Research paper

Temporal integration in absolute identification of musical pitch

I-Hui Hsieh, Kourosh Saberi *

Department of Cognitive Sciences, The Center for Cognitive Neuroscience, University of California, Irvine, CA 92697-5100, United States

Abstract

The effect of stimulus duration on absolute identification of musical pitch was measured in a single-interval 12-alternative forced-choice task. Stimuli consisted of pure tones selected randomly on each trial from a set of 60 logarithmically spaced musical note frequencies from 65.4 to 1975.5 Hz (C2–B6). Stimulus durations were 5, 10, 25, 50, 100, and 1000 ms. Six absolute-pitch musicians identified the pitch of pure tones without feedback, reference sounds, or practice trials. Results showed that a 5 ms stimulus is sufficient for producing statistically significant above chance performance. Performance monotonically increased up to the longest duration tested (1000 ms). Higher octave stimuli produced better performance, though the rate of improvement declined with increasing octave number. Normalization by the number of waveform cycles showed that 4 cycles are sufficient for absolute-pitch identification. Restricting stimuli to a fixed-cycle waveform instead of a fixed-duration still produced monotonic improvements in performance as a function of stimulus octave, demonstrating that better performance at higher frequencies does not exclusively result from a larger number of waveform cycles. Several trends in the data were well predicted by an autocorrelation model of pitch extraction, though the model outperformed observed performance at short durations suggesting an inability to make optimal use of available periodicity information in very brief tones.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Pitch; Music; Memory; Temporal integration; Psychophysics

1. Introduction

This is a brief report on the ability of absolute-pitch (AP) musicians to identify the pitch of short-duration pure tones without an acoustic referent. It was motivated by two questions. First, what is the minimum stimulus-integration-time required to encode pitch for retrieval from long-term memory? Theories of AP encoding, supported by psychophysical and neuroimaging evidence have proposed an in-built association between stored pitch representations and linguistic processes in facilitating the retrieval and labeling of pitch (Levitin, 1994; Zatorre, 2003; Zatorre et al., 1998; Deutsch et al., 2006). It is *a priori* unknown whether involvement of cognitive mechanisms requires a minimally

sustained stimulus for invoking the hypothesized conditional association between pitch and linguistic representations beyond that which is required for simple sensory discrimination.

Second, how does pitch salience decline as a function of duration given the uncertainty trade-off between temporal and spectral resolution (Gabor, 1947)? All previous studies of the effects of duration on pitch perception have either used subjective measures (Doughty and Garner, 1947) or have in some form employed frequency discrimination tasks, i.e., the time-derivative of pitch perception (Doughty and Garner, 1948; Campbell, 1963; Ritsma et al., 1966; Pollack, 1967; Henning, 1970; Metters and Williams, 1973; Moore, 1973; Patterson et al., 1983; Freyman and Nelson, 1986; Beerends, 1989; Robinson and Patterson, 1995; Gockel et al., 2005). While studies of frequency discrimination have provided valuable information on distinguishing between temporal and spectral mechanisms of pitch processing (Moore, 1973; Gockel et al., 2005) sensory

---

Abbreviation: AP, absolute pitch
* Corresponding author. Tel.: +1 949 824 6310; fax: +1 949 824 2307.
E-mail address: saberi@uci.edu (K. Saberi).

discrimination tasks are inherently insensitive to identifying bias effects, and hence, cannot be exclusively relied on to study pitch salience. A constant pitch shift resulting potentially from changes in stimulus duration or peripheral excitation patterns will remain undetected in a discrimination task since such tasks are based on relative cues.

The current study examines the ability of AP musicians to identify the pitch of pure tones in a single-interval 12-alternative forced-choice task for tone durations from 5 ms to 1 s. No feedback, reference sounds or practice trials were used at any point during the experiment. To our knowledge, this is the first study of its type to *directly* assess pitch salience and identity as a function of duration without referent sounds.[1] A number of interesting findings have emerged from this study which is contrasted to predictions of an autocorrelation model of pitch extraction. As will be discussed, several trends in the data are well predicted by the model, including a monotonic decline in pitch-identification ability with decreased stimulus duration, improvements in performance with increasing stimulus frequency, and diminishing *rate* of improvement with increasing octave number from which stimuli are selected. Autocorrelation, however, predicts better than observed performance for the shortest-duration stimuli at the highest octaves, suggesting an inability to make optimum use of available periodicity cues.

