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: Sequences of clicks comprising a dominant regular interval were investigated psychophysically. The first sequence pattern consisted of one regular interclick interval followed by two random intervals. The second sequence pattern consisted of one regular interval with a randomly interspersed click, followed by a single random interval. These stimuli had a single normalized autocorrelation peak of identical height at the regular interval. They were also equated in average rate. In a pitch matching experiment, the stimuli were found to have different pitches even though the regular interval was identical. The pitch shift persisted when the stimuli were high-pass filtered at 3 or 6 kHz and low-pass noise was added. The combined effects of auditory filtering and hair-cell transduction provide a possible basis for this perceptual shift. This emphasizes the need for physiologically-based models for pitch perception.

PACS numbers: 43.66.Ba, 43.66.Hg
Date Received: April 14, 2001 **Date Accepted:** July 31, 2001

1. Introduction

The perception of the pitch of complex sounds has often been associated with periodicities present in the physical waveform. These periodicities can be detected in power spectrum-like representations of the stimulus (Helmholtz, 1877) or in time-domain representations (Schouten, 1940). More sophisticated models have since been proposed to take into account the effects of auditory filtering and spike generation at the level of primary auditory nerve fibers (Licklider, 1951; Goldstein, 1973). In spite of the higher plausibility of these later models, a surprisingly large amount of data can be explained by the waveform-based models. The autocorrelation of the waveform predicts the pitch and pitch strength of many deterministic or stochastic signals (Cariani and Delgutte, 1996; Yost 1996) and gives qualitatively similar results to all-order interspike interval distributions in populations of auditory nerve fibers (Cariani and Delgutte, 1996).

Recently, data concerning the pitch of click trains challenged the view that waveform autocorrelation would be a good first approximation for pitch models. Kaernbach and Demany (1998) studied the perception of click trains with first-order periodicity (regular intervals between successive clicks) and higher-order periodicity (regular intervals between non-successive clicks). They described two kinds of stimuli with a single peak in the waveform autocorrelation function. The first stimulus contained a regular interval followed by two random intervals. This was termed KXX. The second stimulus contained a regular interval formed by the addition of two random intervals, followed by another random interval. This was termed ABX. A simple autocorrelation of the waveforms would predict equal pitch

strength for the two stimuli. However, KXX was easier to discriminate from a random click train than ABX.

The ABX and KXX stimuli of Kaernbach and Demany (1998) were matched for a discrimination task, but they have either similar autocorrelations and different click rates or identical click rates and autocorrelation peaks that differ by one octave. In the present study, we compared KXX and ABX stimuli with identical waveform autocorrelation peaks and average rates. While listening to these new stimuli, it appeared not only that the pitch strength changed, but also that the actual pitch value shifted. Accordingly, a pitch-matching experiment was performed, and it confirmed the initial observation of a consistent pitch shift. A simulation of auditory nerve fibers (section 4) showed that the perceptual pitch shift could be encoded by differences in the interspike interval statistics in response to these two stimuli.

2. Methods

2.1 Stimuli

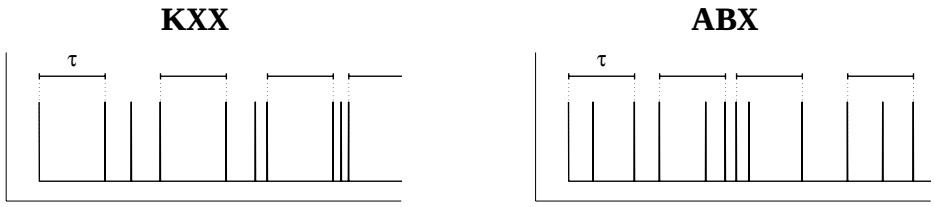

Fig. 1. Stimuli used in the experiments. The regular interval is denoted as τ. This example corresponds to the broadband condition. For a given τ, both stimuli have the same average rate of clicks.

