Proceedings of the 25th Anniversary Conference of the European Society for the Cognitive Sciences of Music, 31 July-4 August 2017, Ghent, Belgium
Van Dyck, E. (Editor)

# Periodicity Detection by Neural Transformation

Frieder Stolzenburg

*Department of Automation and Computer Sciences, Harz University of Applied Sciences, Wernigerode, Germany*
fstolzenburg@hs-harz.de

## ABSTRACT

It is well-known that during the hearing process in the ear and in the brain an acoustic stimulus, e.g. a musical harmony, is transformed in a highly non-linear way. This can be studied by comparing the frequency spectrum of the input stimulus and its response spectrum in the auditory brainstem. The latter shows additional frequencies which are not present in the input spectrum, in particular the periodicity pitch frequency (also known as missing fundamental). The aim of this paper is to find out relevant factors that lead to this occurrence of the periodicity pitch. The most important factor during the neural transformation seems to be the transformation of the input signal into pulse trains (spikes) whose maximal amplitude is limited by a fixed uniform value. This is shown by comparing the response frequency spectrum in the brain with one computed by an artificial neural network.

## I. INTRODUCTION

During the hearing process in the ear and the brain an acoustic stimulus, e.g. a musical harmony consisting of several complex tones, is transformed in a highly non-linear way. The input signal, which may be a superposition of periodic signals of certain frequencies and their overtones, undergoes a neural transformation in the brain.

Lee *et al.* (2009, 2015) investigated this by measuring the auditory brainstem response to several musical intervals. They report experiments with several musical intervals, discovering that the phase-locking activity to the temporal envelope is more accurate (i.e. sharper) in musicians than non-musicians. For instance, (a) the perfect fifth A2–E3 (i.e. tones with 110 and 166 Hz and approximate frequency ratio 3:2), shows the highest response in the brainstem at about $55.3 \approx 110/2$ Hz, and (b) the minor seventh F#2–E3 (93 and 166 Hz, frequency ratio 9:5) at about $18.5 \approx 93/5$ Hz. In both cases, the additionally occurring frequencies coincide very well with the periodicity pitch frequencies of the respective musical intervals.

Recent results from neuroscience demonstrate that periodicities of complex chords can be detected in the human brain by a system of several neurons (Langner, 1997, 2015). Firstly, there are oscillator neurons showing regularly timed discharges in response to stimuli, not corresponding to the temporal structure of the external signal, i.e. intrinsic oscillation. The oscillation intervals can be characterized as integer multiples $n \cdot T$ of a base period of $T = 0.4$ ms with $n \geq 2$ for endothermic, i.e. warm-blooded animals (Langner, 2015, Chapter 5). The external signal is synchronized with that of the oscillator neurons, which limits signal resolution.

Secondly, there are trigger neurons that transfer signals without significant delay. In contrast to them, integrator neurons respond with a certain amount of delay. In the dorsal cochlear nucleus, periodic signals are transferred with different delays. There, onset latencies of integrator neurons up to 120 ms have been observed (Langner and Schreiner, 1988).

When the delay corresponds to the signal period, the delayed response and the non-delayed response to the next modulation wave coincide. By this procedure, the missing fundamental tone and hence periodicity pitch can be detected in the brain. Both groups of neurons are synchronized by the oscillator neurons.

According to Langner (1997, 2015), pitch and timbre (i.e. frequency and periodicity) are mapped temporally and also spatially and orthogonally to each other in the auditory midbrain and auditory cortex as a result of a combined frequency-time analysis that is some kind of autocorrelation mechanism by comb-filtering, including phase locking, which means that phase differences among different signals can be neglected.

But the question remains why actually the periodicity pitch occurs in the frequency response spectrum in the brain, although it is present neither in the original signal nor in a superposition of the signal with delayed versions thereof, because summation of signal waveforms does not alter the frequencies in the respective spectra, only their amplitudes. In addition, autocorrelation may introduce only overtones into the spectra, i.e. integer multiples of the frequencies in the original signals, but not subharmonics like the periodicity pitch frequency.

