The temporal dynamics of pitch perception and what they reveal about processing mechanisms

KATRIN KRUMBHOLZ AND NICHOLAS R. CLARK

MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, UK

Hearing impairment can severely restrict the ability to communicate through speech in noisy en-vironments. One of the most important cues for segregating wanted from unwanted sounds is temporal regularity, or harmonicity in the frequency domain, giving rise to the perception of pitch. However, pitch is a strong segregation cue only in the low-frequency region, where harmonic components are spectrally at least partially resolved. In contrast, spectrally unresolved pitch, produced by high-frequency sounds, is a much weaker segregation cue. This and other differences led to the assumption that resolved and unresolved pitch are processed by different mechanisms – a spectral one for resolved pitch and a temporal one for unresolved pitch. The aim of this study was to test this assumption by measuring the temporal dynamics of pitch perception in the resolved and unresolved regions.

For that, the threshold for the detection of a gap in the autocorrelation function of iterated rippled noise was measured as a function of the pitch value and the spectral region of the stimulus. The minimum detectable gap duration would be expected to be largely independent of pitch value, if pitch were processed spectrally. Contrary to this expectation, we found that the minimum detect-able gap duration decreases with increasing pitch value in an approximately reciprocal way, suggesting that pitch is processed temporally even in the low-frequency region. The experimental data are compared to predictions from models of auditory temporal processing.

INTRODUCTION

Pitch or harmonicity is one of the strongest cues for sound source segregation, a function that is crucial to our ability to listen in noisy environments (e.g., Darwin and Hukin, 2000). However, pitch is a strong segregation cue only in the low-frequency region, where harmonics are at least partially resolved by the cochlear filters (Carlyon, 1996; Vliegen *et al.*, 1999). This led to the assumption that the pitch of resolved harmonics is processed by a spectral mechanism, which evaluates the peaks in the distribution of activity across the tonotopic array (for review, see de Cheveigne, 2005). The aim of this study was to test this assumption by measuring the temporal dynamics of pitch processing in the low-frequency region.

METHODS

For that, we measured the threshold for detecting a temporal gap in the pitch strength of broadband tonal sounds. In monaural and binaural hearing, gap detection has been used as a measure of temporal resolution. The sounds were rippled and iterated rippled

Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007). T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen (Eds.).
ISBN: 87-990013-1-4. Print: Centertryk A/S.

noises (RN, IRN). RN is produced by delaying a copy of a noise by *D* ms and adding it back to the original. The process is iterated *N* times to produce IRN. The perceived pitch corresponds to the reciprocal of the delay, *D*. To introduce a pitch gap in RN, a portion of the delayed noise was replaced by an independent noise sample of the same intensity. In IRN, a portion of the entire stimulus was replaced by noise.

IRN was produced with $N = 8$ iterations. In one condition, the pitch gap detection threshold was measured for broadband RN and IRN (lowpass-filtered at 5 kHz) as a function of delay ($D = 1$, 2, 4, 8 and 16 ms). In another condition, pitch gap detection was measured for bandpass-filtered IRN (1-kHz bandwidth) as a function of the filter centre frequency (0.5, 0.75, 1, 1.5, 2.5 and 3.5 kHz) and a single delay of 8 ms. The stimuli were generated digitally with a 25-kHz sampling rate and a 24-bit amplitude resolution using Matlab and TDT System 3. Pitch gap detection thresholds were measured with an adaptive 2A2IFC procedure using a 3-down 1-up rule.

RESULTS

If the pitch of resolved harmonics were processed spectrally, the pitch gap detection threshold might be expected to be determined by the duration of the cochlear-filter impulse responses and thus be independent of pitch. If, on the other hand, pitch were processed temporally, the pitch gap threshold would be expected to increase with decreasing pitch, or increasing delay, D. Figs 2 and 3 show that threshold did indeed increase with increasing delay. For IRN, pitch gap threshold increased from 0.76 ms at D = 1 ms to 9 ms at D = 16 ms, on average; for RN, threshold increased from 3.9 to 312 ms for the same range of delays. The function relating pitch gap threshold to delay, D, was approximately linear when plotted in log-log coordinates and was best approximated by $\text{Thr} = 0.7 \cdot D^{0.9}$ for IRN and $\text{Thr} = 3.3 \cdot D^{1.6}$ for RN.

