On the consistency of tapping to repeated noise

Christian Kaernbach
Laboratoire d'Audiologie Expérimentale, INSERM unité 229, Université de Bordeaux II, 33 076 Bordeaux Cedex, France and Institut für Neuroinformatik, Ruhr-Universität Bochum, Postfach 10 21 48, 4630 Bochum 1, Germany

(Received 5 June 1991; revised 30 January 1992; accepted 20 April 1992)

Repeated noise at 1–4 cycles per second evokes an effortless heard rhythmic sensation which is often heard as "clanks" and "rasping." Tapping in synchrony with the period of the perceived structure is easy and consistent within one presentation. The present study addresses the question of whether the tapping to presentations at different times is consistent across presentations and across subjects. Nine listeners from three countries were presented with repeated Gaussian noise samples in 300 separate cyclical presentations. Nine samples of Gaussian noise with sample lengths ranging from 500 to 700 ms were used. In each of the presentations, one of these samples was selected at random and presented cyclically with transientless juxtapositions. The listeners were instructed to tap in synchrony with the perceived structure (i.e., once per period). Tapping to later presentations of a given sample was found to be consistent with prior tapping to the same sample: In most cases, one or two different tapping points per noise sample could be reproduced in different presentations. In the case of two possible tapping points in different presentations, the two points are usually far away from each other (most likely half a period away). The correlation between subjects is noticeable, although not perfect. The correlation between subjects of the same country is not significantly higher. The noise generating algorithm is given explicitly to allow subsequent studies to use exactly the same noises.

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

INTRODUCTION

In 1963, Guttman and Julesz reported that an iterated, uninterrupted section of a random waveform (Gaussian noise) with section lengths to about 1 s can easily be recognized as repeating itself. They described the perception of these repeated noises (RN) as "whooshing" for period lengths between 250 and 1000 ms. The more one listens, the perceived structure divides itself into distinct events such as "clanks" and "rasping." It thus becomes easy for the listener to tap rhythmically to the perceived periodicity.

The tapping to RN was studied by Limbert and Patterson (1982; Limbert, 1984). The authors studied the consistency of the tapping during one long presentation. They presented four subjects with 16 different RN stimuli, each between 250- and 2000-ms length. Each time one of these 16 waveforms was selected and presented cyclically with transientless juxtapositions. They asked the subjects to tap 100 times to the perceived periodicity, instructing the subjects to always tap at the same point of the period. They found that during each presentation, consistent tapping occurred at one specific point of the noise sample. For short samples (250 ms), the variability was higher while tapping to repeated noise than while tapping to click trains. For long samples (2000 ms) however, the subjects were even more precise while tapping to repeated noises. For medium sample lengths, the precision of tapping was the same for repeated noises and click trains.

The subjects had to repeat each condition four times. These sparse data did not allow for a comprehensive comparison of the tapping points over replications and across listeners. While some of their noise samples showed a certain measure of consistency, others did not.

It is the aim of the present study to enlarge statistics on this point and to examine how far there is a correlation between the tapping points in different presentations and of different subjects. A certain consistency of the tapping points over replications and across listeners is a precondition to study the physical features evoking the perceived events. The present study uses only eight taps per presentation. Eight taps are quite sufficient to determine the tapping point. The small number of taps per presentation also allows for more presentations per subject. Thus each noise sample could be presented about 30 times per subject. Instead of four subjects, the study uses nine to achieve more reliable conclusions.

The subjects were taken from three different countries to investigate the cultural influence on RN perception. The cultural influence of different countries, especially the fact of being exposed to different phonems and different syntactical units, causes differences in the detection and recognition of complex sound features as well as in their perceptual organization (Cutler *et al.*, 1983; Cutler, 1991). Fowler (1979) and Marcus (1981) investigated the perceptual centers of perceptually regular sequences of speech sounds. Digit sequences were presented to listeners who could adjust the interdigit intervals until the digit sequence was perceptually isochronous. The deviations from a purely acoustical isochrony were systematic and similar for all listeners. These deviations were at least partial due to knowledge about speech production (assumed articulatory movements to produce that word). The knowledge about speech production (and thus its influence on the perception of an emphasis) should differ for listeners with different linguistic backgrounds. These differences could then influence the perception and organization of RN stimuli.

