DETECTION OF AUDITORY PERIODICITY: COMPARING BEHAVIORAL DATA AND A DECISION ALGORITHM BASED ON A NEURAL NET

Carsten Bogler, University of Leipzig, Germany
Email: psy97grp@studserv.uni-leipzig.de

Christian Kaernbach, University of Leipzig, Germany
Email: Christian@Kaernbach.de

Abstract

We present psychophysical data on fast auditory periodicity detection in the millisecond range (temporal pitch) that rule out a simple first-order inter-spike interval model. We then present a neural sequence learner that examines fast spike patterns as they are supposed to occur with periodic auditory signals and measures the degree of their regularity. The output of the neural net is evaluated by a decision algorithm that is able to decide between more and less regular sequences. The data of this virtual decider prove to be compatible with the psychophysical data.

A periodic sound with the fundamental frequency f_0 consists of a series of harmonics f_0, 2 f_0, 3 f_0, etc. or only of some of the harmonics (e.g. in the case of the missing fundamental). There are two complementary mechanisms which accomplish the perception of the periodic sound's pitch:

- Spectral pattern recognition acts on the neural excitation pattern produced in the inner ear (cochlea) by the lower, resolvable harmonics (≤15 f_0).
- Temporal periodicity analysis interprets the temporal structure of the excitation of the cochlea. This is the only operative mechanism for harmonics above 15 f_0, which can no longer be resolved by the cochlea and do not produce any interpretable spectral pattern.

Our research concentrates on the second mechanism.

Psychophysics

In psychoacoustical experiments on temporal periodicity analysis there must not remain any spectral cues that might assist the subject in detecting the periodicity. A periodic stimulus with fundamental frequency f_0 should therefore be high-pass filtered at 15 f_0 or higher so there is no energy in the frequency range left that could give rise to spectral pattern recognition. Reconstitutions of the lower harmonics caused by nonlinearities of the inner ear should be masked by low-pass filtered noise.

The resulting stimulus sounds markedly periodic and can easily be distinguished from aperiodic stimuli with equal spectra. A 2.5% increase in f_0 will be detected (Houtsma and Smurzynski, 1990). They obtained best results with click-like stimuli. Kaernbach and Bering (2001) showed that with improved stimuli the jnd for purely temporal stimuli can be as low as 1.3%. Furthermore, these stimuli conveyed musical information. These results can be explained by the high regularity of the resulting temporal structure of the neural excitation: spikes or short bursts of spikes in equal temporal distances $\tau = 1 / f_0$. With our capability to detect the periodicity and to judge its pitch we are able to interpret temporal structures in the neural excitation which are as fast as 500 Hz.

According to the autocorrelation theory of Licklider (1951), the output of a particular cochlea channel is transferred to a fast line as well as to a delay line (see Figure 1). An array of coincidence neurons calculates the autocorrelation of this output for certain delays. Let the center cell of Figure 1 be the 10-ms cell. It will see the signal on the fast line, and its delayed version 10 ms later on the delay line. A 100-Hz periodic click train will excite this neuron simultaneously via both lines and cause it to fire. The 9-ms cell at its left will never see simultaneous input on both the fast line and the delay line. This model might provide an explanation of how we distinguish a periodic click train from a random click train.

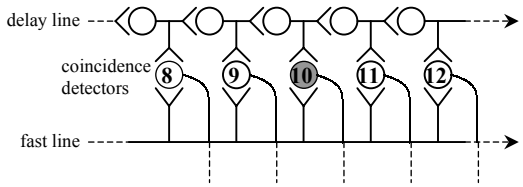

Figure 1: The autocorrelation model of Licklider (1951)

This model has been put into question by psychophysical results based on "second-order periodicity". Imagine a click train where the distance of all successive odd clicks is 10 ms. The even clicks divide these 10-ms intervals randomly in varying portions. Successive intervals could be 2,8,6,4,3,7,5,5... ms, with 2+8 = 6+4 = 3+7 = 5+5. If this signal is high-pass filtered and masked in its low-pass portion in order to test the temporal periodicity analysis, it cannot be distinguished from a random click train (Kaernbach and Demany, 1998).

