Physiological correlates of the perceptual pitch shift for sounds with similar waveform autocorrelation

Daniel Pressnitzer, Alain de Cheveigné
Institut de Recherche et Coordination Acoustique/Musique-Centre National de la Recherche Scientifique (CNRS),
1 place Stravinsky, 75004 Paris, France
Daniel.Pressnitzer@ircam.fr, Alain.de.Cheveigne@ircam.fr

Ian M. Winter
The Physiological Laboratory, University of Cambridge, Downing Site, Cambridge CB2 3EG, England
imw1001@cus.cam.ac.uk

Abstract: A perceptual experiment shows that random click trains with a uniform interclick distribution can be reliably pitch-matched to pseudo-periodic click trains. The pitch matches cannot be explained on the basis of mean rate, power spectrum, or autocorrelation of the waveform. The matches are qualitatively, but not quantitatively, consistent with the most common interspike interval present in responses of single units from the ventral cochlear nucleus of anaesthetised guinea pigs. The physiological recordings also demonstrate that at the level of the cochlear nucleus, similar cues are found in either first-order or all-order interspike interval statistics.

PACS numbers: 43.64.Qh, 43.66.Hg [DWG]
Date Received: February 25, 2003 **Date Accepted:** October 28, 2003

1. Introduction

Many hearing theories posit a temporal mechanism to extract the regularities contained within complex sounds. The nature of the mechanism is still a matter of debate. Licklider (1951) proposed autocorrelation as a way to detect all types of regularities present in neural spike trains. The peak in autocorrelations of auditory nerve spike trains was found to produce correct pitch estimates for a wide range of stimuli (Cariani and Delgutte, 1996). However, the validity of the autocorrelation model has been questioned in several studies. For instance, Kaernbach and Demany (1998) and Kaernbach and Bering (2001) presented listeners with click trains that contained regular intervals, either between consecutive clicks (first-order) or nonconsecutive clicks (second-order). First-order regularities were much more discriminable from random click trains than were second-order regularities. Assuming that for such stimuli, the waveform reflects the activity within the auditory nerve, they interpreted their results as contradicting the autocorrelation model.

Using related pseudo-periodic click trains, Pressnitzer *et al.* (2001) found that similar waveform autocorrelations can produce different pitches. This pitch-shift effect could be qualitatively modelled by including a non-linearity before computing the autocorrelation, highlighting the difference between a waveform autocorrelation and autocorrelation of the output of an auditory model. Pressnitzer *et al*. (2001) also found reliable pitch matches to click trains with second-order regularities, the same regularities that Kaernbach and colleagues had shown to be difficult to discriminate from random click trains. This led to the hypothesis that random click trains (with a bounded, uniform interclick distribution) could produce some sense of pitch. The aim of the present study is twofold: 1) to examine whether random click trains may have a pitch 2) and to see how the perceptual results relate to responses of single neurons at an early stage in the auditory pathway.

2. First- and second-order regularities in click trains

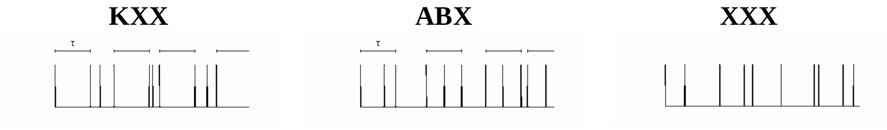


Fig. 1. Click stimuli used in the present experiments. KXX has one first-order regular interclick interval of duration τ, followed by two random intervals; ABX has one second-order regular interclick interval of duration τ, followed by a single random interval; XXX is made of random intervals of maximum duration τ.

The three types of sound used in this study are illustrated in Fig. 1. For KXX, a regular interval of duration τ is followed by two random intervals drawn from the distribution $[0,\tau/2]$. For ABX, a regular interval of duration τ is split in two by a randomly interspersed click and then followed by a single random interval drawn from the distribution $[0,\tau]$. Finally, for XXX, all intervals are drawn uniformly from the distribution $[0,\tau]$. These stimuli and denominations were introduced by Kaernbach and Demany (1998). We modified their original definition of KXX, however, so that for a same regular interval, τ, all stimuli have the same average rate. The stimuli differ in the nature of the regularity they carry. KXX is said to contain a first-order regularity, ABX a second-order regularity, and XXX no regularity at all.