I. METHOD

A. Subjects

Nine subjects were chosen out of three countries (D: Germany; F: France; C: China). Their ages ranged from 24 to 40. All subjects reported normal hearing. In each nationality group, there were two male subjects (numbered 1 and 2) and one female subject (numbered 3). Subjects D1, F1, and F2 had experience in psychophysical experiments. All subjects had lived in France for at least half a year at the time the study was conducted. In addition to their native language, all subjects also spoke French and English. All three of the Chinese subjects came from Peking and spoke the same Chinese dialect. Each subject did 100 runs with other RN stimuli to practice precise tapping. Only one subject (D1, the author) was already familiar with RN stimuli before the study.

B. Stimuli

Each run presented one out of nine samples of Gaussian noise as RN (i.e., cyclically iterated with transientless juxtapositions). A random portion of the first cycle was skipped to avoid identical starting points in subsequent presentations. The nine noise samples were digitally generated using an algorithm described in Appendix A. N1 to N9 were generated using index 1 to 9 of this algorithm, and the sample length was 500 ms for N1 to N3, 600 ms for N4 to N6, and 700 ms for N7 to N9. The resulting noise (e.g., RN1 = N1N1N1...) was converted by a 16-bit converter at a rate of 20 kHz. The standard deviation of the Gaussian noise was 10% of the convertible range. The resulting spectral power density was 24 dB SPL per Hz.

C. Procedure

The subjects were seated in a sound-proof booth. The RN stimuli were presented diotically via Sennheiser 2002 headphones. The subjects were asked to listen a few seconds to each RN stimulus before starting tapping. They then had to tap in synchrony with the perceived periodicity. They should tap once per period at whatever they perceived as emphasis. In case of ambiguity (two striking events, which could both be perceived as emphasis) they should choose the one that they perceived as being better defined in time (e.g., rather the clank than the rasping). The nine different possible RN runs were done in random order, although we prohibited the succession of two identical RN stimuli. Each presentation (run) started automatically 2 s after the preceding one. It ended when the subject had tapped eight times. Preliminary studies on tapping to click trains had shown that this is quite efficient in determining the tapping point of a run. Onset and offset ramps were cosinusoidal and lasted 20 ms. Pilot experiments had shown that it made no difference whether the runs were done the same day or two months later. Nevertheless, the runs followed a fixed schedule: 300 runs in three sessions of 100 runs each at three different days within one week. The 300 runs were chosen at random from the nine samples, so each sample was presented to each subject about 33 ± 6 times.

II. RESULTS

A. Tapping point histograms

Mean and variance of cyclical data are most easily determined in the plane of the complex numbers. The n tapping points t_i corresponding to the n taps of a run (in this study, $n=8$) were interpreted as n complex unit vectors $v_i = \exp(2\pi i\, t_i/\tau)$ with τ being the period length of the sample. The phase Φ of the sum vector $V = \Sigma v_i$ reveals the average tapping point. The length $l = |V|$ of the sum vector is maximal ($l = n$) for perfect in-phase tapping and is expected to be about $\sqrt{n}$ for random tapping. For nearly perfect tapping ($2\pi t_i/\tau = \Phi \pm \sigma$ with σ being small), the expected length is given by

$$l = n\cos(\sigma) \approx n(1-\sigma^2/2). \quad (1a)$$

Equation (1a) can be inverted to get the standard deviation:

$$\sigma \approx (2-2l/n)^{1/2} \quad \text{for } \sigma \text{ being small.} \quad (1b)$$