2. Experiment I: absolute-pitch identification as a function of stimulus duration

2.1. Materials and methods

2.1.1. Subjects

Six AP musicians (five females) served as subjects. They were recruited from the UCI campus community through flyers and announcements in music-performance classes. Their ages ranged from 19–27 (mean = 22) and all had begun formal music training between the ages of 4 and 6 years. They were paid an hourly wage for participation. The protocol for experiments on human subjects was approved by the University of California's Institutional Review Board.

2.1.2. Screening

AP ability was verified through a screening test using pure tones and piano notes in a single-interval 12-alternative forced-choice task. Pure tones were 1000 ms in duration with 100 ms rise-decay times. On each trial of a 50-trial run, a pure tone (or piano note) was selected randomly (with replacement) from a set of 60 musical note frequen-

cies in the range C2–B6 (65.4–1975.5 Hz, equal-tempered scale) and presented with the constraint that successive notes were at least 2 octaves ± 1 semitone apart. The level of a 1 kHz tone was set to 70 dB SPL (A weighted). No other level changes were made and thus the levels of other stimuli were set by this base amplitude and the headphone transfer function. All stimuli in the tested frequency range were clearly audible and at a comfortable listening level. Subjects responded on a graphical user interface by pressing one of 12 pushbuttons labeled with the 12 musical notes $(C, C\#, D, \ldots)$. No practice trials were allowed and no feedback was provided at any point during the entire study. A criterion of 90% accuracy for piano notes and 75% for pure tones (chance = 8.3%) was used to qualify a subject as AP (Baharloo et al., 1998; Miyazaki, 1989; Ross et al., 2004).

2.1.3. Stimuli and procedure

Pure tones were generated at a sampling rate of 44.1 kHz using Matlab software (Mathworks) and presented diotically through Bose headphones (model QCZ, TriPort) in a double-walled steel acoustically isolated chamber (Industrial Acoustics Company; interior dimensions $1.8 \times 1.9 \times 2.0$ m).[2] Piano notes, used only in the screening task, were digitally recorded from a 9-foot Steinway grand piano at UCI's Music Department. Notes were recorded at a sampling rate of 44.1 kHz using a 0.5-in. microphone (Brüel & Kjær Model 4189), a conditioning amplifier (Nexus, Brüel & Kjær), and a 16-bit A–D converter (Creative Sound Blaster Audigy 2ZS). Spectral analysis of the recorded notes confirmed that the piano was in tune. Pure tones used in the main experiment had durations of 5, 10, 25, 50, and 100 ms with linear rise-decay times of 1 ms and a zero starting phase.[3] Tone frequencies were randomly selected on each trial from the set described above. The total energy of the tone was held constant and thus the stimulus power was halved when duration was doubled.

The experiment was run in a block design with stimulus duration fixed within a run. Each subject completed six runs of 50-trials for each of five durations. Subjects were required to respond within 2 s of stimulus presentation.

---

[1] The strength of pitch or its "salience" depends on a number of factors, one of which is duration. Other factors, such as spectral edge, iteration pattern, dichotic coherence, and spectral composition are additional important determinants of pitch salience (Wightman, 1973; Terhardt, 1979; Terhardt et al., 1986; Parncutt, 1994; Bilsen and Goldstein, 1974; Yost, 1996a,b).

[2] We selected pure instead of complex tones partly because absolute identification of the pitch of harmonic complexes may be based on the pitch of individually resolved harmonics that are related to the fundamental by an octave (i.e., 1st and 3rd harmonics are the same musical note as the fundamental), and because our main goal was not to distinguish between mechanisms of encoding virtual pitch and pitch of individual components, but to determine the minimum duration required for absolute identification of musical pitch.
[3] The use of a set of 60 frequencies in a 12-alternative no-feedback design makes it impossible to base judgments on non-pitch cues. Nonetheless, to verify that identification is unaffected by phase and relative level, we measured performance of three AP subjects for a 10 ms pure tone whose phase was randomized on each trial by 2π and whose level was randomized by 12 dB. All subjects performed above chance with proportion corrects of 0.24 and 0.23 for the fixed and randomized conditions respectively (chance = 0.083). A *t*-test comparison showed no significant effect of phase/level randomization on absolute-pitch identification ($t(2) = 2.21$, ns).