These findings are not compatible with the autocorrelation model (Figure 1): The 10-ms cell should notice that every second click had a predecessor exactly 10 ms earlier. The findings suggest that the temporal periodicity analysis resorts to the statistics of the first-order inter-spike intervals (ISI). These statistics are heavily disturbed by the even clicks, resulting in a uniform distribution between 0 and 10 ms without a peak at 10 ms. This would explain why second-order periodicity cannot be detected: Apparently, the mechanism for temporal periodicity analysis is not capable of doing simple arithmetic like 2+8 = 6+4 = 3+7 = 5+5.

There is a simple example that proves that this first-order ISI model fails to work correctly. Instead of using one interval of 10 ms that is randomly divided we work with two different but fixed intervals. A click train with a periodic sequence of the two intervals (e.g. 6,10,6,10,6,10... ms) has to be compared with a random sequence of 6-ms and 10-ms intervals. In this case the first-order ISI statistics are the same for both stimuli. However, as shown in the following experiment, human subjects can easily tell the difference.

Experiment 1

The stimuli consisted of sequences of clicks. These clicks were high-pass filtered at 2000 Hz. Possible distortion products were masked with low-pass filtered noise. The clicks formed intervals of 6 (henceforth A) or 10 (B) ms. Three different types of regular click sequences were used: ABABAB... (i.e. periodic(A,B), henceforth ||:AB:||), AABBAABB... = ||:AABB:||, and ||:AABBAB:||. These were to be compared to random click sequences. The random sequences consisted of A and B intervals in random order and could have a maximum of two identical intervals in succession. The click sequences could have 4 different lengths (15, 30, 50 or 70 intervals). The subjects listened to two different click sequences, a regular one and a random one, and the task was to decide whether the first or the second was the regular one. Four students in the age range of 19 to 26 (3 female, 1 male) participated in three sessions in each of which 60 trials for each condition (3 different sequences against random x 4 lengths) were performed. The first session was considered as training and did not contribute to the data that were analyzed.

Results

As seen in Figure 2, ||:AB:|| can be easily discriminated from random click sequences. It's harder to discriminate ||:AABB:|| from random sequences but it's still possible. The only sequence the 4 subjects couldn't discriminate from a random sequence was ||:AABBAB:||. For short sequences of this type, the performance is somewhat lower ($p<0.1$) than 50%, indicating, that subjects perceived this sequence as less regular than random sequences. The performance for ||:AB:|| and ||:AABB:|| gets better for longer click trains but the main difference is between the click trains with a length of 15 and 30 intervals.

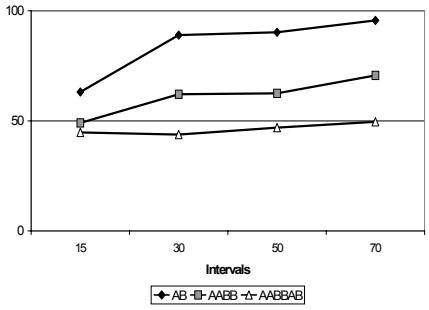

Figure 2: Results of Experiment1. Mean hit rate (%) in dependency of stimulus length (number of intervals) and regularity of stimulus

Experiment 2

In Experiment 1 we only had four subjects. This and the fact that the chosen lengths of the click trains have a bad resolution for the interesting lengths (i.e. between 15 and 30 intervals) lead us to do Experiment 2. The same kind of stimuli were used for this experiment and also the task for the subjects was the same. We now presented 10, 20, 30, 40 and 50 intervals to the subjects. For the click trains with a length of 10 and 20 intervals only, ||:AB:|| and ||:AABB:|| vs. random sequences were presented. Ten students in the age range of 19 to 26 (9 female, 1 male) three of which had already participated in Experiment 1 participated in four sessions in each of which 60 trials for each condition (3 or 2 different sequences against random x 5 lengths) were performed. The first session was considered as training and did not contribute to the data that were analyzed.