If the subject had erroneously touched the tapping key without the intention of tapping, there would be at least one tap out of the usual rhythm. Consequently, σ would be high. Here, 6% of the runs had $\sigma > 0.2\pi$ and were excluded from the data. Figure 1 presents histograms for each subject and noise sample. The abscissa shows the tapping points t_i (in bins of 25 ms), and the ordinate shows the frequency of this tapping point to be tapped. One or two preferred tapping

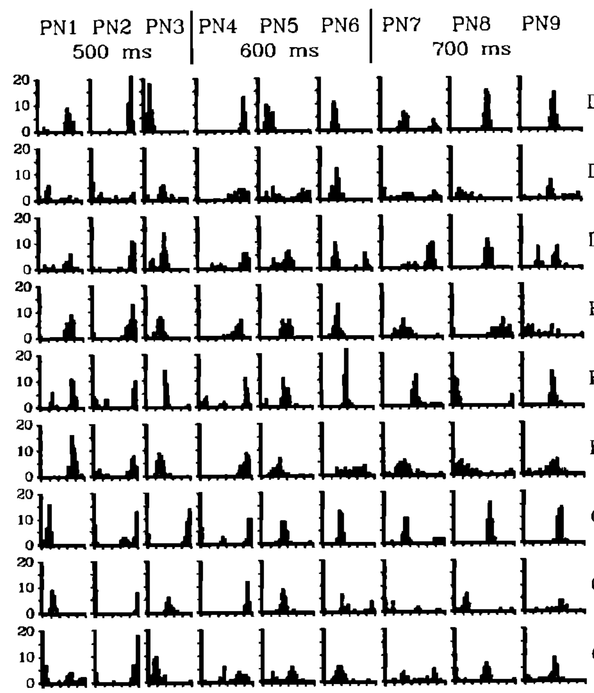

FIG. 1. The tapping point histograms for nine different repeated noises and nine subjects. The histograms show the frequency (ordinate) of a certain tapping point (abscissa, running from zero to 2π, in bins of 25 ms) of a sample to be perceived as emphasis. One or two preferred tapping points could usually be reproduced.

TABLE I. Variance ratios and correlation averaged over all samples. The first row shows the average variance ratios for each subject. The second row shows the weighted version of the variance ratio. The third row shows the shifts that were applied to all data of one subject in order to account for the differences in the anticipation times. Then follow the average correlation coefficients for each pair of subjects. The average over all 36 coefficients is 0.43. Intranation coefficients are bold, coefficients between female subjects are italic.

Subject		D1	D2	D3	F1	F2	F3	C1	C2	C3	Average
Variance ratio		8.1	2.4	4.7	4.1	7.5	3.7	7.7	4.4	3.6	5.1 ± 2
Weighted		18.2	5.2	10.5	11.9	15.1	10.4	17.5	8.8	8.2	11.8 ± 4.1
Shift (bins)		1	0	0	2	−1	1	0	−1	2	
Correlations:											
	D2	**0.36**									
	D3	**0.56**	**0.26**								
	F1	0.38	0.25	0.47							
	F2	0.37	0.43	0.47	**0.63**						
	F3	0.50	0.44	*0.38*	**0.49**	**0.60**					
	C1	0.53	0.32	0.43	0.20	0.26	0.29				
	C2	0.30	0.69	0.41	0.33	0.53	0.50	**0.40**			
	C3	0.60	0.40	*0.59*	0.51	0.37	*0.40*	**0.49**	**0.37**		

points could usually be reproduced by the subjects. A comparison of the histograms of all subjects for one noise sample shows in general a remarkable, although not perfect, correlation. However, visual comparison by itself is not very significant. More can be learned from the correlation coefficients which we will derive in the next sections.