Lee *et al.* (2009, 2015) mention combination tones as a possible cause. They are artificially perceived when two real tones with two frequencies $f_1 < f_2$ sound at the same time and have percepts corresponding to the frequencies $f_1 - k \cdot (f_2 - f_1)$ where $k$ is a positive integer. Combination tones are derived from the distortion products (cf. Hartmann, 1997, Chapter 22) generated by the non-linear behavior of the auditory system. However, there are many more combination tones than tones occurring in the response spectra, and the value of $k$ yielding the periodicity pitch frequency may be rather high.

## II. AIMS

The aim of this paper is therefore to find out more precisely the relevant factors that lead to the occurrence of the periodicity pitch in the response spectrum of a signal. Reasons may be (among others):

1. Phase-locking induced by oscillator neurons with intrinsic oscillation frequency, different from the frequencies in the input;
2. Autocorrelation or distortion products which may be realized by superposition of the input signal and a delayed version of it;
3. The transformation of the input signal into pulse trains (spikes) whose maximal amplitude is limited by a fixed uniform value.

We already briefly discussed reasons #1 and #2 in the introduction (Section I). On the one hand, they allow to derive the periodicity pitch, but on the other hand they do not exactly explain why the periodicity pitch frequency is physically present in the auditory brainstem response. Let us thus inspect reason #3 in more detail:

In the brain, spikes are created when the action potential of a neuron crosses some threshold, namely if the net excitation, received by a neuron over a short period of time, is large enough. As a side effect, the amplitude of the signal is limited by this procedure. As we will demonstrate next (in Sections III and IV), this kind of distortion of the original signal yields the desired result, because then only specific combination tones are present in the frequency spectrum.

## III. METHOD

The stimuli from other studies are used as input to a theoretical model and compared with the corresponding response spectra in the brain. Lee *et al.* (2015) use an electric piano sound (Fender Rhodes) recorded from a digital synthesizer. The stimuli are binaurally presented to adult subjects through insert headphones, and the responses are collected using several scalp electrodes. The waveform of the original stimulus, in this case the perfect fifth mentioned in the introduction, is shown in Figure 1 in blue. As one can see, the signal has an overall period length of about 18.1 ms. It corresponds to the periodicity pitch of 55.3 Hz (cf. Section I).

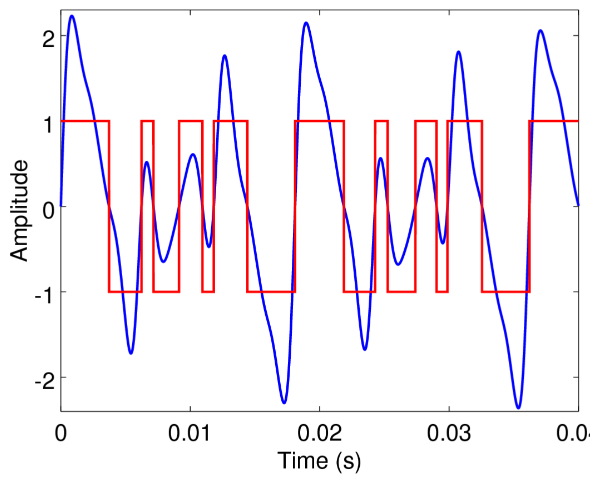


**Figure 1. Perfect fifth (electric piano sound): original signal (blue); amplitude-limited response (red).**

In the following, we introduce a recurrent artificial neural network model for generating and perceiving such periodic waveforms. Generally, an artificial neural network consists of many neurons which are connected with each other. In recurrent networks, e.g. echo state networks (Jaeger, 2001, 2007), the neurons may be recursively connected and the activation of each neuron changes over time. If the neurons $x_1, \ldots, x_n$ are connected to a neuron y, then it holds $y(t+\tau) = g(w_1 \cdot x_1(t)+\ldots+w_n \cdot x_n(t))$. Here, $w_1, \ldots, w_n$ are weights, $\tau$ is a discrete time constant, and g is the so-called activation function.

