Temporal and spectral basis of the features perceived in repeated noise

Christian Kaernbach
Laboratoire d'Audiologie Expérimentale, INSERM unité 229, Université de Bordeaux II, Hôpital Pellegrin, Place Amélie Raba Léon, 33 076 Bordeaux Cedex, France, and
Institut für Neuroinformatik, Ruhr-Universität Bochum, 44780 Bochum, Germany[a)]

(Received 24 June 1992; revised 22 December 1992; accepted 19 March 1993)

When a half-second segment of a noise is played repeatedly, it initially creates a "whooshing" perception. With longer listening, however, individual features like "clanks" and "rasping" emerge. It is easy to tap the period of the perceived structure. This offers a possibility to investigate the mechanisms underlying the perception of these distinct features. The present study addresses the subject of the temporal and spectral extent of the physical correlates of these percepts. Five subjects participated in this study, and their tapping is in notable, although not perfect correlation. The physical correlate of the features can be confined in time to intervals as small as 100 ms. This segment of the stimulus is processed largely independently from the rest of the noise sample. Spectral processing is, in general, local. Some features, however, are spread over more than one octave. In 3 cases out of 25, across-channel processing is apparent.

PACS numbers: 43.66.Mk, 43.66.Ki [HSC]

INTRODUCTION

Guttman and Julesz (1963) investigated the lower limit of periodicity detection. They found that a random waveform with an infratonal periodicity from 20 cycles per second (cps) down to 1 cps can be effortlessly perceived as periodic. Whereas in the range of 20 to 4 cps, these sounds are heard as "motorboating," they are heard as "whooshing" in the range of 4 to 1 cps. In the whooshing region, distinct perceptive events such as "clanks" and "rasping" emerge with longer listening.

There have been attempts to find salient features in the spectrograms of repeated noises which could be identified as the physical basis of the events heard in repeated noises (Limbert, 1984; unpublished analysis of own data). They failed to identify the physical correlate of the perceived events. Other studies on temporally (Brubaker and Warren, 1987, 1990) or spectrally (Warren and Bashford, 1981; Bashford and Warren, 1990) modified repeated noises did not reveal whether the physical correlate of the perceived events was confined in time and spectrum, or whether a kind of holistic pattern processing takes place.

It is the aim of the present study to determine the temporal and spectral extent of the physical correlate of the "clanks" and "rasping." Two experiments study the temporal (I) and spectral (III) extent of the portion of the repeated noise segment that leads to the perception of the feature. The comparison of the data reveals intersubject differences. These could be due to differences in the perception of the noise sample or in the interpretation of the perceived structure. Experiment II is a supplement to experiment I and tries to resolve this ambiguity.

I. EXPERIMENT I: TEMPORAL EXTENT

Brubaker and Warren (1987) raised the question whether the detection of the periodicity inherent to repeated noise is based solely on the detection of the recurrence of a "singularity" or whether it is based on holistic processing of the entire pattern. They presented subjects with repeated noises made up of three frozen segments of noise (A, B, and C) in two different cyclical orders (ABCABC··· vs ACBACB···). The subjects could readily distinguish these two series. This would have been impossible if the subjects could only detect the recurrence of a single "singularity," for instance in segment A. The detection of several singularities, however, would allow the subjects to distinguish the two orders. The subjects would simply compare the patterns of the recurring components as they do with fast recycled sequences of three or four successive sounds (e.g., high and low tones, buzzes and hisses; cf. Warren and Ackroff, 1976).

The question of Brubaker and Warren was directed toward the *detection* of the periodicity. Their task, however, involved not only the detection of the periodicity but also higher-order processing (comparison) of the perceived periodic structure. The original question "On which parts of the noise sample does the *detection* of periodicity depend?" thus remains unanswered. Experiment II may be seen as an answer to this question.

Experiment I, on the other hand, was designed to answer a slightly different question: "On which parts of the noise sample does the *perception* of the periodically recurring distinct events depend?" Does it make sense to think that the entire period might influence the perception of the features of repeated noise? There are examples where the perceived starting point (and thus emphasis) of a cyclical structure depends on the entire pattern. Imagine a repeating two-tone melody (e.g., AABABBAAAB repeated cy-

---

[a)] Please send correspondence to this address.

clically and rather fast). Preusser (1972) found that the start or the end of the longest run of identical tones (i.e., the triple A, in the example above) would define the perceived starting point which was perceived as emphasized. In this case all of the sequence is important to determine the perceived emphasis.

It is the aim of experiment I to find out whether the perceived events result from such holistic pattern processing or whether they are due to features in the noise sample which can be confined to some temporal interval.