B. Autocorrelation

Let k_j ($j = 1,...,m$) be the number of entries in one out of the m bins of a histrogram ($m = 20, 24,$ or 28). The expected value $E(k) = \Sigma k_j/m$ is equal to K/m, where $K = \Sigma k_j$ is the total number of all entries in this histrogram. Let us define the measured probability density p_j to be

$$p_j = k_j m/K. \tag{2}$$

The expected value $E(p)$ is equal to 1. The variance V_p is then given by

$$V_p = E(p^2) - E^2(p) = E(p^2) - 1. \tag{3}$$

Random tapping would lead to an expected variance of $V_{p_r} \approx m/K$. However, if the histogram shows a concentration of tapping points, the observed variance will be higher. The variance ratio $V_p/V_{p_r} = V_p K/m$ of observed to random variance tells, to which degree the subject was capable to restrict his responses to a few bins. Table I shows in its first row for each subject the average variance ratio over all nine samples. Especially subjects D1, F2, and C1 show high variance ratios.

Let p_j be the measured probability density of one of the histograms. Let p'_j be the probability density of the same histogram cyclically shifted for t bins. The autocovariance $V_{pp}(t)$ and the autocorrelation coefficient $a(t)$ are then given by

$$V_{pp}(t) = E(pp') - 1,$$

$$a(t) = V_{pp}(t)/(V_p V_{p'})^{1/2} = V_{pp}(t)/V_p. \tag{4}$$

The average autocorrelation over the entire cycle is zero. Figure 2 shows the autocorrelation averaged over all samples of equal length as a function of the time shift t. The three panels stand for different sample lengths. The thick histogram bars show the average of all subjects, whereas the thin bars show the average of the three best subjects (D1, F2, and C1). The high precision of retapping a certain tapping point in a later presentation is reflected by the fact that neighboring bins are well correlated.

The three best subjects show a sharper autocorrelation diagram than the average. In this way a second-order phenomenon becomes more evident: If the subjects do not tap their preferred tapping point, they are likely to tap half a cycle away from their normally preferred point. This behavior was previously hypothesized by Limbert (1984), although he had not sufficient statistics to make it evident. The

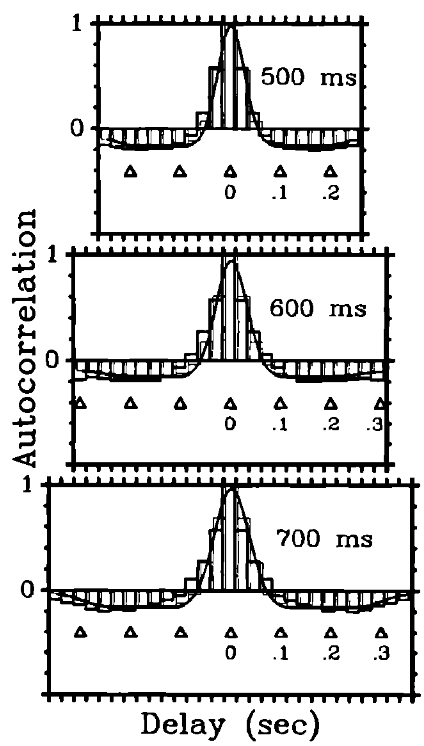

FIG. 2. The autocorrelation (ordinate) as a function of the time delay (abscissa, running from $-\pi$ to π), averaged over all samples of equal length. The thick histogram bars show the average of all nine subjects, whereas the thin ones are averaged only about the most consistent subjects D1, F2, and C1. Neighboring bins are well correlated. The fitted curve refers to the thin histogram bars. There exists a second peak half a period away (i.e., at $-\pi \equiv \pi$), indicating a type of rhythmic enhancement.

solid lines represent curves fitted to the thin histogram bars. The curves are built from two Gaussian distributions, one at zero and the other half a period away. The second peak appears halved in Fig. 2, since this is exactly the point where the cyclical histogram was cut. The longer the sample, the more prominent the second peak: for 700 ms (lowest panel), its height is already more than 10% of the height of the main peak. One could think of explaining this phenomenon by mutual suppression of neighboring features. The subject would be kept from tapping near a strong feature. He would tap either the strong feature or another feature far away. The autocorrelation should then show valleys to the left and to the right of the main peak and reach an asymptotic value for points far away from the main peak. For short periods the broad valley centered at zero could resemble a peak at π. The second peak should nevertheless be much broader than the main peak. In reality, however, the second peak is as narrow as the main peak. Therefore, a sort of "rhythmic enhancement" should be considered: perceived features at a certain point of the noise sample also focus the attention of the subject to the opposite point of the cycle.

