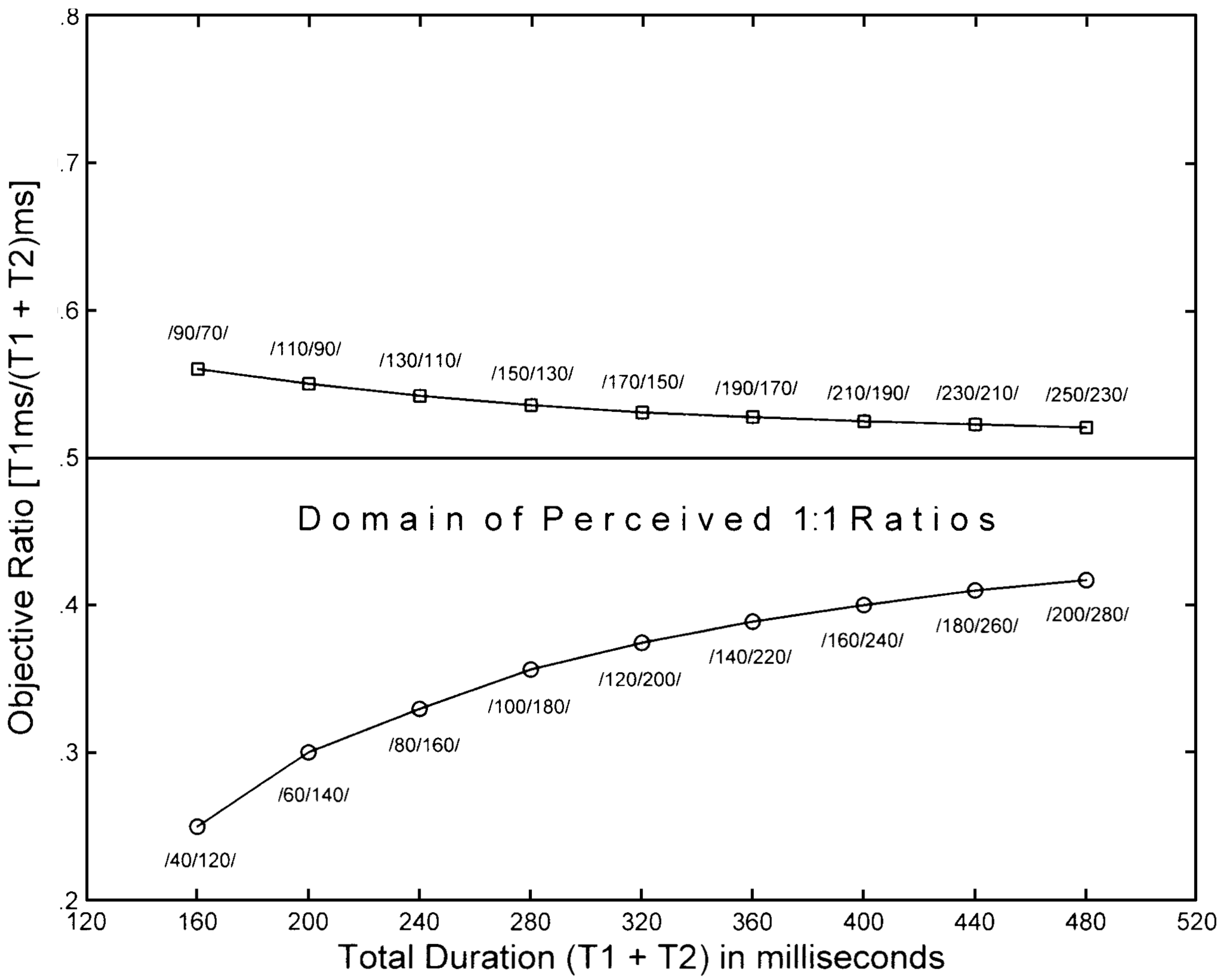

FIG. 13. Estimated domain of perceived 1:1 ratios as a function of the total duration of /t1/t2/ sound patterns and the objective ratio between t1 and t2 expressed as t1/(t1 + t2).

440, and 480 ms. We limited the range of total durations between 160 and 480 ms for two reasons. Two-interval patterns shorter than 160 ms are not realistic in music, and in patterns longer than 480 ms TS vanishes rapidly.