The stimuli were generated as time domain, unipolar click-trains. For KXX, a regular interval of length τ was followed by two random intervals drawn from the uniform distribution $[0,\tau/2]$. For ABX, a regular interval of length τ was formed by the succession of two random intervals and followed by a single random interval drawn from the uniform distribution $[0,\tau]$. The distributions of the random X intervals for KXX and ABX stimuli ensure that both have the same average rate for an identical value of τ. Whereas the ABX stimulus is identical to the one defined in Kaernbach and Demany (1998), our KXX definition is different with respect to the value of K (τ instead of $\tau/2$) and the range of X ($[0,\tau/2]$ instead of $[0,\tau]$). Examples of the waveforms for the KXX and ABX stimuli as defined here are shown in Fig. 1.

Normalized long-term autocorrelation and spectra for these stimuli are shown in Fig. 2. The autocorrelation peak is identical for the two stimuli, although the background activity differs. The spectral peaks are around multiples of $1/\tau$ in both cases, except for the first peak that is skewed toward lower frequencies for KXX. This skew can be accounted for by an additional peak caused by third-order intervals. For KXX, the third-order intervals KXX, XKX, and XXK all have a mean of $3/2\tau$ and a variance of $\tau^2/24$. For ABX, the third-order intervals ABX, BXA, and XAB also have a mean of $3/2\tau$, but variances of $\tau^2/12$, $\tau^2/4$, and $\tau^2/12$, respectively. The narrow distribution of third-order intervals for KXX produces the additional spectral peak, or equivalently ripples at multiples of $3/2\tau$ in the autocorrelation.

In addition to the broadband condition, two high-pass filtered conditions were investigated where the first, irregular spectral peak was absent. Filtering was done in the frequency domain by deleting all frequency components below a given cutoff frequency (3 kHz and 6 kHz) and replacing them with components with equal amplitude but random phases drawn from a Rayleigh distribution. The result of this operation is a high-pass click-train combined with low-pass white noise, and the overall spectral envelope is flat. In this case, the autocorrelation functions for KXX and ABX are virtually identical (see Fig. 4 below) and so are the power spectra.

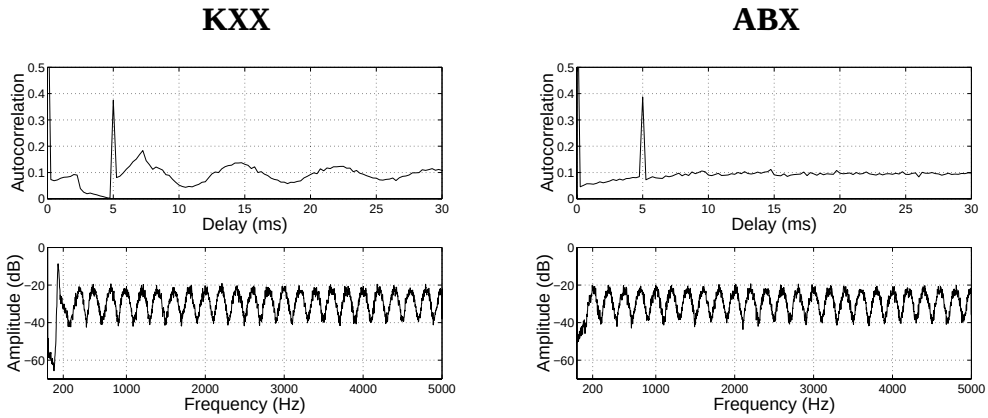

Fig. 2. Normalised autocorrelations and spectra for KXX and ABX stimuli. The stimuli are shown in the broadband case with τ=5 ms. Analysis are averaged over 50 repeats. Autocorrelation binwidth is 250 µs.

Stimuli were 409.6 ms long and included a 5 ms squared-cosine rise and fall time. They were produced on a TDT system II AP2 board and played out diotically through a DD1 digital-to-analog converter at a sampling rate of 20 kHz. Stimuli were attenuated (PA4) and lowpass-filtered at the Nyquist frequency (FT6 anti-aliasing filter) before being presented though AKG K-240-DF headphones. Attenuations were set for a presentation level of 60 dB SPL for the 3-kHz high-pass, 200-Hz periodicity condition. No attempt was made to keep loudness constant across conditions. Sound examples are available in Mm.1. to 3.