C. Weighted correlation

The variances calculated above do not consider neighboring responses as similar. Two entries at, say, bins 8 and 9 have as little in common as two entries at bins 8 and 17. An appropriate weighting will link entries in neighboring bins and thus improve correlation analysis. Let us define the weighted probability density estimate x_j to be

$$x_j = \sum_{d=-3}^{3} w_d (k_{j+d}) m/K, \quad \text{with} \quad \sum_{d=-3}^{3} w_d = 1. \quad (5)$$

Here, d is the distance between two bins, and w_d is the weight that links two entries which have the distance d. The expected value $E(x)$ is again equal to one. The weights used in this study were 0.25, 0.20, 0.125, and 0.05 for $|d| = 0, 1, 2, 3$ (see Fig. 3). These weights were chosen in view of the main peak of the autocorrelation functions. The weighted estimate x is thus a smoothened version of the original estimate p: It is void of all narrow-spaced random fluctuations. In this way, features in the histogram corresponding to the weight function (Fig. 3) will show up more clearly. The variance V_x is given by

$$V_x = E(x^2) - 1. \quad (6)$$

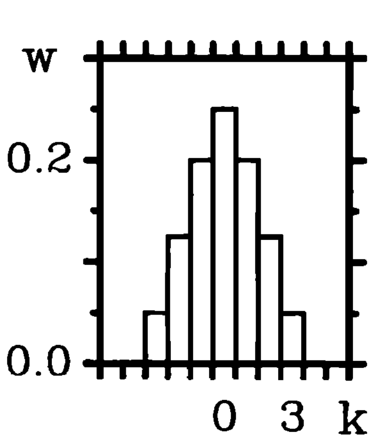

FIG. 3. The weights used in the weighted correlation analysis.

TABLE II. The weighted variance ratios, averaged over subjects.

Length (ms)	500			600			700		
Sample	RN1	RN2	RN3	RN4	RN5	RN6	RN7	RN8	RN9
Variance ratio	10.6	12.9	13.5	10.2	10.8	13.7	8.1	14.4	11.6
Group average		12.3			11.6			11.4	
Total average					11.8 ± 1.9				

The expected variance of a histogram resulting from random tapping is much smaller for x than for p, as much of the random fluctuations will be smoothened out: $V_{x_r} = V_{p_r} \Sigma w^2 \approx 0.18 V_{p_r}$. However, fluctuations in the histograms due to consistent tapping will not smoothen out. The weighted-variance ratio $V_x / V_{x_r} = V_x K/m/\Sigma w^2$ is shown in the second row of Table I. It is much higher than the unweighted one. Ratios up to 18 reflect the subjects capability to restrict their tapping to a few tapping points. Table II shows the weighted-variance ratios averaged over subjects instead of over samples. This average indicates for each sample the degree of "ease" the subjects had to reproduce the same tapping point. There exists a slight tendency of decreasing variance ratios for longer samples. This tendency should not be given too much importance since the third group (700 ms) contains the highest and the lowest entry as well. The standard deviation of the average values is much lower for the values averaged over samples (Table II: $\sigma = 1.9$) than for the values averaged over subjects (second row of Table I: $\sigma = 4.1$). Consequently, the variability in the present study was due more to the subjects than to the samples. The subject variability could be reduced by training and the selection of "good" subjects with high variance ratios.