A. Method

The tapping task allows us to study repeated noises in the whooshing range. The repeated noises were presented to the subjects, and the subjects had to tap the perceived period (i.e, once per period). Limbert and Patterson (1982; Limbert 1984) showed that tapping to periodic noises occurs consistently at one point within one presentation. A study by the author (Kaernbach, 1992) has shown that this tapping point may be reproduced with a high probability in a later presentation of the same noise segment. Thus slight modifications to the noise segment may be introduced and tapping to the modified noise sample may be compared to the original tapping point. This allows us to use the tapping task to study the physical correlate of the perception of repeated noise.

Repeated noise was presented to the subject. The presentation ended when the subject has tapped eight times. The next presentation started after a pause of 2 s. The tapping points were interpreted modulo the length τ of the repeated noise segment. If the resultant eight tapping points (all from the interval $[0,\tau]$) had a standard deviation of more than 0.1τ, this trial was discarded. A trial lasted on the average for about 12 s. Onset and offset ramps were cosinusoidal and lasted 20 ms. A random portion of the first cycle was skipped to avoid identical starting points on successive runs.

The repeated noises were digitally generated and converted by a 16-bit converter at a rate of 20 kHz with a low-pass filter at 10 kHz. The spectral power density of the Gaussian noise was 24 dB SPL per Hz. The numerical noise-generating algorithm is described in detail in (Kaernbach, 1992). The noise used throughout this study corresponds to the first six seconds of noise one of this algorithm.

Six seconds of Gaussian noise were prepared and stored.[1] On each trial, a short portion of this noise was cut out and presented repeatedly with transientless juxtapositions.[2] The length of this portion was selected at random from the values 400, 600, or 800 ms. Its position was random within the 6 s. This procedure can be seen as clipping windows out of the 6 s of noise. This allows us to examine the perception of events, such as clanks, as a function of the position and the length of the window. The position of the window was chosen such that no two successive presentations had overlapping windows.

Five subjects participated in this study. All subjects reported normal hearing. They had at least 1 h of training hearing repeated noises. The subjects were seated in a sound-proof booth and the repeated noises were presented diotically via Sennheiser 2002 headphones. The subjects were asked to tap the perceived period. They were told to ignore any apparent resemblance to a former presentation. The subjects were instructed to tap in synchrony with whatever they found to represent the period best. They reported tapping in synchrony with long patterns ("whooshing") occasionally, but in most cases to shorter, nearly punctuate events ("clanks"), as this was easier to synchronize.

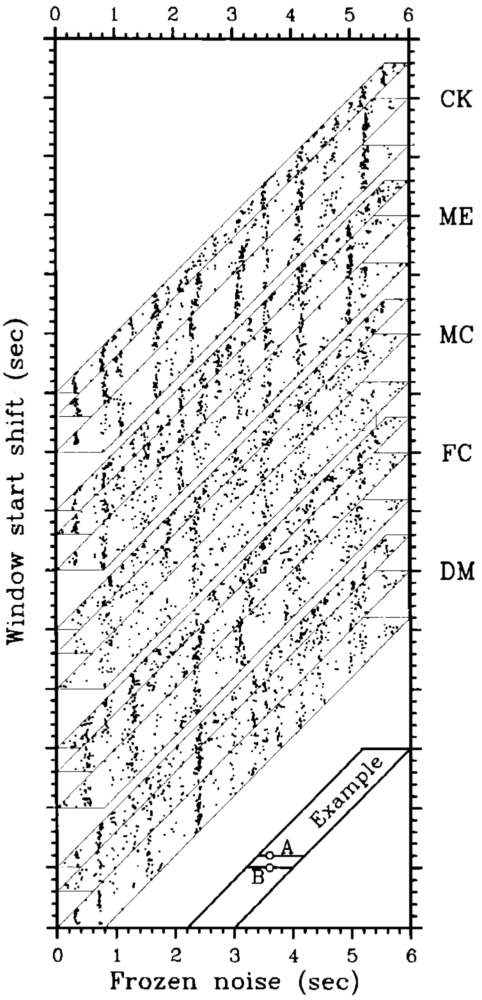

FIG. 1. Dot-plot of the tapped noise instances (horizontal axis) of experiment I. A vertical offset is added corresponding to the starting point of the window. For each subject, three diagonal bars present the tapping points for repeated noises of 400-, 600-, and 800-ms period length. The example bar at lower right is explained in the text. The vertical alignment of the data points shows that the perception of the features only depends on the contiguous presentation of a rather small segment of the noise sample.

B. Results

All subjects performed about 1000 runs. Each run can be described by the position of the window, its length, and by the noise instant indicated by the subjects tapping. Figure 1 shows the results in the form of a diagonal-bar diagram. The horizonal axis represents the 6 s of frozen noise. When the subject points out a certain point in a given noise segment by tapping it, this corresponds to a specific noise instant on this axis. The vertical axis corresponds to the starting point of the window.