Failure to respond was scored as an error (though this rarely occurred). The task and response interface were identical to that used for screening. Prior to each run, one of the five experimental conditions was randomly selected for that run. This procedure was continued until every condition was run at least once before a second set of runs began, and until a total of six runs per duration per subject was completed.

2.2. Results

Top panel of Fig. 1 shows averaged results from six subjects as a function of stimulus duration. Filled symbols represent proportion correct identification of the target note, with the dashed horizontal line representing chance performance equal to 0.083. Open symbols show proportion correct identification to within one semitone. We show this latter measure both because it provides additional information on the extent of pitch misidentification, and because it is a commonly reported measure for AP musicians. Chance performance for this measure is 0.25. For durations up to 100 ms, each point is based on 1800 pitch estimates. The 1 s condition is from the screening procedure (300 estimates). Error bars show ±1 standard deviation.

There is a clear monotonic increase in performance as a function of duration, with an initial rapid improvement from 5 to 25 ms followed by a more gradual increase. Even the shortest-duration stimuli produced performance significantly above chance (filled circle at 5 ms; $t(5) = 5.19$, $p < 0.01$; open circle $t(5) = 4.99$, $p < 0.01$). Subjects also showed an improvement in performance when tone duration was increased from 100 to 1000 ms suggesting integration times exceeding 100 ms ($t(5) = 3.38$, $p < 0.05$ filled symbols; $t(5) = 3.63$, $p < 0.05$ open symbols).

Middle and bottom panels of Fig. 1 correspond respectively to filled and open circles of the top panel, and show joint effects of duration and musical octave on pitch-identification performance (2nd through the 6th octave). Each curve is labeled with its appropriate octave number. Curves associated with successive octave numbers are shown with alternating solid and dashed lines to facilitate visual comparison. These data show that for a given duration, higher octave stimuli produce better identification performance. A 2-way repeated-measures ANOVA on the data of the middle panel showed a significant effect of octave ($F_{4,20} = 99.04$, $p < 0.001$), duration ($F_{4,20} = 355.27$, $p < 0.001$), and interaction between duration and octave ($F_{16,80} = 18.88$, $p < 0.001$). Similar results were obtained for the data of the bottom panel, with significant effects of octave ($F_{4,20} = 94.05$, $p < 0.001$), duration ($F_{4,20} = 361.87$, $p < 0.001$), and their interaction ($F_{16,80} = 16.94$, $p < 0.001$).

To determine if there were pitch shifts (or biases) we analyzed semitone errors from the entire dataset and found that subjects were as likely to make a 1-semitone error above a target note (probability of 0.5083) as they were to make a semitone error below the target. Analyzed as a function of stimulus duration, the posterior probabilities
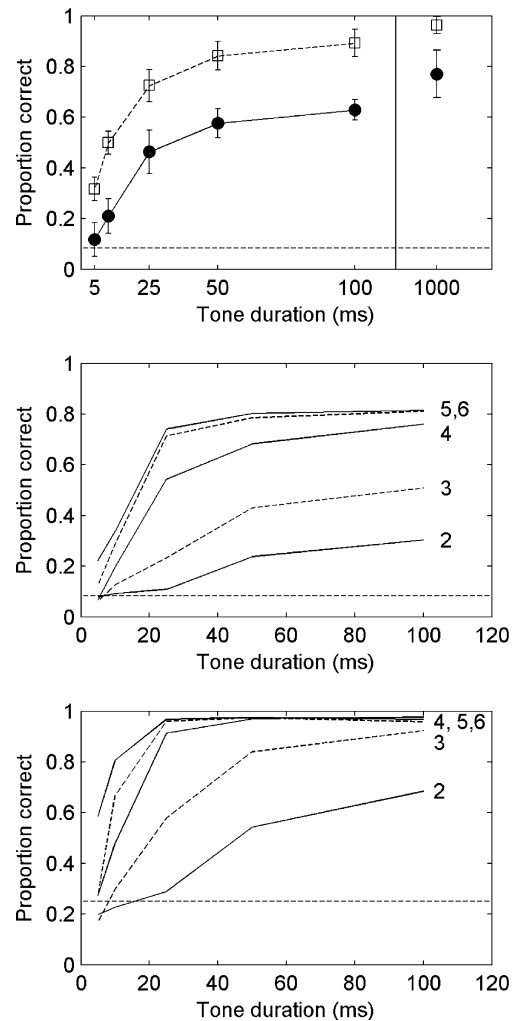