Let y_j be the weighted probability density of a second histogram. The covariance V_{xy} and the correlation coefficient r are given by

$$V_{xy} = E(xy) - 1, \quad r = V_{xy}/(V_x V_y)^{1/2}. \quad (7)$$

Rhythmic tapping generally occurs in anticipation to the waveform features causing the perceived events. The amounts of anticipation differ from subject to subject (for a review, see Schmidt, 1968). This becomes clear when we study the subjects F1 and F2: Although F2 taps in good correlation with F1, he always taps just a little bit later. To allow for a correction of these differences, small shifts of one or two bins were assigned to each subject. These were then applied to all histograms of this subject equally. The exact amount of the shift (third row of Table I) was adjusted to maximize the overall correlation. The last part of Table I shows the average correlation coefficients over all nine samples for each pair of subjects.

D. Dependence on culture and sex

The average of all correlation coefficients of Table I is 0.43, with the single coefficients ranging from 0.2 to 0.7. The average correlation between subjects of the same country (0.463) is slightly higher than the overall average. This is

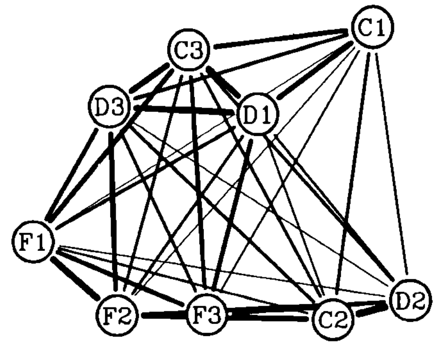

FIG. 4. Graphical representation of the relations between the subjects. A good average correlation of the preferred tapping points between two subjects is expressed as proximity of the two-dimensional positions and with a heavy connecting line. There is no evident cultural grouping.

due to the French subgroup, which with 0.575 has a relatively high intranation correlation. The female subjects show an average correlation of 0.456, which is only slightly higher than the overall average. Figure 4 shows a graphical representation of the relations between the subjects. The positions were chosen so that the distances correspond as well as possible to $-\log(r)$ (i.e., the better the correlation, the closer the positions). Furthermore, the line width was chosen accordingly to the correlation. Figure 4 illustrates that the dependence on cultural origin or sex, if it exists at all, is negligible. The intranation differences of RN perception are comparable to the internation differences. The high intranation correlation of the French subgroup may be due to chance. But it should be taken into account that the study suffered from the fact that all subjects lived in France for at least one year. If the study had been conducted in three different countries, the differences might have been more obvious.

III. CONCLUSIONS

The tapping to RN stimuli is not only consistent within one single presentation (Limbert and Patterson, 1982), it is also reproducible in further presentations of the same stimulus. Some subjects show a higher degree of consistency than other subjects who show a lower (but still remarkable) consistency. This seems to be correlated to musical and/or psychophysical training. Tapping is thus a valid approach to study the perception of repeated noise if the subjects are preselected for consistent tapping—ambiguous samples might produce two preferred tapping points. In this case, the second tapping point is most likely half a period away. This suggests a sort of "rhythmic enhancement" of features half a period away from the main feature. For instance, if a sample contained the features a, b, c, and d in approximately equal distances, with a being the most important feature (i.e., ...d **a** b c d **a** b...), it is quite likely that the subject would report *c* to be the second best feature.

Furthermore, there is a noticeable correlation[1] across subjects. This indicates that there exits a common base of periodicity detection. Thus it is possible to address the physical features causing the perceptual events.

What are the stimulus features providing the base for the perceived emphasis? Limbert (1984) looked for peaks in the short-term power spectrum of the noise samples corresponding to the tapping times. There were no significant peaks found at the corresponding tapping times.

The phase of the noise samples does not seem to play an important role. Warren and Wrightson (1981), Warren (1982), and Patterson *et al.* (1983) examined the perceptual effect of the phase shifts. In subsequent periods, the phases of all components of the noise sample were inversed or shifted by some phase angle. For the sample lengths relevant to the present study, this did not alter perception.