To better understand this representation of the data, let us look at the example bar in the lower right-hand corner of Fig. 1. Imagine a run where the 800-ms window from 3.4 to 4.2 was presented. The result would be plotted somewhere on the upper horizontal line A. If in this run, instant 3.6 was tapped by the subject, the resulting data point would lie at the circle on this line. If in a second run, the window was from 3.2 to 4.0, the result would be plotted somewhere on line B. If again instant 3.6 was tapped, this would result in the circle on line B. Vertically aligned points thus denote that a certain noise instant is tapped regardless of its exact temporal context.

The data for each subject are plotted in three bars corresponding to the three possible period lengths. There is a strong tendency toward vertical alignment in the data. This is valid up to the borders of the bars, indicating that the perceived events remain stable for windows containing them just at their beginning or just at their end.[3]

It appears that the temporal context of the feature can be changed to a great extent without affecting its perception. The end points of the vertical lines tell us something about the temporal extent of the physical correlate of the perceived features. Imagine a feature that could be perceived only if the 200 ms before the tapping as well as the 200 ms after the tapping were presented contiguously. The corresponding vertical line would then not touch the borders of the diagonal bar. In Fig. 1 the vertical lines touch the borders, although they are sometimes a little bit bent (cf. 0.8 s for CK, 400- and 600-ms bars). This indicates that the temporal extent of the feature is rather small. In general, the features are limited to 100 ms. The inherent imprecision of the tapping task does not allow us to determine the temporal extent more accurately. The bending of some lines could be due to a certain temporal spread of those features: If a part of the feature is cut away, the center of gravity will move.

The window length seems to have little influence on the perceived events: The lines found in the 800-ms bars may in general be found in the other two bars as well. There are, however, additional lines in the 400-ms bars, which are not seen in the 800-ms bars. This can be understood by the relative dominance of the perceived events. For instance, subject ME perceives two striking events at 4.2 and at 5. There seems to be a minor striking event at 4.5. The 800-ms segments containing this event will, however, contain one of the other two, which are dominant and thus suppress the response to the event at 4.5.

There is remarkable consistency of the lines for different subjects. Several lines can be found for three or more subjects at the same place (0.3, 0.8, 2.4, 3.6, 4.2, 5.2). But there are also remarkable differences between the subjects. For instance, at 5.0 there seems to be nothing special to subject CK, whereas for subject ME this is a very clear-cut and dominant event. This is reminiscent of the results of Kaernbach (1992), where the correlation between subjects was notable, although not perfect. The next section addresses the origin of the intersubject differences.

II. EXPERIMENT II: INTERSUBJECT DIFFERENCES

In the experiments of Preusser (1972), the signal was simply a sequence of tones differing in pitch. There is probably much less variation in the perception of this stimulus than in the perception of repeated noises. The responses to these stimuli were nevertheless variable. This indicates that this variability is associated with a more central process. The intersubject differences of tapping to repeated noise could, however, be due to perceptual differences as well. The following experiment employs a detection task to look for such perception differences. Subjects CK and ME participated in this study.

A. Method

The same 6 s of frozen noise as in experiment I were prepared and stored. At a random position in these six seconds, a 200-ms segment was cut. [That is, a starting point was chosen at random from the interval (0–5.8) at sampling rate resolution] The 200-ms length was chosen to ensure that it holds a characteristic feature. This segment was presented as an only partially frozen repeated noise. To this end the length of a quasiperiod was selected from the values 600,700,...,1000 ms. This quasiperiod was built by starting each time with the 200-ms frozen noise segment and filling it up to the selected length with nonrepeating ("running") noise. Let A be the selected segment. A 600-ms quasiperiod could be represented as ABCADEAFGA··· with the other letters standing for other nonrepeating 200-ms segments of noise (not taken from the 6-s master).

The task of the subject was to detect the periodicity. The noise was presented, and the subject had 12 s to start tapping the perceived period. If the subject started tapping within these 12 s, the presentation was prolonged as necessary. If the subject did not start or if the subject started and the tapped period was not in accordance with the presented quasiperiod, the trial was judged unsuccessful. The criterion for the discard was the same as in experiment I: If the standard deviation of the tapping points (modulo length of the quasiperiod) was more than 0.1τ, the trial was judged unsuccessful. This guaranteed that successful tapping was only possible if the subject "got" the period. The crucial result was the percentage of successful tapping. The tapping point was of no importance.[4] The variations of the length of the quasiperiod served in this experiment simply to disorient the subject regarding the actual periodicity; they are disregarded further on. Both subjects performed about 1000 runs. The probability that a noise instant from the 6000-ms master will be contained in the selected 200-ms segment is $200/6000=1/30$. From this it follows that each noise instant was contained in about $1000/30\approx 33$ presentations (standard deviation $\sqrt{33}\approx 6$).