Results

Results of Experiment 2 are shown in Figure 3. It can be seen in Figure 3a that again ||:AB:|| can be easily detected. Even click trains that are as short as 10 intervals–that is about 80 ms long–sound more regular than random click trains of equal length. Again, ||:AABBAB:|| sounds less regular than random click trains ($p<0.01$). This is an result we didn't expect. The performance on ||:AABB:|| in this experiment is a bit worse than in Experiment 1. Figure 3b shows the ||:AABB:|| performance for single subjects. As can be seen, there have been only 3 good "||:AABB:|| detectors". Two of them had already participated in Experiment 1. Their performance didn't change during the two experiments, so these effects aren't training effects, as one might suppose. The other subjects had near-chance performance on the ||:AABB:|| sequences. The main difference in the performance of detecting ||:AABB:|| sequences between the two experiments is the ratio of good and bad ||:AABB:|| detectors. However, as can be seen in Figure 3b, there are subjects that are able to detect these sequences very well.

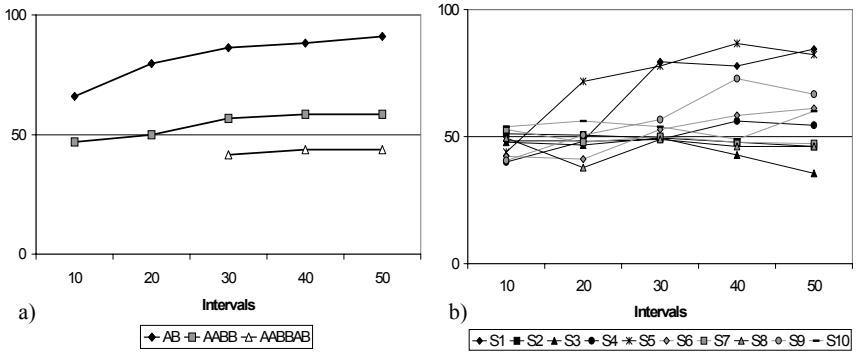

Figure 3: Results of Experiment 2. a) Average hit rate (%) over all subjects in dependency of stimulus length (number of intervals) and stimulus type. b) Hit rate (%) on ||:AABB:|| for each single subject.

Neural model

The model we introduce here is an extension of the above mentioned first-order ISI model. It is a simplification of a neural network model by Kaernbach and Mohlberg (1994) that is able to learn the regularity in a periodic click train via fast synaptic plasticity. First we will present a model that is able to learn the regularity of ||:AB:|| sequences and extend it later, so it can also learn the regularity of ||:AABB:|| and – depending on the parameters – even of ||:AABBAB:|| sequences.

The scheme in Figure 4a shows an array of ISI-sensitive cells including a 6-ms and a 10-ms cell. The stimulation of the 6-ms cell would be optimal by a spike that was proceeded by another spike 6 ms earlier. These cells are connected all-to-all, with connection delays as long as the sensitive interval of the receiving cell. The connection $w_{10,6}$ from the 6-ms to the 10-ms cell would thus have a delay of 10 ms. Let us consider the activity of this network when stimulated with a periodic click train consisting of the intervals 6,10,6,10,6,10...ms. In addition to the stimulation by the output of the cochlear channel at the end of the 10-ms interval the 10-ms cell will furthermore see the output of the 6-ms cell via $w_{10,6}$ which will arrive at the same time if the latter fired at the end of the proceeding 6-ms interval. Hebbian learning rules (cf. Rumelhart & McClelland, 1986; see also Table 1) determine how the connections learn their weights. The connections $w_{10,6}$ and $w_{6,10}$ will be strengthened (see

Figure 4a). This leads to a stabilization of the activity of the neural net for this periodic input. If we, on the other hand, present this network with a random sequence of 6-ms and 10-ms intervals, these connections would be less strong than with stimulus ||:AB:||. This won't help the neural net to reach a stable activation or a stable oscillation of activation as will be shown later.

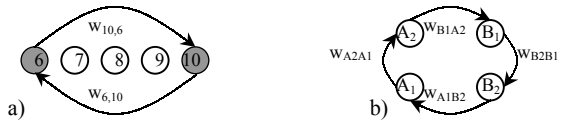

Figure 4: Strengthened connections of the neural net for a a) ||:6,10:|| and a b) ||:AABB|| sequence. In the second panel, a circle (e.g. A1) does not necessarily represent a single cell. It can also stand for a group of cells.