Brubaker and Warren (1987) addressed the question of whether the detection of infratonal periodicity is based on the detection of a reoccurring singularity or on a holistic processing of the entire pattern. Their results indicate that at least recognition involves holistic pattern processing. Preusser *et al.* (1970) and Preusser (1972) studied the perceptual organization of repeated auditory patterns. In short repeating sequences of few tones (e.g., repeat cdddccd) the emphasis would most likely be perceived at the beginning of the longest run of identical tones (i.e., c**ddd**ccd). This procedure is a type of holistic pattern processing meant to find the emphasis of the period. The stimuli are, however, not very similar to repeated noises.

The physical nature of the features evoking the perceived events seems to be much more complicated. Further studies on this topic will be described in a future paper.

ACKNOWLEDGMENTS

Egbert de Boer, Yves Cazals, and Laurent Demany helped the author with many fruitful discussions to formulate the goals of this study. The author is grateful to Roy Patterson, Richard Warren, and Brad Brubaker for helpful comments to the manuscript. Jean-François Rouanet constructed an efficient tapping key of ingenious simplicity.

APPENDIX: GAUSSIAN NOISE INDEXING SYSTEM

The Gaussian noise samples used in this study were digitally generated using a simple, reproducible setup. One integer, the "index," suffices to identify a member of a random number generator family, which generated this noise. Using the same amplitude, sampling frequency, and sample length, the same noise can be reproduced. The algorithm is provided in this Appendix to allow further studies to use exactly the same noise samples.

Most random number generators produce a uniform distribution of values from [0,1]. To produce Gaussian noise, these values have to be transformed to Gaussian-distributed values. This can be done with the Box–Muller transformation. Given two independent random numbers X and Y, uniformly distributed in [0,1], one gets two independent normal distributed random numbers U and V (mean 0, standard deviation 1) as follows:

$$U=[-2\ln(X)]\cos(2\pi Y), \quad V=[-2\ln(X)]\sin(2\pi Y). \tag{A1}$$

The commonly used type of random number generator is the linear congruential generator. Given the modulus M, the multiplier A, a constant C, and the starting value K_0, further values of K follow from

$$K_{n+1}=(AK_n+C)\bmod M. \tag{A2}$$

TABLE AI. The values of the random series generated by Eqs. (A1) to (A4) for verification reasons. Finally presented was U_1, V_1, U_2, V_2, ... after multiplying with the intended standard deviation. The latter was 10% of the highest convertible value ($2^{15}-1$), so excessions of the conversion range could hardly ever occur. Really converted were thus 211, − 3725, 5346, − 1363, ... for the first series.

i	n	K_n	L_n	U_n	V_n
1	1	280718689	407496342	0.06451	− 1.13695
	2	130062562	515535454	1.63167	− 0.41610
	500	351784289	413111291	0.11226	− 0.91262
2	1	503673792	179541249	− 0.16618	0.28335
	2	527791743	452322272	0.10142	− 0.15436
	500	184200468	210021800	− 1.13400	0.92386

If the constants are chosen appropriately, this will give random numbers uniformly distributed with $0 \leqslant K < M$. These can be transformed to the interval [0,1] by division by M.

The correct choice of the modulus M, the multiplier A, and the constant C is all but trivial. Knuth (1969) discusses the advantages and possible problems of a lot of possible choices. The random generators proposed here are constructed closely following his suggestions. Regarding the computing speed it would be best to choose $M = 2^b$, where b is the number of bits per computer word. The modulus operation becomes then rather simple. Regarding the reliability though, it is better to choose M to be prime. Any multiplier will then give the maximum period, the right-hand digits of the resulting numbers are as random as the left-hand digits, and one is no longer concerned about the potency of the random sequence (a measure of the independence of successive numbers), which is infinite. This choice is preferable since computer power seems sufficient today. The multiplier A should be larger than $\sqrt{M}$, preferably larger than $M/100$, but smaller than $M - \sqrt{M}$. The constant C should be chosen so that C/M is approximately equal to $1/2 - \sqrt{3}/6$, since then the serial correlation will be very low. All calculations must be done exactly, with no roundoff error. If these principles are followed, the generator will produce sufficiently independent random numbers. The following section will deal with the definition of a family of independent generators and its implementation with common high level languages.