The constraints put on the stimuli are reflected in their interclick statistics. We computed interval histograms either between successive clicks (first-order) or between all possible pairs of clicks (all-order). This analysis is identical to histograms used by physiologists to characterize spike trains, except that the computation is done here on clicks rather than spikes. We term the first- and all-order interclick histograms FOI-clicks and AOI-clicks, respectively. AOI-clicks is equivalent to the autocorrelation of the waveform.

Figure 2 shows that the FOI-clicks from ABX and XXX are similar and do not display any peak, whereas there is a distinct peak at τ for KXX. In contrast, all stimuli exhibit a peak around τ for the AOI-clicks, including XXX. The autocorrelation of a bounded uniform interval distribution is not flat and is expected to exhibit such a feature. The peak is, however, rather small.

Kaernbach and Demany (1998) and Kaernbach and Bering (2001) showed that the KXX-XXX discrimination was easy (using their original definition of KXX), whereas ABX-XXX was almost impossible. Considering that the all-order statistics are visibly different for ABX and XXX, they concluded that any temporal mechanism should be blind to higher-order intervals. Pressnitzer *et al.* (2001) showed that listeners could consistently pitch-match KXX and ABX, which indicates that ABX does evoke some sort of pitch. The following psychophysical experiment was designed to test the hypothesis that XXX can also be pitch-matched to other sounds.

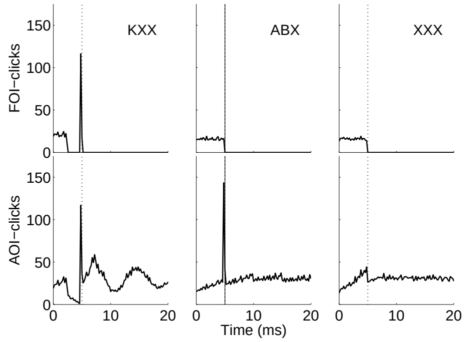


Fig. 2. First- and all-order interval statistics for the click trains, in units of intervals/s. The value of τ is 5 ms, so all click rates are 400 clicks/s. Average over 50 repeats, bin width 200µs.

3. Pitch-matching of random click trains

3.1 Methods

The stimuli used were those described in Fig. 1. A reference, regular interval duration of τ=5 ms was chosen. The stimuli were 409.6 ms long, with 25 ms onset and offset ramps ($\cos^2$). Stimuli were sampled at 44.1 kHz and high-pass filtered above 3 kHz (highpass Butterworth filter, 96 dB/octave). Lowpass noise was introduced to mask potential distortion products between 0 and 3 kHz. The noise was obtained by filtering gaussian noise with a lowpass filter having the same, 3-kHz cutoff frequency, and the level was adjusted so that the whole stimulus had a flat spectral envelope. The overall stimulus level was 60 dB SPL.

The experimental task was a pitch comparison. A 2AFC adaptive procedure was used to minimize judgments bias (Jesteadt, 1980). Listeners heard a pair of sounds, the reference and the test, and they had to decide which one had a higher pitch. Listeners could thus use any pitch cues and not necessarily musical pitch cues. The order of reference and test was random from trial to trial. According to the listener's response, the τ-value of the test stimulus was adaptively modified (2-down, 1-up). Eight reversals were measured, and the thresholds were estimated as the mean of the last 4 reversals. Two adaptive tracks were interleaved. In one of them, the test sound started with a τ 50% higher than the reference, in the other it started with a τ 50% lower. The mean of the two thresholds for the two tracks constitute one match value.

Listeners ran two kinds of experimental blocks. The test sound was always XXX, but in one kind, the reference was KXX, and in another kind, the reference was ABX. Each listener performed 5 blocks of each type in a random order. Three normal-hearing listeners participated in the study.