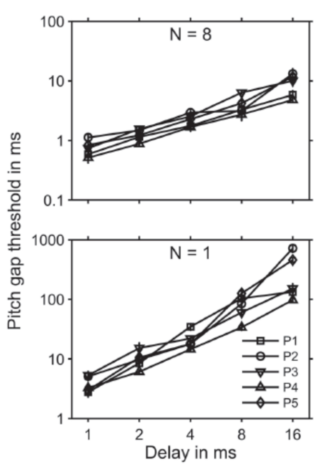

Fig. 1: Individual pitch gap detection threshold for IRN (A) and RN (B) as a function of delay, *D*, plotted in log-log coordinates. Different symbols denote different participants.

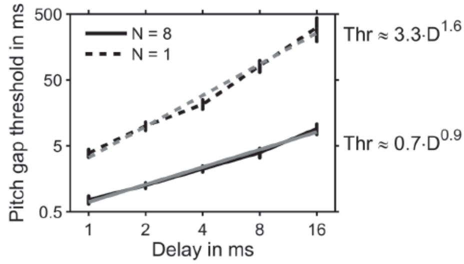

Fig. 2: Average pitch gap detection threshold as a function of delay (black lines) and log-log regression lines (grey) for IRN (solid lines) and RN (dashed lines).

Fig. 3 shows that pitch gap detection performance decreased with increasing filter frequency above 0.75 kHz. On average, the pitch gap threshold increased from 4.3 ms at 0.5 kHz to 69 ms at 3.5 kHz. There was, however, considerable inter-individual variability in the effect of filter frequency: In two out of five participants, gap threshold increased sharply between 1 and 1.5 kHz and plateaued between 2.5 and 3.5 kHz; in two participants, threshold kept increasing even at the highest filter frequencies, and in one participant, threshold increased roughly linearly (in log-log coordinates) over the entire range of frequencies. The increase in gap threshold is probably related to a decrease in pitch strength with increasing filter frequency, which may be due to the loss of phase locking towards higher frequencies (Yost *et al.*, 1998; Hall *et al.*, 2003).

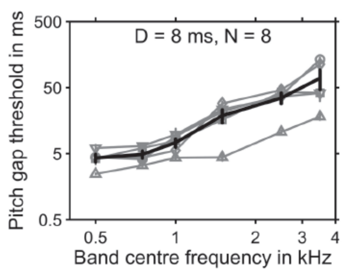

Fig. 3: Average (black line) and individual (grey lines) pitch gap detection threshold for bandpass-filtered IRN (D = 8 ms, N = 8) as a function of the filter centre frequency.

MODELLING

While a spectral explanation of the increase in the pitch gap detection threshold with increasing delay cannot be firmly excluded until a spectral model has been tested, the current results seem more compatible with a temporal explanation. Most temporal pitch models are based on some form of autocorrelation (Licklider, 1951). The running autocorrelation function of a stimulus is obtained by multiplying the original with a delayed version of the stimulus, and then submitting the resulting cross product to a leaky integration (e.g., Meddis and O'Mard, 1997). The time constant of the integration process has previously not been a function of the autocorrelation lag (for review, see Wiegrebe *et al.*, 2000). We implemented a physiologically motivated version of the running autocorrelation model by (i) bandpass filtering the stimulus between 0.45 and 8 kHz to simulate the middleear filtering, (ii) passing it through a 64-channel gammatone filterbank with filter frequencies evenly spaced on an ERB scale between 0.1 and 4 kHz to simulate cochlear filtering, (iii) halfwave rectifying, square-law compressing and lowpass filtering (F_c = 1 kHz, 4th order) the signal to simulate the auditorynerve activity pattern, and (iv) applying a running autocorrelation to the resulting simulated activity pattern, using either a fixed integration time con-stant, τ, or one that varied linearly with delay.