Fig. 1. Absolute-pitch identification as a function of stimulus duration. Top panel shows averaged data from six AP musicians (1800 trials per point). Filled symbols represent proportion-correct identification of a note in a single-interval 12-alternative forced-choice task without feedback (chance = 0.083 shown as the dashed horizontal line). Open symbols show correct identification to within a semitone (chance = 0.25). Error bars are ±1 standard deviation. Middle panel shows performance for exact identification of musical notes organized by the octave number from which stimuli were selected (corresponding to the filled symbols in the top panel). Bottom panel show proportion correct identification of a target note to within one semitone sorted by octave number (corresponding to the open symbols in the top panel).

of making a semitone error above the target note for stimulus durations of 5, 10, 25, 50, and 100 were 0.54, 0.56, 0.52, 0.46, and 0.46, respectively, possibly indicating a marginal upward pitch shift as stimulus duration is decreased.

To determine if better pitch identification at higher octaves is related to the number of waveform cycles (i.e., periods) in a fixed-duration tone, performance is plotted in the top panel of Fig. 2 as a function of number of cycles. Filled and open symbols show, respectively, proportion-correct note identification and correct identification to within a semitone. The number of cycles is defined as the product of frequency and duration, rounded to the nearest integer (hence 0 cycles represents a waveform containing
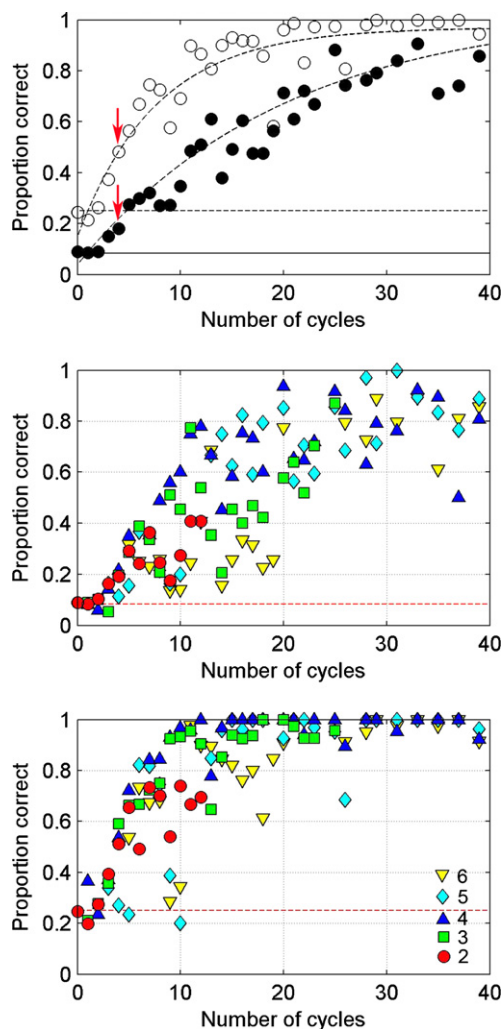
Fig. 2. Absolute-pitch identification as a function of the number of waveform cycles (periods) rounded to the nearest cycle. Top panel: averaged data from six subjects. Open and filled symbols represent exact identification and identification to within a semitone, respectively. The curves are least-square fits of a modified-exponential function from 9000 pitch estimates (see text). Solid and dashed horizontal lines represent chance performance for filled and open symbols respectively. Middle panel: same data as the filled symbols of the top panel (exact identification), with the data re-plotted by octave number. Bottom panel: same data as the open symbols of the top panel (open symbols) re-plotted by octave number (see legend).

less than half a cycle). The curves are least-square exponential fits $p = \alpha - e^{\beta(n-\mu)}$ where n is the number of cycles and α, β and μ are free parameters. The solid and dashed horizontal lines represent chance performance for the two conditions ($p = 0.083$ and 0.25). Note that the best-fitting curves do not cross chance performance at zero cycles (but at approximately 1 cycle) suggesting that pitch information does not begin to accumulate until at least one waveform cycle is completed. A 2-way repeated measures ANOVA on the data of the top panel showed a significant effect of number of cycles ($F_{32,64} = 28.71$, $p < 0.001$), condition ($F_{1,2} = 111.69$, $p < 0.01$), and their interaction ($F_{32,64} = 2.27$, $p < 0.01$).