Mm. 1. The sequence [KXX-ABX-pause] is repeated three times, with different random realisations, for the broadband case and a 200-Hz periodicity. (530 Kb)

Mm. 2. As Mm.1 but with a 3 kHz high-pass cutoff (530 Kb)

Mm. 3. As Mm.1 but with a 6 kHz high-pass cutoff (530 Kb)

2.2 Procedure and listeners

A double staircase, interleaved, adaptive procedure was used (Jesteadt, 1980). Sounds were presented in pairs, and listeners had to decide which one had the higher pitch. Each trial consisted of a reference KXX stimulus with a fixed τ, to be compared with an ABX stimulus with an adaptively varied τ. Both stimuli were in the same filter condition (broadband, 3 kHz or 6 kHz high-pass). The order of KXX and ABX was randomized between trials.

Two different adaptive tracks were randomly interleaved. The lower track started with a τ 75 % longer for ABX than KXX, and a 2-down, 1-up adaptive rule was used (τ 75 % shorter and a 2-up, 1-down rule for the upper track). Eight reversals were measured for each track. The value of the ABX τ was altered by 15% in the first 2 reversals and by 5% in the last 6 reversals. Values were averaged from the last 4 reversals of each track to produce an estimate of the pitch match. The lower and upper track typically converged toward two different matches, with a separation between them of typically 8%. A single run (2 matches) was performed by each listener for each regular interval and filter condition. All conditions were randomized across subjects.

Three listeners participated in the experiment: 2 male and 1 female, aged from 24 to 37 years old. All had normal hearing (<10 dB HL). L1 is the first author. L2 and L3 performed a short training session (4 randomly selected runs) before data collection began. Listeners were seated in a double-walled, sound-attenuating booth. No feedback was provided.

3. Results

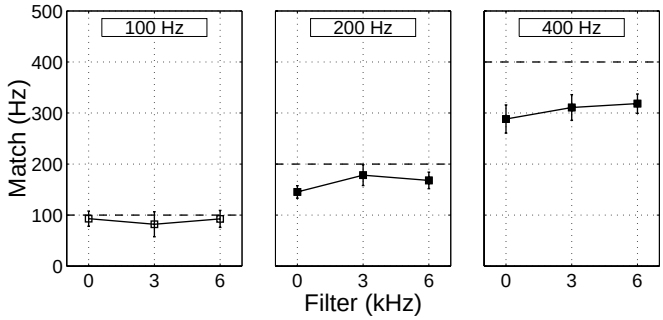

Fig. 3. Experimental results. The ABX regular interval, $1/\tau$, that produces the same pitch as KXX with a given $1/\tau$ (each panel) is shown. The horizontal dashed line indicates no pitch shift. Standard deviations are shown as vertical bars, and filled symbols indicate a significant deviation from a match to the same τ ($p<0.05$, two-tailed t-test).

Mean results for the three listeners are shown in Fig. 3. The horizontal dashed line in the graphs represent the expected results if listeners were matching the stimuli according to the length of the regular interval τ. In the condition where the periodicity was 100 Hz (τ=10ms), no significant deviation from this expected match was observed. However, in both the 200 and 400 Hz conditions (τ = 5 and 2.5 ms, respectively), a significant deviation was observed (two-tailed t-test, $p<0.05$). The match is obtained when the ABX stimulus has a longer τ than the KXX stimulus. This corresponds to a downward pitch shift for KXX compared to ABX. The shift is largest for the 400 Hz condition.

In the broadband condition, the first spectral peak differs between KXX and ABX (Fig. 2) which might explain the observed pitch shift. When high-pass filtering the stimuli, however, this cue is ruled out. In the 3-kHz high-pass condition, a pitch shift was observed even though KXX and ABX produced spectral peaks and dips at the same locations. When the high-pass cutoff was 6-kHz, all spectral peaks were unresolved for both the 200 and 400 Hz conditions (Houtsma and Smurzynski, 1990), and the shift was again observed.

4. Model

We investigated the effect of early stages of auditory processing on the autocorrelation functions (ACFs) for KXX and ABX. Stimuli were first passed through a gammatone filter centered on 4.5 kHz with an equivalent rectangular bandwidth matched to the critical band at this center frequency (Glasberg and Moore, 1990). As can be seen from the middle panels of Fig. 4, this smears the ACF but does not introduce any shift in the peak of activity. However, when the ACF is computed after gammatone filtering and half-wave rectification (HWR), activity appears at longer delays for KXX but not ABX stimuli. Further simulations varying the filter center frequency showed that this effect is observed in all auditory filters that contain unresolved harmonics of $1/\tau$. An interpretation of this finding is that the HWR non-linearity reintroduces the lower part of the spectrum for KXX within each channel and, thus, the difference in background of autocorrelation activity.