The model presented in Figure 4a is, however, not capable of learning sequences more complicated than ||:AB:||. The extension we have to do to our model so that it can learn the regularity of sequences of higher Markov order (||:AABB:|| etc.) is to stimulate more than just one cell by a single click. We need a couple of cells that are tuned to the same ISI. The cells could then build subgroups in order to represent this more complex stimulus. For ||:AABB:|| the net would build two A subgroups and two B subgroups so that w_{A2A1}, w_{B1A2}, w_{B2B1} and w_{A1B2} would be strengthened which would lead the net to stabilize its activity (see Figure 4b). Exemplary simulation results for this net are shown in Figure 5.

$c_j(t-1)$	$c_i(t)$	Δw_{ij}
1	1	+3/4 λ
1	0	-1/4 λ
0	1	0
0	0	0

Table 1: Hebb learning rule for the changing of weights between neurons. Status of neuron j in time step t-1, status of neuron i in time step t, and change of the strength of the connection from neuron j to neuron i. With λ the learning speed can be varied.

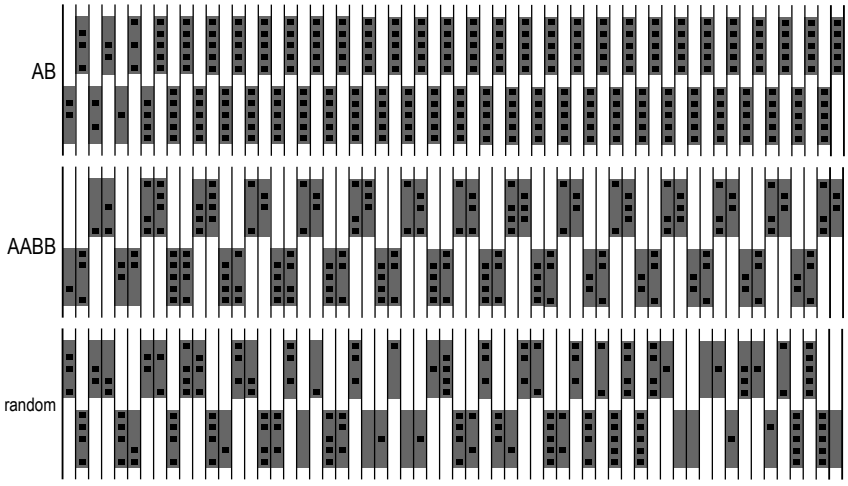

Figure 5 Simulation results for 60 intervals of the stimuli ||:AB:||, ||:AABB:||, and a random sequence. Each column monitors the output of 10 neurons for one time step (interval). The gray shading indicates the input to the cells, and the black dots represent firing cells.

Decision algorithm

A major goal of the present study was to complement the simplified network model described above with a decision algorithm that would allow to simulate behavioral data. As can be seen in Figure 5, for regular stimuli the net will sooner (||:AB:||) or later (||:AABB:||) oscillate in a stable way. This feature of the net will be used by our decision algorithm to decide for the more regular stimulus. Over a certain period of time the standard deviation of the number of firing cells is calculated. High standard deviations result from unstable activation patterns (cf. Fig 5, bottom row), whereas stable oscillations lead to low standard deviations. The standard deviation is then divided by the mean firing rate (standard deviation / mean = relative standard deviation). The smaller this value, the more stable the net. This calculation permits us to simulate a virtual subject: We present the net with a regular and a random sequence. The output of the net is then evaluated by the decision algorithm: The relative standard deviations of the number of firing cells are compared for the two sequences, and the smaller value indicates the more regular sequence. That is how our virtual subject makes its decision, which can be false or true. The percentage of correct responses can then be compared to behavioral results.

Simulation

Each data point in Figure 6 is the result of 1000 comparisons. For the simulation we varied three parameters: the learning speed, the number of neurons optimized for the same ISI and the memory of each neuron (see Details of the Network). Learning speed is defined by the factor with which we multiply the amount in which the weights are changed (see Table 1). With these three parameters we are able to control the number of clicks needed for a correct decision, as well as the relative performance for simple (||:AB:||) and complex (||:AABB:||, ||:AABBAB:||) sequences. The simulation data presented here demonstrate that it is possible to obtain data similar to the results of Experiment 1. In this example, the neural net with the decision algorithm is able to discriminate ||:AB:|| and ||:AABB:|| sequences from random sequences. Even the performance less 50% for ||:AABBAB:|| found in Experiment 1 can be simulated. Individual differences as found in Figure 3b can be reflected by assuming different sets of parameters for different subjects.