Middle panel of Fig. 2 shows performance as a function of the number of waveform cycles and musical octave (corresponding to the filled circles of the top panel). Bottom panel shows performance accuracy to within one semitone (corresponding to the open circles of the top panel). Dashed lines represent chance performance. Contrary to the effects shown in Fig. 1, there does not appear to be an octave effect once data are normalized by number of cycles. A repeated-measures ANOVA on conditions for which data from all octaves were available (cycles 5, 6, 7, 8, 10) was not statistically significant ($F_{4,16} = 2.94$, $p = 0.16$; middle panel). The absence of an octave effect after normalization by number of cycles may be a result of the high variability observed in the data even though each panel shows data from 9000 pitch estimates.

t-Test analyses on the data of the top panel of Fig. 2 showed that the first point at which performance is statistically significantly above chance is 4 cycles ($t(5) = 5.14$, $p < 0.01$, open circles; and $t(5) = 3.76$, $p < 0.05$, filled circles). These points are marked by arrows. Scores for all cycle numbers above 4 (i.e., 5–40) were also statistically significantly above chance. t-tests at 3 cycles approached significance (p values of 0.10 and 0.06, respectively for open and filled circles). Most studies that have investigated the effects of duration on frequency or pitch-contour discrimination (Sekey, 1963; Henning, 1970; Ronken, 1971; Moore, 1973; Freyman and Nelson, 1986; Patterson et al., 1983; Robinson and Patterson, 1995) suggest that subjects require a minimum of 4–10 cycles of a sinusoid for detection of a *change* in pitch or melody. The current study places this limit for pitch *identification* in the lower range of estimates obtained from frequency or melody discrimination tasks.

To our knowledge, only two studies have reported pitch extraction from 0.5- or 1-cycle tones. Sipovsky et al. (1972) reported a 30-Hz frequency discrimination threshold for a 0.5-cycles 1.5-kHz pure tone. Mark and Rattay (1990) reported frequency discrimination thresholds for 1-cycle tones in the order of one semitone. However, results from both these studies are likely based on spectral-edge or bandwidth pitch artifacts.[4] Robinson and Patterson (1995) used complex stimuli consisting of glottal vowels with a fundamental at one of only four musical notes and found chance performance at 1 cycle of the complex, but above-

---

[4] Sipovsky et al.'s (1972) result is not likely based on pitch extraction from a periodicity cue but rather from a spectral-edge cue associated with the stimulus duration and possibly timbre cues. Given a fixed number of cycles, changing the stimulus frequency will results in a change in duration and a detectable bandwidth- or spectral-edge pitch associated with the burst duration (i.e., 1/d) and separate from fine-structure periodicity. This is particularly true for very brief tones. We verified this by listening to trains of partial-cycle random-phase tones with ascending or descending frequencies and found that a pitch change is easily identifiable even for 0.1 cycle waveforms. In addition, Sipovsky et al. used fixed-phase 0.5-cycle tones which in a 2IFC task using a single tone frequency (1.5 kHz) will likely allow discrimination based on timbre. Mark and Rattay (1990) also used zero-phase fixed-cycle tones in a 2IFC discrimination paradigm.

chance timbre identification at 1 cycle. They also reported no octave effect (for octaves 2–4) for complex sounds with a fixed number of cycles, but did report a decline in performance for octave 1 which they attributed to the lower limit of pitch perception. As will be described, our findings from Experiment II contradict their conclusions.