3.2 Results

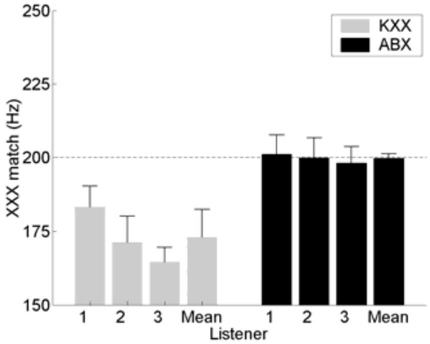


Fig. 3. Results of the pitch-matching experiment. Mean and within-listener standard deviations are plotted (1-3), as well as the average of matches (Mean) and between-listeners standard deviations. The dotted line indicates a 200 Hz match that would correspond to equal click-rates between stimuli.

Results of the perceptual experiment are shown in Fig. 3. The value plotted is the τ-value of XXX at the matching point. For ABX, all three listeners reached their matching point around 200 Hz. Because the reference τ was 5 ms, this indicates that XXX and ABX have the same τ when they are perceived as having a similar pitch. The standard deviation within and between listeners is relatively small, which indicates that at least one pitch cue could be reliably used by listeners to perform the task. A first kind of cue consistent with the match could be the average click rate of the stimuli. Another kind could be linked to the interclick statistics of the stimuli: both ABX and XXX produce a peak at the regular interval, albeit of different height.

The results are quite different for KXX. Here, all listeners matched a longer τ-value for XXX. There is a larger spread between subjects than for ABX, which indicates that possibly more than one cue was available for the judgments. However, the mean match is clearly lower than 200 Hz. This demonstrates that the average rate was not the main cue for the match; other cues linked to the temporal characteristics of the stimuli must have been used. The following physiological experiments examine the possible nature of these cues.

4. Encoding of click trains in the ventral cochlear nucleus

4.1 Methods

Single units recordings were collected from the ventral cochlear nucleus of the anaethetised guinea pig. The physiological methods used have been described in detail elsewhere (Winter and Palmer, 1995). We have recorded responses to the stimuli described in Fig. 1, with an interval τ of 5 ms, from 25 units. Experiments have been carried out under the terms and conditions of the project licence issued by the United Kingdom Home Office to the third author.

Upon isolation of a single unit, estimates of best frequency (BF) and threshold were obtained using audio-visual criteria. Single units were classified by their peri-stimulus time histogram shape in response to suprathreshold BF tone bursts, their interspike interval and discharge regularity. The unit population was subdivided into three different types; primary-like, chopper, and onset. We used the coefficient of variation of the discharge regularity to classify a unit as primary-like or chopper (Young *et al.*, 1988). To identify a unit as an onset unit we have used the classification scheme of Winter and Palmer (1995).

The stimuli were sampled at 40 kHz and high-pass filtered at 3 or 6 kHz according to the BF of each unit. Lowpass noise was added up to the cutoff frequency. Stimulus duration and gating were the same as used for the psychophysics. The sound level was adjusted to produce a reasonably high discharge rate. This meant that the sound level could vary across units between 30 and 50 dB suprathreshold. Peri-stimulus time-histograms were collected for 25 nonidentical repeats of the stimuli. First-order interspike histograms (FOI-spikes) and all-order interspike histograms (AOI-spikes) were computed.

4.2 Results

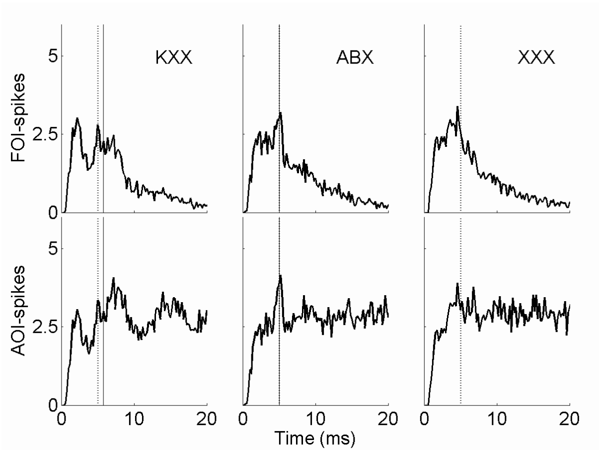
To estimate the nature of the temporal information across the population of neurons, we will present average FOI-spikes and AOI-spikes. The averages were done for the three main types of units: primary-like (N=5); chopper (N=14); and onset (N=6). All onsets were onset-choppers, whereas choppers were either transient (N=7) or sustained (N=7). Differences in individual firing rates exist between units, but we verified that a rate normalization would not change the histograms' main features described below.