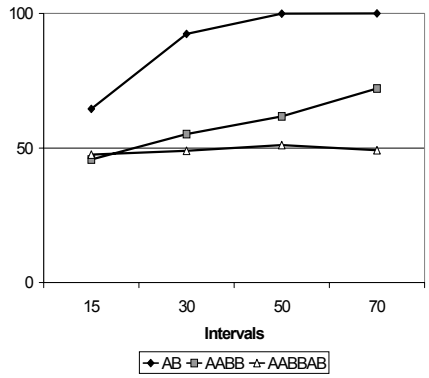

Figure 6: Results of the simulation. Mean hit rate (%) in dependency of stimulus length (number of intervals) and regularity. (Net parameters: 2 blocks of 3 neurons; λ=0.015; memory=16 time steps)

Discussion

Click sequences like ||:AB:|| and ||:AABB:|| can be detected against random click sequences as more regular by subjects as well as by our model. ||:AABBAB:|| seems to sound less regular than random click sequences. By varying the amount how the past firing rate of a neuron is considered to calculate the future firing rate (see details of the network), we can also simulate this finding.

For further experiments it is planned to vary the length of the intervals. Instead of taking 6 and 10 ms for A and B, we could use 7 and 9 ms, or 5 and 11 ms. This will necessitate a change in our model. So far the neurons did either get input from the cochlea or not. If there is a difference in performance depending on the similarity of A and B we will have to reintroduce a Gaussian input function (Kaernbach & Mohlberg, 1994). It appears, however, that the decision algorithm developed with the simplified model can straightforward be applied to the more realistic Gaussian extension.

Details of the Network

A cells sums up a potential V_i that consists of the input by the other neurons in the previous time step and an external field ϕ_i. It fires if V_i exceeds a threshold η=1.

$$V_i(t) = \sum_{j=1}^{N} w_{ij} c_j(t-1) + \phi_i(t-1), \qquad c_i(t) = \Theta(V_i(t) - \eta). \quad (1)$$

The external field $\phi_i(t)$ represents a cell's stimulation by a click at time t. It comprises a fixed component ψ and a noise component $\nu_i(t)$ which is reduced depending on the mean activity $\langle c_i \rangle_x$ of the cell i over the last x time steps (memory). It gets thus the smaller the more the cell fired during its recent history. These components are added and multiplied with the input function (*I*) (1 or 0 depending if the cell is tuned to the ISI or not).

$$\phi_i(t) = [\psi + \upsilon_i(t)\mu_i(t)] * I(i), \qquad \mu_i(t) = (1 - \langle c_i \rangle_x)^2. \quad (2)$$

The synaptic weights are modified according to Hebbian learning rules (see Table 1), keeping them nonnegative. After the update, the weights are renormalized such that the sum rule $\sum_{j=1}^{N} w_{ij} = 0.9\eta$ holds. The fixed component ψ of the external field was set to 0.6, the noise component $\nu_i(t)$ was chosen uniformly from [0, 0.7]. There were no connections from cell i to cell i.

References

Houtsma, A.J.M., & Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.*, 87, 304-310.

Kaernbach, C., & Bering, C. (2001). Exploring the temporal mechanism involved in the pitch of unresolved harmonics, *J. Acoust. Soc. Am.*, 110, 1039-1048.

Kaernbach, C., & Demany, L. (1998). Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *J. Acoust. Soc. Am.*, 104, 2298-2306.

Kaernbach, C., & Mohlberg, H. (1994). A neural sequence-learning model to explain auditory periodicity analysis, in *Proceedings of the ICANN 1994*, Sorrento, Italy. Eds. M. Marinaro and P. G. Morasso (Springer, New York), vol. II, 909-912.

Licklider, J.C.R. (1951). A duplex theory of pitch perception. *Experentia*, 7, 128-134.

Rumelhart, D. E. & McClelland, J. L. (1986). *Parallel distributed processing*, vol. 1, MIT Press, Cambridge, MA.