3. Experiment II: absolute-pitch identification from a fixed-cycle waveform

An important question is whether better pitch identification as a function of octave number results from the larger number of cycles at a fixed-duration. This explanation has previously been proposed to account for the ability to discriminate "faster" the virtual-pitch of a missing-fundamental complex tone relative to its fundamental (Patterson et al., 1983). Experiment II was designed to determine if better pitch identification at higher octaves is exclusively dependent on the number of cycles or will also be observed when cycle number is constant. The experiment was partly motivated *post hoc* by results of model simulations described in Section 3.3. The model predicted better performance at higher octaves, even when normalized by number of waveform cycles. Results from Experiment I, in contrast, had not shown a significant octave effect when data were plotted as a function of the number of cycles (Fig. 2; middle and bottom panels). This may have possibly been due to the small sample size at each cycle number and octave (∼70). In Experiment II we examined absolute-pitch identification using a 5-cycle tone. We selected 5 cycles because results of Experiment I showed that it produces performance that is clearly above chance but also below asymptote, increasing the likelihood of observing a range of performances by avoiding ceiling and floor effects.[5]

3.1. Methods

Three AP subjects who had served in Experiment I were employed for Experiment II. The apparatus and procedures were the same as those used in Experiment I except for the following. First, the number of waveform cycles was fixed at 5. Second, to eliminate frequency-dependent differences in signal-to-noise ratio resulting from the headphone transfer function, we digitally filtered each tone with the inverse of the headphone transfer function measured using a pair of complementary 512-point Golay codes (Golay, 1961; Zhou et al., 1992). Third, the total energy of the tone was held constant to correct for the decrease in energy resulting from shorter-duration higher-frequency tones given a fixed number of cycles. Fourth, the level of each tone on each presentation was additionally roved by

12 dB. Fifth, no rise-decay ramp was imposed on the stimuli. Sixth, the starting phase was randomized. Each subject completed 15 runs of 100 trials each.

3.2. Results

Fig. 3 shows results of this experiment for the 3 subjects and their averaged performance (bottom-right panel). Data are shown for both exact identification of musical note frequency and identification to within a semitone. Results show a clear monotonic effect of stimulus frequency with higher-octave stimuli producing significantly better performance. Thus, better absolute-pitch identification at higher frequencies is not exclusively a result of the larger number of waveform cycles. An ANOVA test showed a significant effect of frequency for exact pitch identification ($F_{4,10} = 8.97$, $p < 0.01$) as well as for identification to within a semitone ($F_{4,10} = 11.21$, $p < 0.01$). We also examined if level randomization (12 dB rove) had an effect on performance by pooling all trials across frequencies and subjects for a given level (12 categories). To calculate a correlation coefficient we rounded the decibel values to the nearest lower integer. The correlation coefficient between level (-12 to 1 dB) and proportion correct pitch identification was $r = -0.2275$. This correlation was not significantly different than zero, $t(11) = -0.774$, $p = 0.455$.

4. Discussion

Moore (1973) in a now classic paper demonstrated that as tone duration is reduced and hence its bandwidth increased, pitch discrimination is not constrained by the evoked excitation patterns along the basilar membrane, providing evidence against a place mechanism of pitch and in favor of temporal encoding at low frequencies. To determine the accuracy with which a temporal model will predict absolute-pitch identification as a function of duration and frequency, we examined performance of an autocorrelation model of pitch extraction with pure-tone stimuli as inputs to the model. The model consisted of a bank of 50 fourth-order GammaTone bandpass filters spaced logarithmically from 50 to 2500 Hz (Holdsworth et al., 1988; Saberi and Petrosyan, 2005). Filter bandwidths were based on human auditory filter estimates measured in notched-noise (Glasberg and Moore, 1990). The filterbank was followed by half-wave rectification, a square-law nonlinearity, and Gaussian noise added independently at the output of each of 50 filters, as well as independently again at the output of each filter during autocorrelation calculations to simulate time-dependent changes in the noise sample resulting from autocorrelation delay. Gaussian noise is a useful means of specifying the probabilistic nature of the model filter output if one assumes that a modest number of auditory nerve fibers form the peripheral neural bases of a given auditory filter. The sum of Poisson point processes which are characteristic of neural-spike timing patterns approaches a normal density as sample size increases to