The average results for the primary-like units are likely to reflect the activity in primary auditory nerve fibers. For this population, the FOI-spikes histograms have a characteristic Poisson-distribution shape, with superimposed peaks corresponding to the regular interval. A peak at 5 ms is clearly visible in FOI-spikes for ABX and XXX even though their FOI-clicks did not display such a feature (see Fig 2). This is because individual neurons do not generally respond with a spike for each and every click. As a consequence, some *second-order interclick* regularities will always be transformed into *first-order interspike* regularities. For all unit types, the total number of spikes has been compared to the number of clicks presented, and this showed that the units responded on average to 24.5% of the clicks. This probability of firing is typical for click trains in the cochlear nucleus (Burkard and Palmer, 1997).

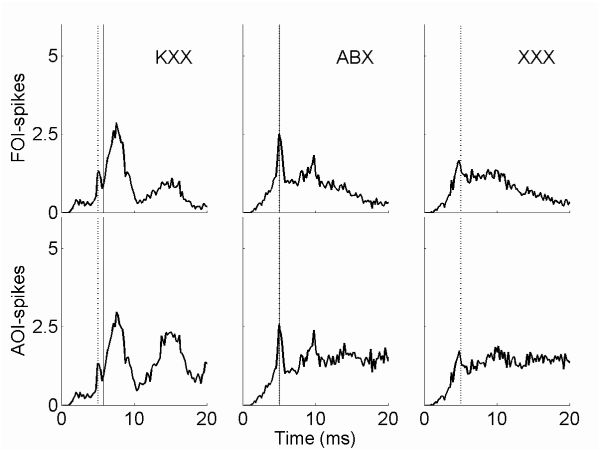
The position of the peak for XXX and ABX corresponds to 5 ms. In contrast, the main peak for KXX is shifted towards longer intervals. This shift provides a qualitative correlate of the perceptual pitch shift as it occurs in the same direction – longer intervals and lower pitch. However, the magnitude of the physiological peak shift (2.5 ms) is larger than the magnitude of the psychophysical pitch shift (0.8 ms).

The results for the onsets and choppers are similar to the ones from the primary-likes, except for an enhanced contrast of the dominant peaks of the distributions for onsets and a substantial reduction in short intervals for both onsets and choppers.

PRIMARY-LIKE



ONSET



CHOPPER

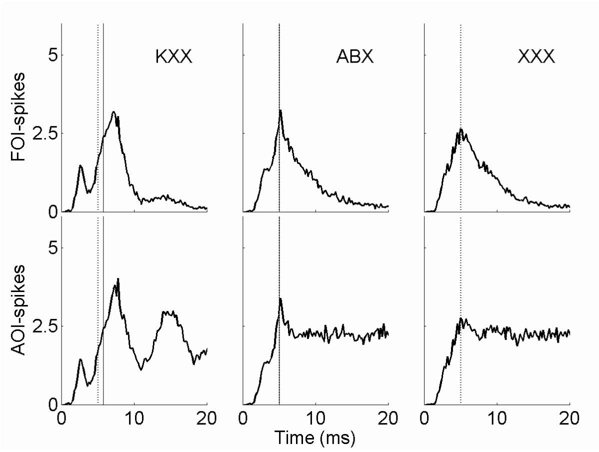


Fig. 4. Physiological results. Averaged first- and all-order interspike histograms are presented in units of interval/s for three main VCN unit types. Bin width is 200 µs. Average firing rate is 98.2 spikes/s. The dotted lines indicate $\tau = 5$ ms, the value used for all stimuli. The solid lines indicate the mean perceptual results. KXX was found to sound lower in pitch than ABX or XXX.