This shift in ACF will be reflected in interspike interval statistics at the level of primary auditory nerve fibers. We have implemented a simulation of an array of high-spontaneous rate auditory nerve fibers with best frequencies ranging from 100 to 8500 Hz. The model had 60 frequency channels, and each channel comprised three stages. The first stage was a gammatone filter; the second stage consisted of a simulated inner hair-cell (Meddis and O'Mard, 1997). In the final stage, spikes were generated using a Poisson process with a 1-ms dead time. The amplitude was set to 60 dB SPL. Fifty fibers per channel were simulated, and 100 repeats with fresh stimuli were averaged for each condition.

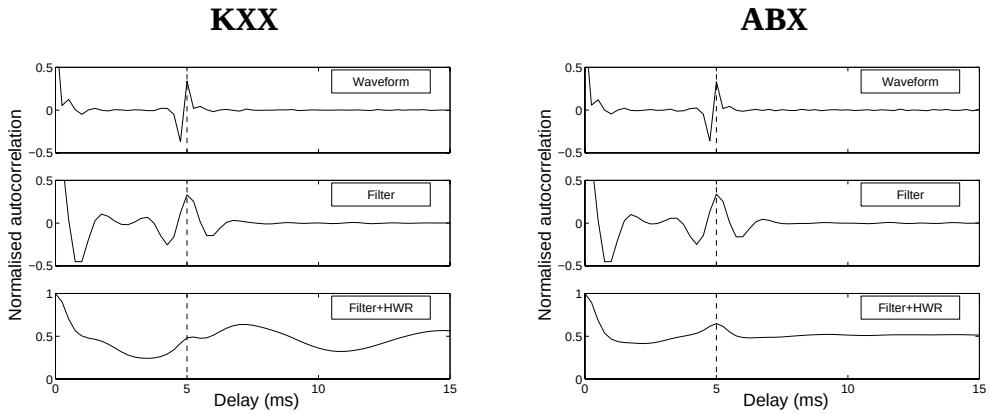

Fig. 4. Long term autocorrelation for KXX and ABX stimuli following different processing stages. Binwidth is 250 μs, 50 repeats. The upper row is the autocorrelation of the waveform for a 3 kHz filter condition and a 5 ms regular interval. The middle row is the autocorrelation following filtering with a single gammatone filter with a center frequency of 4.5 kHz. The lower row is the autocorrelation after the same filter followed by half-wave rectification (HWR).

All-order and first-order interspike interval distributions were calculated from the individual spike trains. The results of this process are shown in the first two columns of Fig. 5. In both the first- and all-order interval analysis, there is a peak in the interval distribution at 5 ms for the ABX stimulus. In contrast, the peak is broader and shifted towards longer intervals for the KXX stimulus. Both peaks are relatively small, but the pitch is also rather weak. The shift agrees qualitatively with the perceptual data.

5. Discussion

The pitch of complex tones, if not the pitch strength, is usually well predicted by the main peak of the waveform autocorrelation function. The results presented in this study are, to our knowledge, the first example of a large pitch shift between two sounds where the main peaks in the waveform autocorrelation were identical. Computer simulations suggest that this pitch shift originates from the combined effects of auditory filtering and neural transduction, and is reflected in interspike interval statistics of auditory nerve fibers. This means that models acting after a simulation of auditory nerve transduction will probably predict the pitch shift (Licklider, 1951; Goldstein, 1973), but waveform-based models will not.

We acknowledge that in the highest filter condition (6 kHz), the pitch cue must have been weak. It would probably not be sufficient to support musical melodies (Pressnitzer *et al.*, 2001). However, matches could be made consistently by the three listeners in the 6-kHz cutoff condition, so a periodicity cue must have remained.

The filtering used in the stimulus generation was sharp, so it is also possible that an edge pitch was created. However, this edge pitch should have been identical for KXX and ABX and is unlikely to explain the present results. The introduction of low-pass noise at a level equal to that of the stimulus also precludes an interpretation of the results in terms of distortion products at the level of the basilar membrane.