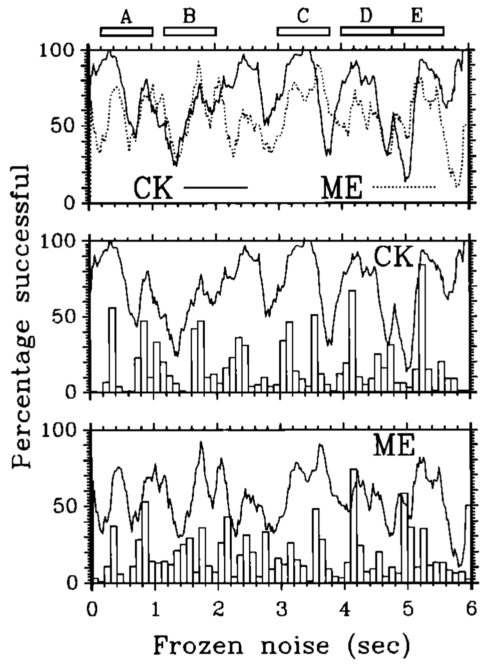

FIG. 2. The upper panel compares the percentage of successful tapping to only partially frozen noise (a 200-ms segment was frozen) for subjects CK and ME. The lower two panels compare, for each subject, this performance with the histogrammed data of experiment I. The *y* axis for the histograms runs from 0 to 100 just as for the percentages. The labeled bars above the panels refer to Fig. 4.

B. Results

Figure 2 shows the results. The percentage of successful tapping is shown over the position of the noise segment. The upper panel compares the data for both subjects, whereas the lower two panels compare, for each subject, the data of this experiment with the data of experiment I.

The success rate was on average better than 50%, but there were large differences depending on the noise segment presented. Whereas some noise segments led to perfect or almost perfect periodicity detection, others seemed hard to detect. This perceptual property of each noise segment is in notable, but not perfect correlation for the two subjects. From 0.5 to 2.1 the results coincide remarkably, even better than could be expected from the sparse statistics. For other parts (2.3 to 3.6, 4.1 to 4.8), the graph of ME is following the graph of CK in parallel at a lower level of performance. This can be understood, as CK is the author and has extreme training in listening to repeated noises. Thus the performance of ME is in general lower or equal to the performance of CK.

There is one remarkable deviation from this trend: At 5.0 the performance of CK goes down to 15%, whereas ME performs successfully in about 45% of the presentations. This corresponds to the missing line at 5.0 in the data of experiment I for CK. This is an example of a clear-cut perceptual difference between the subjects CK and ME. Subject CK did not simply choose to ignore the event at 5.0, preferring 5.2 instead, which would be a possible interpretation of the data from experiment I; he could not perceive this event at all.

The lower two panels show for both subjects the percentage of successful tapping in comparison with the histogrammed tapping points of experiment I. The histogram peaks always correspond to high success rates, but the opposite is not true. High success rates at 3.45 do not lead to tapping this instant in the window experiment: The events at 3.1 and 3.6 seem to suppress the event at 3.4. It is remarkable that the results of the tapping task (experiment I) are compatible with the results of the detection task (experiment II). This is a justification for using the tapping task to study the feature perception.

It is interesting to note an informal observation of the subjects doing this task: It is much harder to detect the periodicity of only partially frozen repeating noise. But once having caught on to the repeating feature, it became clearer and clearer. Even for features hard to detect (i.e., with the subject starting tapping after nearly twelve seconds), the event got so clear that at the end one wondered how one could have missed it before. As the periodicity is partially destroyed, its detection is impaired. This could be seen as an indication that the *detection of periodicity* is based on the entire period. Once having caught the periodicity, however, the features emerge with the same prominence as if the noise were entirely periodic. The *perception of features* thus can be based on only small segments of the noise sample.

III. EXPERIMENT III: SPECTRAL EXTENT

The third experiment was designed to determine the physical correlate of the perceived features in more detail: It determines the spectral range in which the physical equivalent to the perceived feature resides.