To produce completely independent noise samples, two independent random number generators are defined per sample:

$$K_{n+1} = (A_k K_n + C_k) \bmod M_k,$$
$$L_{n+1} = (A_l L_n + C_l) \bmod M_l, \qquad \text{(A3)}$$

specifiying K_0, A_k, C_k, and M_k as well as L_0, A_l, C_l, and M_l. To guarantee the independence of the series, the moduli M_k and M_l should be different. They should be as large as possible to get a long period. But it should be kept in mind that AM (following the suggestions for A this is about $M^2/100$) should be calculable without roundoff error. Many processors manage 64-bit integer arithmetics, but high level languages like FORTRAN or PASCAL usally do not give access to these capabilities. The IEEE standard for double precision floating point arithmetics (e.g., DOUBLE PRECISION in FORTRAN) supports precise integer arithmetics up to 2^{53}. This allows M to be as large as 2^{29} and A as large as 2^{23} (i.e., $\approx M/64$). The moduli were therefore chosen among the primes below 2^{29} from Knuth (1969). To avoid copying errors, the other parameters were chosen to be simple potency expressions. The following parameters result:

$$K_0 = 26^6, \quad A_k = 24^5 - i, \quad C_k = 22^6, \quad M_k = 2^{29} - 33,$$
$$L_0 = K_0, \quad A_l = 24^5 + i, \quad C_l = C_k, \quad M_l = 2^{29} - 43, \qquad \text{(A4)}$$

where i is the index of the member of the generator family. The differences in the multiplier of different generators from that family will make them sufficiently independent. K_n and L_n are then transformed to $X_n = (K_n + 1)/M_k$ and $Y_n = (L_n + 1)/M_l$, and these following Eq. (A1) to U_n and V_n. Each sample consists of the values $U_1, V_1, U_2, V_2, \ldots, U_j, V_j$, where j is half the sample length. Given the index i, the sample length $2j$, the sampling frequency, and the standard deviation, the sample can be reproduced exactly. Table AI gives some values of the first two samples for verification reasons.

[1]One would not expect a perfect correlation even if the perception of RN were identical for all listeners. This is due to the probabilistic nature already inherent in the rhythmic organization of the perceived patterns, since cyclical patterns do not possess an unequivocal starting point. Thus it cannot be decided whether the source of the observed ambiguity is located in the organization process alone or in the preceding perception process as well.

Brubaker, B. S., and Warren, R. M. (1987). "Detection of infratonal repetition of frozen noise: Singularity recognition or pattern recognition?" J. Acoust. Soc. Am. Suppl. 1 **82**, S93.

Cutler, A. (1991). "Linguistic rhythm and speech segmentation," in *Language, Music, Speech and Brain*, edited by Carlson, Nord, and Sundberg (MacMillan, New York).

Cutler, A., Mehler, G., Norris, D., and Segui, J. (1983). "Language-specific comprehension strategy," Nature **304**, 159–160.

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

Fowler, C. A. (1979). " 'Perceptual centers' in speech production and perception," Percept. Psychophys. **25**, 375–388.

Knuth, D. E. (1969). *The Art of Computer Programming. Vol. 2* (Addison-Wesley, Reading, MA).

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.

Marcus, S. M. (1981). "Acoustic determinants of perceptual center (P-center) location," Percept. Psychophys. **30**, 247–256.

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.

Preusser, D., Garner, W. R., and Gottwald, R. L. (1970). "Perceptual organization of two-element temporal patterns as a function of their component one-element patterns," Am. J. Psychol. **83**, 151–170.

Schmidt, R. A. (1968). "Anticipation and timing in human motor performance," Psychol. Bull. 70, 631–646.

Warren, R. M. (1982). *Auditory Perception. A New Synthesis* (Pergamon, Oxford).

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