In the study by Kaernbach and Demany (1998), the stimuli were designed to test first-order versus all-order interval hypotheses in the encoding of pitch. The present data and model simulations suggest that stimulus *interclick* order should be distinguished from auditory nerve *interspike* order, because of the stochastic nature of spikes generation. The pitch shift we observed could to be due to third-order interclick statistics, which, if correct, means that interclick intervals beyond first-order can influence pitch. The effect can, however, be predicted by either first- or all-order interspike statistics. For the present data at least, the issue of the interspike order used in the neural encoding of pitch is unresolved.

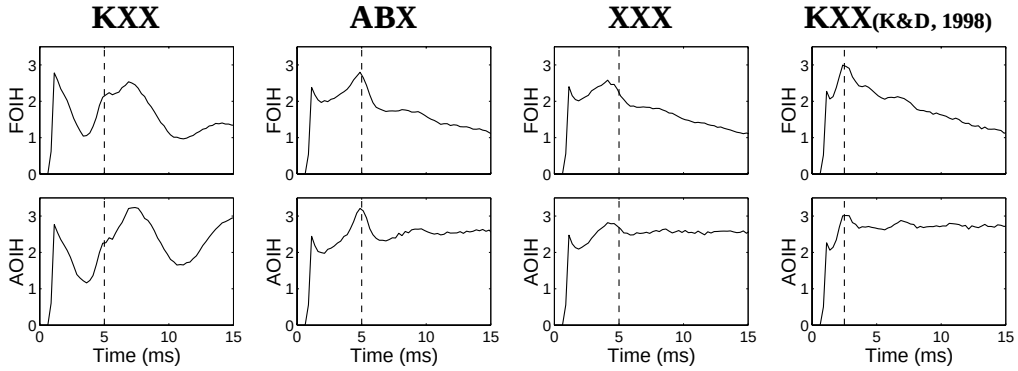

Fig. 5. First- and all-order interspike interval histograms (FOIH and AOIH, respectively) from a simulated auditory nerve fiber array. Binwidth is 250μs. Stimuli have a 6 kHz cutoff and τ is equal to 5 ms. From left to right: KXX and ABX as defined in section 2.1; a random click train XXX with intervals uniformly drawn from [0,τ]; a KXX stimulus as defined in Kaernbach and Demany (1998) with a regular interval at τ/2 and Xs drawn from [0,τ].

The simulations also provide a possible basis for the discrimination data of Kaernbach and Demany (1998). The last three columns of Fig. 5 show the model output for the stimuli used in their study. The backgound activity for KXX is now more similar to ABX, but KXX produces a peak at τ/2. The XXX, random stimulus does not produce a flat interval distribution. It rather displays a peak near τ. Remember that only the first-order stimulus interclick distribution is uniform for XXX; the second-order interclick distribution has a triangular shape centered on τ. This is reflected in the simulated interspike distributions for XXX, which are peaky and more similar to ABX than to KXX. The simulations would be consistent with the observation that both ABX and XXX possess a comparable weak pitch and are thus difficult to discriminate (Kaernbach, personal communication). Such a hypothesis would need to be tested by further psychophysical and physiological investigations.

Acknowledgments

We would like to thank Christian Kaernbach, Bill Yost, and Martin McKinney for helpful comments on a previous version of the manuscript. This work was supported by the Centre National de la Recherche Scientifique (CNRS) and the Wellcome Trust.

References

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.

Glasberg, B.R., and Moore, B.C.J. (**1990**). "Derivation of auditory filter shapes from notched noise data," Hear. Res. **47**, 103-138.

Goldstein, J.L. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496-1515.

Helmholtz, H.L.F. (**1877**). *On the sensation of tone as the physiological basis for the theory of music*. 2nd edition (1954), translated A.J. Ellis (1885) from German 4th Ed. New York: Dover.

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

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.

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., Patterson, R.D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch,". J. Acoust. Soc. Am. **109**, 2074-2084.

Schouten, J.F. (**1940**). "The residue and the mechanism of hearing," Proc. K. Ned. Akad. Wet. **43**, 991-999.

Yost, W..A. (**1996**). "Pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **100**, 3329-3335.