Several experiments have studied spectrally modified versions of repeated noise. Guttman and Julesz (1963) found that high- or low-pass filtering of the signal does not degrade the perception. Warren and Bashford (1981) showed that repetition is heard at all center frequencies when a $\frac{1}{3}$-octave filter is swept through broadband repeated noise. Bashford and Warren (1990) showed that bandpass filtered repeated noises can be memorized and can be recognized when presented later embedded in broadband parents. These experiments show that the phenomenon of detection of these long periodicities is relatively impervious to filtering. They do not tell us whether the perception of the distinct features is based on small spectral regions (e.g., a single channel) or on across-channel correlations (e.g., as found in comodulation experiments; for a review see Moore, 1990). The following experiment tries to answer this question.

A. Method

The upper two panels of Fig. 3 explain the way the stimuli were constructed. Imagine the upper-left-hand panel to represent a given frozen noise segment of length τ. Let us choose a cut frequency f and delay the low-pass part in time for a random amount Δt. The resulting sample (upper-right-hand panel) will have shifted all features which were situated in the low-pass portion. Tapping may be described in the high-pass coordinate **t** or in the low-

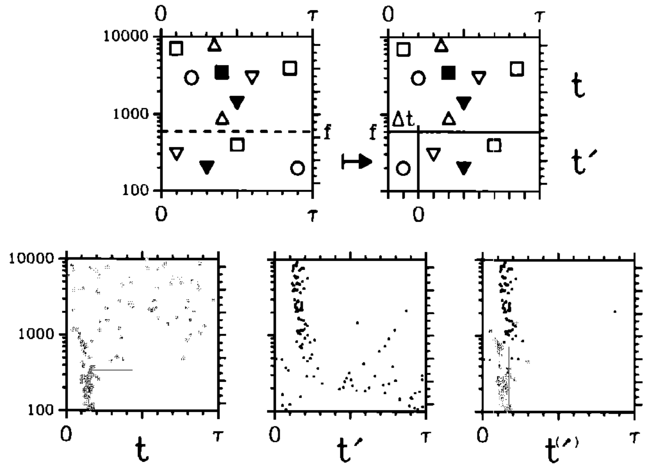

FIG. 3. Stimulus construction and data analysis for the cut-and-delay experiment. The vertical axis shows the frequency of the cut between high-pass and low-pass. The horizontal axis shows the time running from zero to the length τ of the repeated noise segment. The upper two panels explain the construction of the stimulus (see text for further explanation). The first two lower panels show example data plotted with respect to the high-pass coordinate $\mathbf{t}$ and with respect to the low-pass coordinate $\mathbf{t}'$. The third panel combines these into one plot, the selection of $\mathbf{t}$ or $\mathbf{t}'$ as a coordinate being performed as described in the text.

pass coordinate $\mathbf{t}'$, which is corrected for the delay: $\mathbf{t}' = \mathbf{t} - \Delta\mathbf{t}$. If the tapped feature was contained in the high-pass part, subsequent presentations with the same cut frequency, but with different amounts for the delay $\Delta\mathbf{t}$, would reproduce the same tapping point $\mathbf{t}$ regardless of the delay of the low-pass. The low-pass coordinate $\mathbf{t}'$, however, would be randomized. If the feature, on the other hand, was contained in the low pass, $\mathbf{t}'$ would be the better descriptor, and $\mathbf{t}$ would be randomized.

The lower panels for Fig. 3 describe the way the data were plotted. Each run is completely described by three values: The cut frequency $\mathbf{f}$, the high-pass tapping point $\mathbf{t}$, and the delay-correlated low-pass tapping point $\mathbf{t}'$. If we knew which of $\mathbf{t}$ and $\mathbf{t}'$ is the relevant coordinate, we could plot the data on a two-dimensional panel. The left-hand and the middle panel of the lower row present example data with respect to $\mathbf{t}$ and with respect to $\mathbf{t}'$. The cut frequency is shown on the ordinate. For cut frequencies below 500 Hz the tapping point is constant with respect to $\mathbf{t}$, whereas above 1200 Hz $\mathbf{t}'$ seems to be the relevant coordinate. Between these two frequencies, the feature seems to split off; both tapping relative to $\mathbf{t}$ and to $\mathbf{t}'$ is possible. To compare high-pass tapping and low-pass tapping more easily, a two-pass plot was designed. The lower-right-hand panel shows the same data plotted with respect to $\mathbf{t}$ (gray dots) or with respect to $\mathbf{t}'$ (black dots), depending on the concentration of data points in their "neighborhood" in the previous plots.[5] In this kind of plot the transition region and thus the spectrotemporal location of the feature is easy to see.