5. Discussion

The present results show that ‘random’ click trains (with intervals drawn from a bounded uniform distribution) can be used for pitch matches. Possible pitch cues exist for random click trains. Mean rate is one of them, but it cannot explain the match between XXX and KXX that was obtained for unequal rates. Spectral cues can be ruled out because the random click train has a flat spectral envelope after high-pass filtering. Thus, other temporal regularity cues were used by listeners during the pitch-matching experiments.

The responses of ventral cochlear nucleus neurons to the click trains showed, on average, a time-locked activity to the clicks. Similar cues were visible in first- and all-order interspike interval distributions. For second-order and random click trains (ABX and XXX), all types of neurons exhibited a peak in their interspike distributions corresponding to the regular interval, τ. This could provide the basis for a pitch cue for those two types of stimuli. A shift in the distribution towards longer intervals was observed in response to the remaining stimulus, KXX. This is in qualitative agreement with the lower perceived pitch of KXX compared to both ABX and XXX. The position of the peak in the interspike distribution, however, does not reflect the magnitude of the perceptual pitch shift. This represents a problem for models of pitch based on the most common interspike interval.

Several hypotheses can be advanced to reconcile the magnitudes of the physiological and perceptual shifts. Carlyon et al. (2002) used click trains with 4 ms and 6 ms intervals and showed that subjects typically heard a pitch corresponding to 5.7 ms, a result not predicted by the most common interval hypothesis. They suggested weighting the first-order interval distributions to give more weight to the longer intervals. However, if applied to our physiological data, this strategy would increase rather than decrease the discrepancy with the psychophysics. Another possibility is to note that more than one peak is visible in the KXX interspike distributions, for primary-like and chopper units (below and above the perceptual match). Judgments based on these peaks could produce an ambiguous pitch somewhere in between. Finally, listeners could achieve a pitch match by adjusting the whole of the underlying distributions, instead of just using peaks. A simple distance measure between histograms (Meddis and O’Mard, 1997) would lead to a prediction more in line with the magnitude of the effect.

The order of the regularity embedded in the stimulus is lost at the level of single auditory nerve fibers. For a pitch mechanism to behave as if it depended on waveform first-order intervals, information must be combined across several fibers. However, we have found no evidence for such a recoding in the responses of primary-like, chopper or onset-chopper units at the level of the ventral cochlear nucleus.

References

Burkard, R., and Palmer, A.R. (**1997**). “Responses of chopper units in the ventral cochlear nucleus of the anaesthetised guinea pig to clicks in noise and click trains,” Hear. Res. **110**, 234-250.

Cariani, P.A., and Delgutte, B. (**1996**). “Neural correlates of the pitch of complex tones. I. Pitch and pitch salience,” J. Neurophysiol. **76**, 1698-1716.

Carlyon, R.P, van Wieringen, A., Long, C.J., Deeks, J.M., and Wouters, J. (**2002**). “Temporal pitch mechanisms in acoustic and electric hearing,” J. Acoust. Soc. Am. **112**, 621-633.

Jesteadt, W. (**1980**). “An adaptive procedure for subjective judgments,” Percept. Psychophys. **28**, 85-88.

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

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

Licklider, J.C.R. (**1951**). “A duplex theory of pitch perception,” Experientia **7**, 128-133.

Meddis, R., and O’Mard, L. (**1997**). “A unitary model of pitch perception,” J. Acoust. Soc. Am. **102**, 1811-1820.

Pressnitzer, D., de Cheveigné, A., and Winter, I.M. (**2001**). “Perceptual pitch shift for sounds with similar waveform autocorrelation,” Acoustical Research Letters Online **3**,1-6.

Winter, I.M., and Palmer A.R. (**1995**). “Level dependence of cochlear nucleus onset unit responses and facilitation by second tones or broadband noise,” J. Neurophysiol. **73**,141-159.

Young, E.D., Robert, J.M., and Shofner, W.P. (**1988**). “Regularity and latency of units in ventral cochlear nucleus: implications for unit classification and generation of response properties,” J. Neurophysiol. **60,**1-29.