The activation function g may be the identity. In this case, one speaks of linear activation. With linear activation and also with deviations thereof (see below), pure cosine and sine waves can be generated by a simple recurrent network consisting of only two neurons (see Figure 2). If we want to generate $x_1(t) = \alpha \cdot \cos(\omega \cdot t+\varphi)$ and $x_2(t) = \alpha \cdot \sin(\omega \cdot t+\varphi)$, respectively, where $\alpha$ is the amplitude, $\omega = 2\pi \cdot f$ the angular frequency, and $\varphi$ the phase shift of the oscillation, then the initial state must be $x_1(0) = \alpha \cdot \cos(\varphi)$ and $x_2(0) = \alpha \cdot \sin(\varphi)$. Applying the trigonometric addition theorems, we obtain the following recursion formulas below where the coefficients correspond to a rotation matrix with rotation angle $\rho = \omega \cdot \tau$:

$$x_1(t+\tau) = \cos(\rho) \cdot x_1(t) - \sin(\rho) \cdot x_2(t)$$

$$x_2(t+\tau) = \sin(\rho) \cdot x_1(t) + \cos(\rho) \cdot x_2(t)$$

More complex waveforms are obtained by superposition, i.e. summation of several different simple waveforms. Usually, a non-linear, strictly increasing sigmoidal activation function g is used in artificial neural networks (cf. Goodfellow *et al.*, 2016), e.g. the logistic function, the hyperbolic or arc tangent, or simply the sign function. By this, the input signal is transformed into a rectangular pulse train with uniform maximal amplitude. For our running example, the perfect fifth, Figure 1 shows in red the amplitude-limited response of the artificial neural network by applying the sign function as activation function g to the input signal.

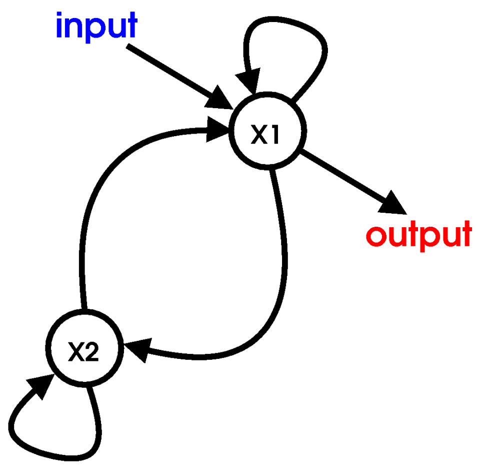


**Figure 2. Recurrent artificial neural network for sinusoid generation and perception. Two neurons suffice. Each neuron can act as input and/or output of an external signal.**

## IV. RESULTS

From the given waveforms, the frequency spectra of the original signal and the amplitude-limited signal can be easily computed. This as well as the implementation of the recurrent artificial neural networks has been done by means of a MatLab/Octave program (Higham and Higham, 2017) written by the author. For this, a Fourier transformation has to be performed (cf. Hartmann, 1997, Chapter 8). In the implementation, the discrete variant is used, the Fast Fourier Transformation (cf. Hartmann, 1997, Chapter 21).

Figure 3 shows the frequency spectra of the perfect fifth (electric piano sound) for the original signal (in blue) and its amplitude-limited response (in red). The periodicity pitch occurs physically in the real brainstem response (see Lee *et al.*, 2015, Figure 5) and the predicted frequency spectrum, although our recurrent artificial neural network model is rather simple. The key point is the non-linear, sigmoidal activation. Neither an autocorrelation analysis nor a complex neural model, e.g. with oscillating neurons (see e.g. Shapira Lots and Stone, 2008, and Lerud *et al.*, 2014), is required.

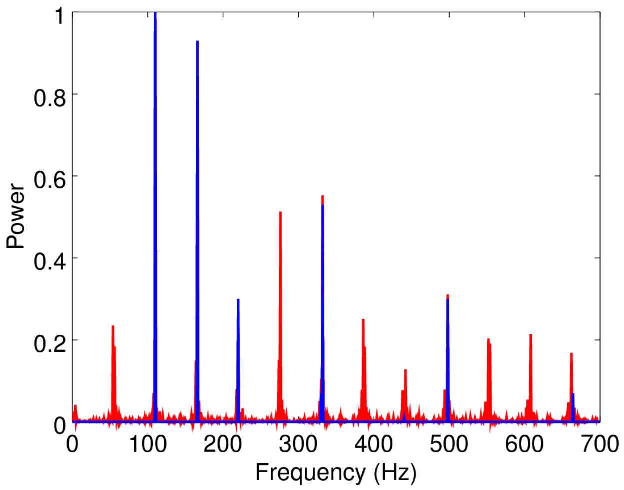


**Figure 3. Frequency spectra of perfect fifth (electric piano sound): original signal (blue); amplitude-limited response (red).**

However, until now only a qualitative comparison of the brainstem response spectrum with the spectrum predicted by the model has been done. Therefore, of course, this point needs further investigation. Nevertheless, both the real brainstem response frequency spectrum and that computed by the theoretical model show:

- The response spectra contain as expected in addition to the original spectrum first and foremost the periodicity pitch frequency, not arbitrary difference tones.
- The peaks in the response spectrum are sharper the more pulse-like the transformed input is.
- The peaks at the periodicity pitch frequencies are more salient for more consonant harmonies. In this case, the periodicity pitch frequency is comparatively high.