During the pilot experiments it was noted that it was rather hard to work with a single noise sample: The partial resemblance of successive runs tended to introduce confusion and bias. Therefore five different noise samples were presented in random order. The samples were cut from the six seconds of noise used throughout this study: The positions are indicated by the labeled bars in Fig. 2. A cut frequency was chosen at random between 100 Hz and 10 kHz, and the low-pass was delayed by a random amount. The low- and high-pass filtering was achieved by a fast Fourier transform (FFT). The filter shape was strictly rectangular. The delay operation was performed by multiplying the lines with the appropriate phase-shift factors. The signal was then resynthesized. The segments were 819.2-ms long (2^{14} data points) to ease FFT processing. Seven hundred to 1700 runs were done by the subjects.

B. Results

Figure 4 shows the resulting two-pass plots. Let us first discuss panel ⟨*B*, *FC*⟩. The feature seems striking, as long as the cut frequency is above 2.2 kHz or below 500 Hz. But there is a clear-cut gap between these two points: The feature appears to be destroyed, and the subject chooses another feature instead. This suggests across-channel processing: A cut between the relevant channels will destroy their temporal relation and will make the feature disappear. There are two other panels suggesting across-channel processing: ⟨*D*, *DM*⟩ and ⟨*E*, *ME*⟩. The other panels show more or less broad transition regions where both tapping to the low-pass or to the high-pass occurs. The transition regions may be rather sharp (e.g., within one third octave band in ⟨*D*, *FC*⟩) or spread as much as 1.5 octaves (e.g., ⟨*A*, *ME*⟩). In the latter case we have to assume some integration process, taking into account evidence from a certain spectrotemporal region.

Some panels show a remarkable consistency between subjects: ⟨*D*, *MC* and *FC*⟩, ⟨*E*, *FC* and *DM*⟩, or ⟨*A*, *CK* and *ME* and *DM*⟩ show good coincidence of the spectrotemporal location of the feature. It is interesting to note that sometimes a good temporal coincidence is due to features in completely different spectral regions (e.g., ⟨*C*, *CK* and *ME*⟩). In this case the subjects seem to have listened to different features, which coincidentally were at the same time.

IV. CONCLUSIONS

When the auditory system is presented with repeated white noise, it will enhance details of this noisy structure which we otherwise would not perceive. These details can be localized on the spectrotemporal plane, and they are generally extended over not more than 100 ms. Their spectral extent varies from rather sharp limited signals ($\frac{1}{3}$ octave band) to signals involving several auditory channels.

White noise seems to be filled with a lot of such potential features. In the nonrepeating case, this multitude of possible perceptive events does not have any structure. The physical basis for perceiving such features does exist, so perception should be possible. Nevertheless, the perceptive system seems to recognize that the structureless signal does not convey any information. It seems that it blocks the perception of these uninformative events. The resulting perception is homogeneous. As soon as the feature pattern