Applying the model to the IRN stimulus revealed that, with a fixed τ, the model is unable to predict the increase in gap threshold with increasing delay. With τ fixed, the predicted threshold function is v-shaped and the residual deviation between predicted and measured thresholds exceeds 3 ms, irrespective of the value of τ (Fig. 45A). In contrast, with τ a linear function of delay, the data are predicted reasonably well, apart from those for the smallest delays (Fig. 4B). It is unlikely, however, that a linear relationship between τ and delay would also be able to account for the RN data, because the gap threshold for RN increased overproportionately with delay.

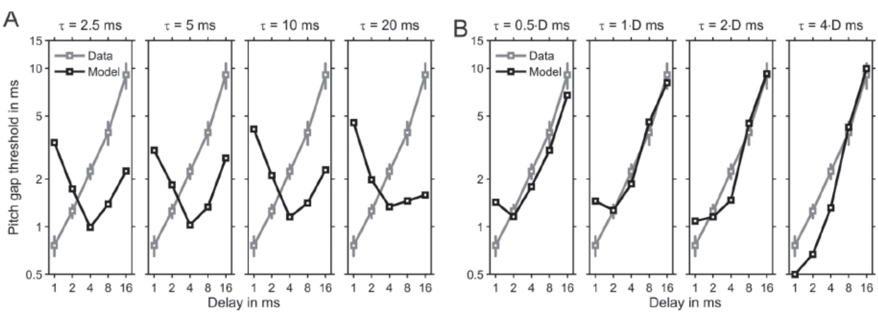

Fig. 4: Model simulations of IRN data for a fixed time constant, τ, (A) and for τ varying linearly with delay, D, (B). Different panels show the results for different parameters (see panel titles).

CONCLUSIONS

The current study showed that the duration required to detect a gap in the autocorrelation function of a regular-inter noise stimulus, such as RN or IRN, increases with decreasing pitch. The results suggest that pitch is processed by a temporal mechanism which analyses serial correlation in the stimulus waveform and integrates the longer, the longer the correlation time period analysed. The finding that gap threshold increased over-proportionately with delay for the RN stimulus, which has a lesser correlation than the IRN stimulus suggests that the integration time constant of the pitch processing mechanism may also depend on stimulus statistics. Gap thresholds also increase with increasing spectral region. Initial modelling results suggest that this is due to the progressive loss of phase locking towards higher frequencies (data not shown).

REFERENCES

Carlyon, R. P. (**1996**). “Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker,” J. Acoust. Soc. Am. **99**, 517-524.

Darwin, C. J., and Hukin, R. W. (**2000**). “Effectiveness of spatial cues, prosody, and talker characteristics in selective attention,” J. Acoust. Soc. Am. **107**, 970-977.

de Cheveigné, A. (**2005**). “Pitch perception models, Pitch: Neural Coding and Perception, Edited by C. Plack and A. Oxenham, New York, Springer Verlag.

Hall, J. W. 3rd, Buss, E., and Grose, J.H. (**2003**). “Modulation rate discrimination for unresolved components: temporal cues related to fine structure and envelope,” J. Acoust. Soc. Am. **113**, 986-993.

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

Vliegen, J., Moore, B. C., and Oxenham, A. J. (**1999**). “The role of spectral and periodicity cues in auditory stream segregation,measured using a temporal discrimination task,” J. Acoust. Soc. Am. **106**, 938-945.

Yost, W. A., Patterson, R., and Sheft, S. (**1998**). “The role of the envelope in processing iterated rippled noise,” J. Acoust. Soc. Am. **104**, 2349-2361.

Wiegrebe, L., Hirsch, H. S., Patterson, R. D., and Fastl, H. (**2000**). “Temporal dynamics of pitch strength in regular-interval noises: effect of listening region and an auditory model,” J. Acoust. Soc. Am. **107**, 3343-3350.