The latter means that the relative periodicity, as defined by Stolzenburg (2015), is relatively low. Relative periodicity denotes the approximated ratio of the period length of the musical harmony (i.e. its periodicity pitch) relative to the period length of its lowest tone component. The good correlation between relative periodicity and consonance has been shown extensively by Stolzenburg (2015). The periodicity pitch frequency of our perfect fifth is approximately 55.3 Hz. This corresponds exactly to the first bigger peak in the response spectrum in Figure 3 (in red).

How does limiting the amplitude of the input waveform introduce additional frequencies into the signal, in particular the periodicity pitch frequency? In order to understand this more precisely, we consider the input signal as a sequence of rectangular pulses with uniform amplitude, as in the brain, and analyze its frequency spectrum. A similar procedure has been undertaken by Ebeling (2007, 2008) in his mathematical model for the analysis of the perception of consonance.

For the perfect fifth with ground tone frequencies of 110 and 166 Hz, we have pulses every $1/110 \approx 9.1$ ms and every $1/166 \approx 6.0$ ms, respectively. After an overall period of approximately 18.1 ms, which corresponds to the period length of the missing fundamental tone, both signals coincide. This can be seen in Figure 4 (in blue). Therefore, the amplitude is not uniform at this point. Limitation of the amplitude is achieved by introducing an additional signal that has the frequency of the periodicity pitch, again consisting of rectangular pulses. This is also shown in Figure 4 (in red). This is the reason why the periodicity pitch frequency is present in the amplitude-limited signal.

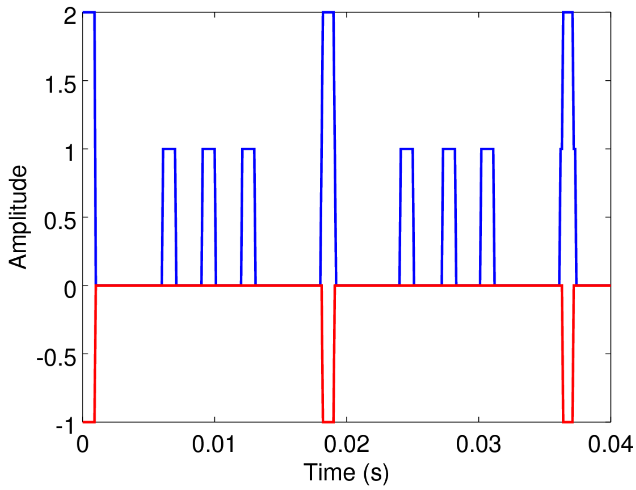


**Figure 4. Perfect fifth with rectangular pulses: original signal (blue); additional periodicity pitch frequency signal to limit the amplitude to a uniform height (red).**

In general, the additional frequencies of a complex harmony comprising several rectangular pulses with different frequencies can be determined as follows. In this context, we assume that the respective frequency ratios are integer fractions. If this is not the case, they can be approximated up to a fixed accuracy, e.g. 1% (see e.g. Forišek, 2007). This procedure yields us the harmonic-series presentation of the given harmony where the real frequencies $f_1, \ldots, f_n$ are mapped to a set of abstract frequencies and their ratios that are small integer numbers (cf. Stolzenburg, 2015, Section 3.2). For instance, we have the following abstract frequency ratios:

(a) 4:5:6 for the major triad in root position
(b) 3:2 for the perfect fifth

In the amplitude-limited signal, all rectangular pulses must have uniform amplitude. This means, whenever two pulses of different frequencies coincide, it has to be compensated (as shown in Figure 4 in red). In consequence, the set of abstract frequencies has to be extended iteratively by all possible greatest common divisors. These extended sets of abstract frequencies for our two examples are (a) 1,2,4,5,6 and (b) 1,2,3, respectively. The amplitudes of the newly introduced abstract frequencies (here: 1,2 and 1, respectively) are -1 in most cases. This holds in particular for the periodicity pitch frequencies of our two examples which correspond to the abstract frequency 1. In general, the absolute value of the amplitudes may differ from 1 in this model calculation. For this, starting with the uniform amplitude $\alpha(f) = 1$ for each abstract frequency f in the extended frequency set, each amplitude is corrected by setting $\alpha(f_1) = \alpha(f_1) - \alpha(f_2)$ whenever the abstract frequency $f_1$ is a divisor of $f_2$. Anyway the periodicity pitch occurs in the frequency spectrum, as desired.