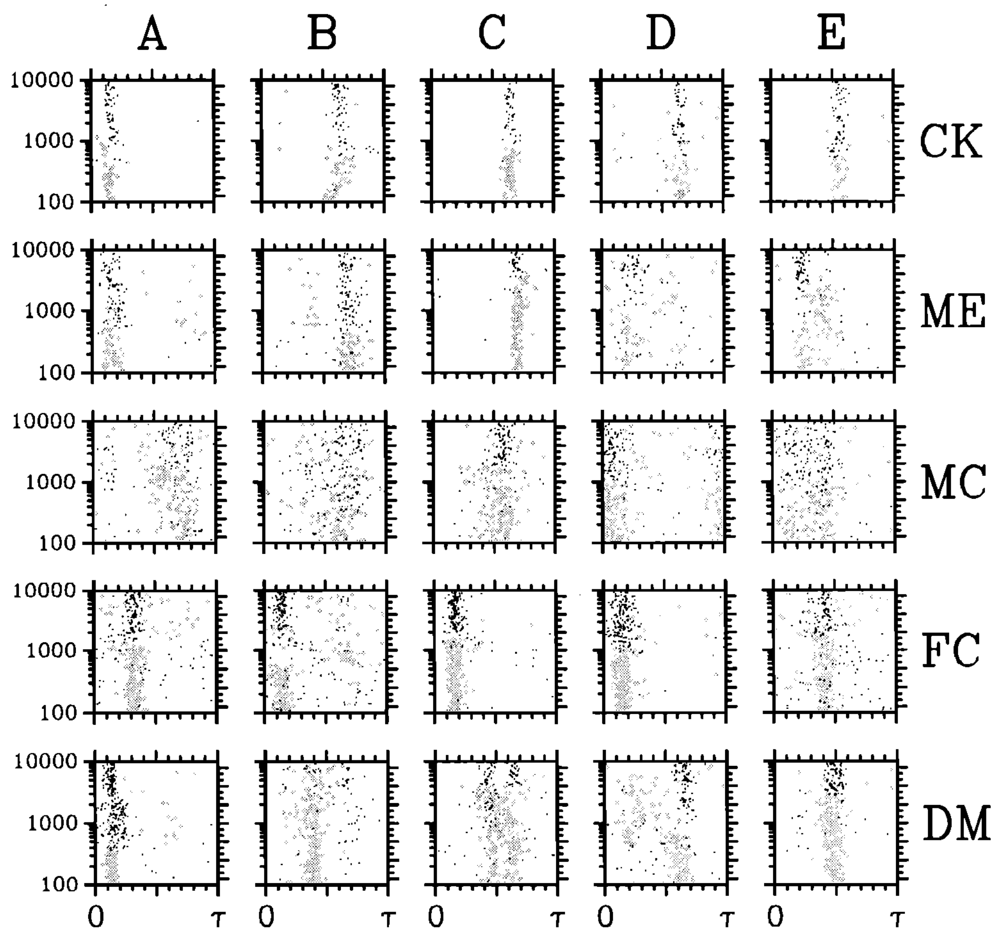

FIG. 4. The cut-and-delay data of all five subjects shown in two-pass plots. The five noise samples A to E correspond to the labeled bars in Fig. 2. The vertical axis represents the cut frequencies. The horizontal axis refers to the high-pass coordinate t (gray dots) or to the low-pass coordinate t′ (black dots). Gray dots thus indicate that the feature was placed in the high-pass, and black dots indicate low-pass features. The transition region is informative for the spectrotemporal extent of the feature.

reappears,[6] the features are taken to be informative, and the corresponding perceptive events emerge.

Guttman and Julesz (1963) proposed that periodicity perception is based on the detection of short-term power-spectrum recurrence. A major difference between the short-term power spectrum and the original waveform is that the phase information is lost. Experiments by Warren and Wrightson (1981) and by Patterson *et al.* (1983) show that repeated noise perception is phase insensitive. So the features could be buried in the short-term power spectrum. Yet Limbert (1984) did not succeed in finding outstanding peaks corresponding to the tapping points. The author of the present study analyzed the short-term spectrum of samples with features constrained in time and spectrum using filters of cochlear bandwidth. He could not find any clear-cut characteristics for such features in the spectrograms.

It appears that the features are not just simple energy concentrations, nor is there only one kind of feature. Features seem to comprise a variety of complicated spectrotemporal structures of the stimulus, some of them involving a single auditory channel, others involving neighboring channels, others spanning well-separated channels. One can explain the fact that listeners sometimes reliably hear the same feature and sometimes, with equal reliability, hear quite different features, by assuming that there are feature detection mechanisms which are marginally excited by the typical white noise source. Depending on the relative sensitivity of these mechanisms, different features can mediate the perception of periodicity.

The perception of repeated noise seems to be based on a learning process. With undisturbed repeating noise, this learning is so fast ("one-shot learning") that one could mistake it for a peripheral detection process. The learning nature of this process becomes clearer when the evidence for a recurrence of the features is reduced as in experiment II: Now, several samples are required before a recurrence is detected. However, once having learned the regularity, the periodic structure becomes evident.

The learning process seems to involve sensory memory. This is not only suggested by the remarkable coincidence of the time limits for sensory memory and for effortless periodicity detection. Two informal observations by the author also strengthen this idea.

(1) Warren and Bashford (1981) reported that trained subjects can detect periodicities as long as ten seconds. The author of the present study is such a subject. It is striking that the events he perceived were nevertheless

contained in approximately 1 s: A 10-s cycle sounded homogeneous for 9 s and structured for 1 s. The detection mechanism seems to concentrate on a particular segment of 1 s length, which will fit into the sensory memory, and tries to detect its recurrence.

(2) In general, n cycles of a noise segment will evoke only $n-1$ repetitions of the perceived events: The first cycle sounds homogeneous. If, however, the subject was trained on this segment, the events will already show up during the first cycle. In this case, there is apparently no need to fill the sensory buffer and to wait for the recurrence of its contents.

Stimuli composed of noise segments with especially dominant features were constructed to try to increase the density of perceived events. This did not work. The features seem to suppress each other, leaving at most 2 or 3 events per cycle. This is reminiscent of the suppression effect found in a previous study (Kaernbach, 1992). On the other hand, it would be interesting to see whether a featureless noise (composed from noise segments with few and weak features) would have special masking properties. Hartmann and Pumplin (1988) showed that noise with reduced fluctuations, so-called low-noise noise, masks less than normal noise. Featureless noise might show similar properties. The difficulty is that one should measure the dominance of the noise segments to select the featureless segments without having the subject learn them all.