Figure 13 portrays the lower and upper 1:1 category boundaries for all nine total durations, and by connecting these lower and upper boundaries one gets an estimate of the domain of perceived 1:1 ratios. It can be seen that 1) a percussionist has a freedom of 50 ms in producing t1 (and t2 accordingly) to provide the listener with a 1:1 percept, and 2) this t1-range of freedom is strongly asymmetric with respect to the objective 1:1 ratio (the 0.5 reference line). One might argue that TS is never so strong as to assimilate t2 completely to t1 and that the pattern therefore cannot be perceived as 1:1. However, also here we have to take a DL into account. Take for instance the pattern /120/200/. The approximate amount of shrinking of t2, as estimated by our previous studies, is 60 ms, which leaves the listener with an effective /120/140/ pattern. That the /120/140/ pattern is perceived isochronously is caused by sensitivity problems. Ten Hoopen et al. (1995) reported that the DL between t1 and t2 in this kind of patterns is about 15 ms. Given that the DL is a 50 % criterion, it is quite conceivable that /120/140/ is perceived as 1:1. One might object that the process of TS also diminishes the DL, but that is not the case. In ten Hoopen, Beumer, and Nakajima (1996) we showed that TS only affected the point of subjective equality (PSE) of the last interval, but not the DL.

FURTHER RELEVANCE TO RHYTHM PERCEPTION AND PERFORMANCE

Gabrielsson (1974) reported that when musicians (two pianists and one percussionist) performed a number of notated rhythms, several characteristic deviations from the musical notation appeared. One conspicuous deviation occurred when the pianists had to play two eight-notes followed by two quarter-notes. They tended to make the first eight-note physically shorter and the second one physically longer (example 7 in Gabrielsson's Table 2, p. 65). The durations of the two eight-notes were 223 and 273 ms, instead of the notated 250 and 250 ms. In a /223/273/ pattern, TS can easily occur because t2 - t1 $<$ 80 ms, and t1 is not much longer than 200 ms, as a result of which listeners (also the performer her/himself) perceive isochrony, though the physical pattern is anisochronous.

A rhythmic deviation comparable to the above example was reported by Stobart and Cross (2000) with ethnic music. They recorded Bolivian ethnic music, the so-called 'Easter songs,' and analyzed the timing of their individual notes. The authors mentioned the possibility that TS affected the perception of those songs: "The duration in Easter songs lie close to the upper threshold (c. 200 ms) of those employed in the experiments demonstrating time shrinking. Nakajima et al. suggest that below this threshold time shrinking is more-or-less unavoidable in perception, becoming less so as the threshold is approached, which might help explain the rhythmic ambiguity encountered by many listeners to Easter songs" (Stobart & Cross, 2000, p. 84). A typical temporal pattern of a three-note rhythmic group the authors recorded was /187/264/ ms, where the difference between the two successive time intervals was about 80 ms, a difference where TS operates maximally, and the iambic duration relation was perceived as subjectively equal. "But, whilst time shrinking may help us to understand how paired iambic durations might be employed to produce a subjective equality of duration in Easter song performance, it does not explain the differences in rhythmic perceptions encountered between Bolivian and European listeners" (Stobart & Cross, 2000, p. 84). It is indeed interesting that they found a big difference in the perceived rhythmic structures between Bolivian and European listeners. Although Bolivian people perceived this temporal pattern in an Easter song as regular and on-beat, European listeners perceived it as a 1:2 time ratio and as anacrustic. The authors related this difference in the perceived rhythmic structure to the difference in accent placements between the languages. The languages in which Easter songs are sung have a stress pattern such that the first syllable of a phrase is treated as a downbeat, while European languages' stress pattern depend upon the duration relations and other features. It seems that assimilation due to TS appeared when the tune was perceived as on-beat and the rhythm pattern was iambic, whereas contrast worked when the tune was perceived as anacrustic and the grouping of the notes was different.

CONCLUSIONS

It is clear that TS, the process of unilateral assimilation (Nakajima et al., 2004), can produce severe perceptual distortions of simple rhythmic patterns. This perceptual distortion is very robust: Even when the sound markers delimiting t1 and t2 vary widely as regards sound intensity (ten Hoopen et al., 1993) and frequency (Remijn et al., 1999), t2 is significantly and conspicuously underestimated. Hence, even when a typical TS two-interval pattern is embedded in a musical rhythmical pattern (that is, with a melodic contour and dynamic accent) deviations from the score can occur. But the deviation from the score might be even more complicated: recently we found that TS emerged in three-interval patterns (/t1/t2/t3/) in a very intricate way (Sasaki et al., 2002). The last interval t3 could not only be shrunk directly by t2 but, depending on the t2-duration, also directly by t1. Even more interesting was that TS could propagate through the pattern: t1 could shrink t2, and as a result the subjective duration of t2 could become appropriate to shrink t3. The opposite could also happen: t1 could shrink t2, by which t2 got a subjective duration that was inappropriate to shrink t3. There were even /t1/t2/t3/ patterns in which the temporal relations were ideal for TS to occur but Gestalt figural grouping inhibited TS.