## V. CONCLUSIONS

In summary, the most important factor during the neural transformation for periodicity detection seems to be the spiking with uniform, limited amplitude (i.e. reason #3, cf. Section 2). Even for random phase difference or slightly mistuned intervals the results do not change. Autocorrelation, intrinsic oscillation, phase-locking, or similar mechanisms (see e.g. Shapira Lots and Stone, 2008, and Lerud *et al.*, 2014) are not needed to explain the response spectra but nevertheless correlate well with the empirical findings. The result that the periodicity pitch appears in the response spectrum and not arbitrary difference tones can be reproduced by Fourier analysis of amplitude-limited pulse trains.

Nonetheless, it remains an interesting research question whether a similar effect can be noted also in the response spectrum when the harmonic tones are not presented simultaneously, but in succession. This should be the subject of future work. In addition, more extensive studies and comparisons with real brainstem responses have to be done. Last but not least, the recurrent artificial neural network model should be developed further.

## REFERENCES

Ebeling, M. (2007). *Verschmelzung und neuronale Autokorrelation als Grundlage einer Konsonanztheorie*. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien: Peter Lang.

Ebeling, M. (2008). Neuronal periodicity detection as a basis for the perception of consonance: A mathematical model of tonal fusion. *Journal of the Acoustical Society of America, 124*(4), 2320-2329.

Fishman, Y. I., Micheyl, C., & Steinschneider, M. (2013). Neural representation of harmonic complex tones in primary auditory cortex of the awake monkey. *The Journal of Neuroscience, 33*(25), 10312-10323.

Forišek, M. (2007). Approximating rational numbers by fractions. In Crescenzi, P., Prencipe, G., & Pucci, G., (Eds.), Fun with algorithms. In *Proceedings of 4th International Conference, LNCS 4475* (pp.156-165). Castiglioncello, Italy: Springer.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning. Adaptive Computation and Machine Learning*. Cambridge, MA, London: MIT Press.

Hartmann, W. M. (1997). *Signals, Sound, and Sensation*. Woodbury, NY: American Institute of Physics.

Higham, D. J., & Higham, N. J. (2017). *MatLab Guide (3rd edition)*. Philadelphia, PA: Siam.

Jaeger, H. (2001). *The "echo state" approach to analysing and training recurrent neural networks*. German National Research Center for Information Technology (GMD), Technical Report 148. With an erratum note, 2010.

Jaeger, H. (2007). Echo State Networks. *Scholarpedia, 2*(9), 2330.

Langner, G., & Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat: I. Neuronal mechanisms, II. Topographical organization. *Journal of Neurophysiology, 60*(6), 1799-1840.

Langner, G. (1997). Temporal processing of pitch in the auditory system. *Journal of New Music Research, 26*(2), 116-132.

Langner, G. (2015). *The Neural Code of Pitch and Harmony.* Cambridge, UK: Cambridge University Press.

Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *The Journal of Neuroscience, 29*(18), 5832-5840.

Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2015). Neural transformation of dissonant intervals in the auditory brainstem. *Music Perception, 32*(5), 445-459.

Lerud, K. D., Almonte, F. V., Kim, J. C., & Large, E.W. (2014). Mode-locking neurodynamics predict human auditory brainstem responses to musical intervals. *Hearing Research, 308*, 41-49.

Shapira Lots, I., & Stone, L. (2008). Perception of musical consonance and dissonance: An outcome of neural synchronization. *Journal of the Royal Society Interface, 5*, 1429-1434.

Stolzenburg, F. (2015). Harmony perception by periodicity detection. *Journal of Mathematics and Music, 9*(3), 215-238. Extended version available at http://arxiv.org/abs/1306.6458.