The present study concentrated on repeating noises in the range from 1 to 2.5 cps, since the tapping task can be used to isolate features in this case. But there are probably the same mechanisms at work for longer periods (called "trying" by Guttman and Julesz). Periods shorter than 250 ms show strong periodicity cues, but the perceived structure does not resolve into distinct features. Furthermore tapping is less accurate, and systematic phase shifts due to a small period error may occur (Limbert, 1984). It is not clear up to now whether the basic mechanisms for periodicity detection are quite different in that region, or whether the difference is due to postprocessing at a higher cognitive stage.

ACKNOWLEDGMENTS

The author is grateful to Laurent Demany, Yves Cazals, and Egbert de Boer for many discussions on the perception of repeated noise. The author wishes to thank Roy D. Patterson, Bill Hartman, Jan Vorbrüggen, and an anonymous reviewer for helpful comments on the manuscript.

---

[1] The length of the master results from pilot experiments. Informal observations showed that 6-s of noise are sufficient to avoid bias by over-learning. More than 6-s would have meant less statistics for each noise instant.

[2] The transientless juxtapositions are quite easy to achieve with sampled white noise. The sampled values of the random waveform of white noise are just random values, following e.g., a Gaussian distribution. Each value is uncorrelated to its predecessor and successor. The waveform is thus already as discontinuous as possible. This will not get worse by cutting and pasting. Once having cut and reassembled a sample, one cannot distinguish the cuts.

[3] One could argue that the events might have been affected by the window position. Their temporal position might have been preserved, but perhaps their sound had changed. Informal experiments don't support this idea: Sliding the window slowly along the 6-s master will make the events appear, remain stable for a while, and disappear, with quality of sound changing little at the beginning and the end of the appearance.

[4] The tapping points were always in or very near to the repeating 200-ms noise segment. This shows that no off-beat tapping occurs. This can also be seen from Fig. 1: Off-beat tapping would result in interrupted lines, starting somewhere in the middle of the diagonal bar, being displaced at the border by one period length, and ending somewhere in the middle of the diagonal bar.

[5] The concentration of data points in the neighborhood of a certain point was compared algorithmically as follows: The two plots with respect to t and to t′ were generated independently. We then summed for each point an attractive "potential" for all other points, with nearby points making smaller (more negative) contributions. The spectral bandwidth of this potential was about 1 octave. Finally, the summed potentials were compared for the two plots, and the smaller one decided which was the appropriate coordinate for this point.

[6] Limbert (1984) shows that a single repetition of a frozen noise segment is sufficient to detect this repetition.

Bashford, J. A., and Warren, R. M. (**1990**). "Pattern recognition within spectrally isolated regions of broadband complex sound," J. Acoust. Soc. Am. **87**, S24.

Brubaker, B. S., and Warren, R. M. (**1990**). "Auditory memory for long-period random waveforms," J. Acoust. Soc. Am. **87**, S24.

Brubaker, B. S., and Warren, R. M. (**1987**). "Detection of infratonal repetition of frozen noise: Singularity recognition or pattern recognition?" J. Acoust. Soc. Am. **82**, S93.

Guttman, N., and Julesz, B. (**1963**). "Lower limits of auditory periodicity analysis," J. Acoust. Soc. Am. **35**, 610.

Hartman, W. M., and Pumplin, J. (**1988**). "Noise power fluctuations and the masking of sine signals," J. Acoust. Soc. Am. **83**, 2277–2289.

Kaernbach, C. (**1992**). "On the consistency of tapping to repeated noise," J. Acoust. Soc. Am. **92**, 788–793.

Limbert, C. (**1984**). "The perception of repeated noise," Doctoral dissertation, Cambridge University.

Limbert, C., and Patterson, R. D. (**1982**). "Tapping to repeated noise," J. Acoust. Soc. Am. Suppl. 1 **71**, S38.

Moore, B. C. J. (**1990**). "Comodulation masking release: spectro-temporal pattern analysis in hearing," Br. J. Audiol. **24**, 131–137.

Patterson, R. D., Milroy, R., and Lutfi, R. A. (**1983**). "Detecting a repeated tone burst in repeated noise," J. Acoust. Soc. Am. **73**, 951–954.

Preusser, D. (**1972**). "The effect of structure and rate on the recognition and description of auditory temporal patterns," Percept. Psychophys. **11**, 233–240.

Warren, R. M., and Ackroff, J. M. (**1976**). "Two types of auditory sequence perception," Percept. Psychophys. **20**, 387–394.

Warren, R. M., and Bashford, J. A. (**1981**). "Perception of acoustic iterance: Pitch and infrapitch," Percept. Psychophys. **29**, 395–402.

Warren, R. M., and Wrightson, J. M. (**1981**). "Stimuli producing conflicting temporal and spectral cues to frequency," J. Acoust. Soc. Am. **70**, 1020–1024.