Because TS operates only at relatively fast tempi, say note durations shorter than 250 ms, and rapidly vanishes with slower tempi, it can easily break down perceptual rhythm constancy (Handel, 1993). Hence, if a musician has to play such fast rhythmic patterns he/she should adjust the timing of the temporal ratios considerably to produce the perceptual temporal impression as intended by the score. We emphasize that TS is not the only process giving rise to nonveridical ratios and rhythms. Also with tempi far too slow for TS to operate, systematic deviations from objective temporal ratios can occur (Nakajima, 1987; Nakajima et al., 1988) as we already discussed above. Recently, Repp, Windsor, and Desain (2002) reported significant deviations from notated interval ratios in the majority of two- and three-note rhythms played by skilled pianists, even at the slowest tempo they used (500 ms quarter notes). The authors described the deviations in terms of assimilation and contrast and noted that many of them replicated results from Fraisse (1956).

Author Note

We are much indebted to Marco Beumer, Jan Bruijn, and Michel Simons for carrying out pilot experiments and assisting in Experiment 1. We very much appreciated that George Papadelis from Thessaloniki (Greece) visited Leiden University (the Netherlands) to communicate his latest work. We thank the Japan Society for the Promotion of Science (JSPS) for selecting Gert ten Hoopen as a short-term 1995-Fellow, and in collaboration, the Netherlands Organization for Scientific Research (NWO) generously provided the traveling fund to Japan. This research was also supported by a Grant-in-aid of the Ministry of Education of Japan, # 07459017 in the fiscal year 1995, and a Grant-in-aid from the Japan Society for the Promotion of Science (JSPS), # 14101001 in the fiscal years 2002–2004 to Yoshitaka Nakajima. This study was also part of Japan's governmental Center of Excellence (COE) program at the Kyushu University (Fukuoka), called "Design of Artificial Environments on the Basis of Human Sensibility." Parts of Experiment 1 were presented at the Fifth Workshop on Rhythm Perception and Production, September 1994, Sheffield, UK. Requests for reprints from the Pacific Rim to Takayuki Sasaki, Miyagi Gakuin Women's University, 9-1-1 Sakuragaoka, Aoba-ku, 981-8557 Sendai, Japan (sasaki@mgu.ac.jp); otherwise, to Gert ten Hoopen, Leiden University, Cognitive Psychology Unit, P.O. Box 9555, 2300 RB Leiden, the Netherlands (tenhoopen@fsw.leidenuniv.nl).

Address correspondence to: Dr. Gert ten Hoopen, Leiden University, Faculty of Soc. and Behav. Sci., Department of Psychology, P.O. Box 9555, 2300 RB Leiden, The Netherlands. E-MAIL *tenhoopen@fsw.leidenuniv.nl* (office), *hoopenten@planet.nl* (home)

References

Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 215–264). San Diego: Academic 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.

Clarke, E. F. (2000, August). *Categorical rhythm perception and event perception.* Paper presented at the 6th International Conference on Music Perception and Cognition, Keele University, UK.

Davison, M. L. (1992). *Multidimensional scaling.* Malabar, FL: Krieger Publishing Company.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception, 32*, 341–365.

Everitt, B. S. (1993). *Cluster analysis.* London: Edward Arnold.

Fraisse, P. (1956). *Les structures rythmique.* Louvain, Belgium: Publications Universitaires de Louvain.

Fraisse, P. (1978). Time and rhythm perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. VIII, pp. 203–253). New York: Academic Press.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). New York: Academic Press.

Friberg, A., & Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *Journal of the Acoustical Society of America, 98*, 2524–2531.

Gabrielsson, A. (1974). Performance of rhythm patterns. *Scandinavian Journal of Psychology, 15*, 63–72.

Handel, S. (1993). The effect of tempo and tone duration on rhythm discrimination. *Perception & Psychophysics, 54*, 370–382.

Iverson, P., & Kuhl, P. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97*, 553–562.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling.* Beverly Hills, CA: Sage Publications.

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

Large, E. W. (2000, August). *Rhythm categorization in context.* Paper presented at the 6th International Conference on Music Perception and Cognition, Keele University, UK.

Miyauchi, R., & Nakajima, Y. (2005). Bilateral assimilation of two neighboring empty time intervals. *Music Perception, 22*, 411–424.

Nakajima, Y. (1987). A model of empty duration perception. *Perception, 16*, 485–520.

Nakajima, Y., Nishimura, S., & Teranishi, R. (1988). Ratio judgments of empty durations with numeric scales. *Perception, 17*, 93–118.

Nakajima, Y., ten Hoopen, G., Hilkhuysen, G., & Sasaki, T. (1992). Time-shrinking: A discontinuity in the perception of auditory temporal patterns. *Perception & Psychophysics, 51*, 504–507.

Nakajima, Y., ten Hoopen, G., Sasaki, T., Yamamoto, K., Kadota, M., Simons, M., et al. (2004). Time-shrinking: The process of unilateral temporal assimilation. *Perception, 33*, 1061–1079.

NAKAJIMA, Y., TEN HOOPEN, G., & VAN DER WILK, R. (1991). A new illusion of time perception. *Music Perception, 8*, 431–448.

PAPADELIS, G., & PAPANIKOLAOU, G. (2004). Mapping the perceptual space between and within musical rhythm categories. In J. Davison (Ed.), *The music practitioner: Research for the music performer, teacher, and listener* (pp. 117–129). Aldershot, Hampshire, UK: Ashgate Publishers.

PARNCUTT, R. (1994). Categorical perception of short rhythmic events. In A. Friberg, J. Iwarsson, E. Jannson, & J. Sundberg (Eds.), *Proceedings of the Stockholm Music Acoustics Conference 1993* (pp. 47–52). Stockholm: Royal Swedish Academy of Music.

PENEL, A., HOLLWEG, C. A., & BRODY, C. D. (2005, July). *Development of a visuospatial representation method and evidence for rhythmic categorical perception.* Paper presented at the 10th Rhythm Perception and Production Workshop, Bilzen, Belgium.

REMIJN, G., VAN DER MEULEN, G., TEN HOOPEN, G., NAKAJIMA, Y., KOMORI, Y., & SASAKI, T. (1999). On the robustness of time-shrinking. *Journal of the Acoustical Society of Japan (E), 20*, 365–373.

REPP, B. H., & LIBERMAN, A. M. (1987). Phonetic category boundaries are flexible. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 89–112). Cambridge: Cambridge University Press.

REPP, B., WINDSOR, W. L., & DESAIN, P. (2002). Effects of tempo on the timing of simple musical rhythms. *Music Perception, 19*, 565–593.

ROSEN, S., & HOWELL, P. (1987). Auditory, articulatory, and learning explanations of categorical perception in speech. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 113–160). Cambridge: Cambridge University Press.

SASAKI, T., NAKAJIMA, Y., & TEN HOOPEN, G. (1998). Categorical rhythm perception as a result of unilateral assimilation in time-shrinking. *Music Perception, 16*, 201–222.

SASAKI, T., SUETOMI, D., NAKAJIMA, Y., & TEN HOOPEN, G. (2002). Time-shrinking, its propagation, and Gestalt principles. *Perception & Psychophysics, 64*, 919–931.

SCHULZE, H.H. (1989). Categorical perception of rhythmic patterns. *Psychological Research, 51*, 10–15.

STOBART, H., & CROSS, I. (2000). The Andean anacrusis? Rhythmic structure and perception in Easter songs of Northern Potosi, Bolivia. *British Journal of Ethnomusicology, 9*, 63–94. AuQ1

TEN HOOPEN, G. (1992, February). *The processing of fast auditory patterns.* Paper presented at the 2nd International Conference on Music Perception and Cognition, Los Angeles.

TEN HOOPEN, G., BEUMER, M., & NAKAJIMA, Y. (1996). What differs between the first and the last interval of a click sequence simulating a mora structure, the DT or the PSE? A replication of Tanaka, Tsuzaki, and Kato (1994). *Journal of the Acoustical Society of Japan (E), 17*, 155–158.

TEN HOOPEN, G., BOELAARTS, L., GRUISEN, A., APON, I., DONDERS, K., MUL, N., et al. (1994). The detection of anisochrony in monaural and interaural sound sequences. *Perception & Psychophysics, 56*, 110–120.

TEN HOOPEN, G., HARTSUIKER, R., SASAKI, T., NAKAJIMA, Y., TANAKA, M., & TSUMURA, T. (1995). Perception of auditory isochrony: Time-shrinking and temporal patterns. *Perception, 24*, 577–593.

TEN HOOPEN, G., HILKHUYSEN, G., VIS, G., NAKAJIMA, Y., YAMAUCHI, F., & SASAKI, T. (1993). A new illusion of time perception—II. *Music Perception, 11*, 15–38.

VAN DER KLOOT, W. A. (1997). *Meerdimensionale schaaltechnieken voor gelijkenis- en keuzedata* [Multidimensional scaling techniques for similarity and choice data]. Utrecht, the Netherlands: Lemma. AuQ2

WINDSOR, W. L. (1993). Dynamic accents and the categorical perception of metre. *Psychology of Music, 21*, 127–140.

[AuQ1] Is question mark correct here? If so, “rhythmic” should be capitalized.
[AuQ2] Hyphen OK here?