---

[5] The tone bursts used in this experiment vary in duration as a function of frequency, thus producing a pitch associated with duration (see Footnote 4). This pitch, however, which could provide a spectral-edge cue in a frequency-discrimination task, is at an incorrect musical note frequency and thus cannot facilitate absolute-pitch identification.
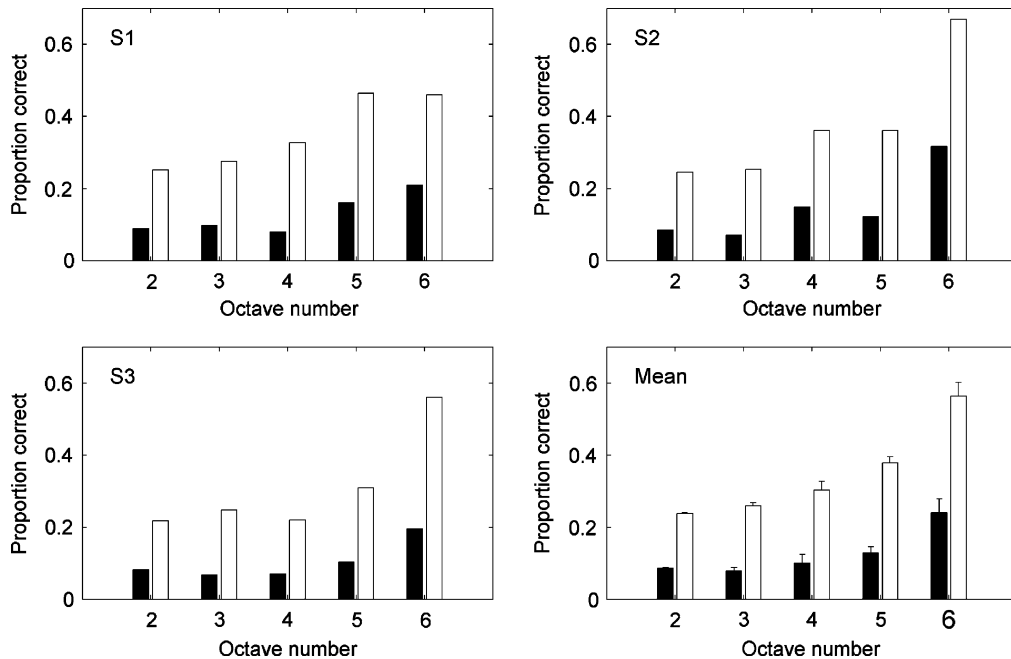
Fig. 3. Results from Experiment II for three subjects and their averaged data (bottom right) from 4500 pitch estimates. Stimuli were 5-cycle pure tones randomly selected from octaves 2 to 6. Black bars show results for exact pitch identification and white bars for identification to within a semitone. Error bars are one standard error.

about 30 (Pitman, 1993). Additive Gaussian noise has also been extensively used as an appropriate distribution for internal limiting noise (Spiegel and Green, 1981; Saberi and Green, 1996, 1997; Durlach et al., 2005; Beutelmann and Brand, 2006). The noise variance was one of two free parameters of the model and was kept constant. The signal output and thus effectively the signal-to-noise ratio within each channel was weighted by a frequency-dependent function representing outer- and middle-ear attenuation. This weighting function was derived from a logistic fit to the data shown in Fig. 2 of Meddis and Hewitt (1991). The filter outputs were followed by autocorrelation within frequency channels, frequency integration, and a decision device. A second free parameter of the model was the variance of a central noise added just prior to the decision. The decision device made one of 12 discrete responses per trial depending on the musical note frequency closest to the frequency designated by the peak of the autocorrelation function.

Fig. 4 shows output of the autocorrelation model, prior to frequency integration, in response to two musical note frequencies A4 (440 Hz) and B5 (987.8 Hz). The input stimuli were 10 ms pure tones. Red arrows mark autocorrelation peaks at the first non-zero lag. The model's pitch estimate is based on the inverse of the period determined from these autocorrelation peaks. Note that this peak is less sharply defined for the lower frequency note, resulting in greater variance of pitch estimation in presence of internal noise. Fig. 5 shows the model's predictions from 10,000 simulated trials per stimulus duration. Top-left panel shows predictions for exact note identification (filled circles) and identification to within a semitone (open circles).
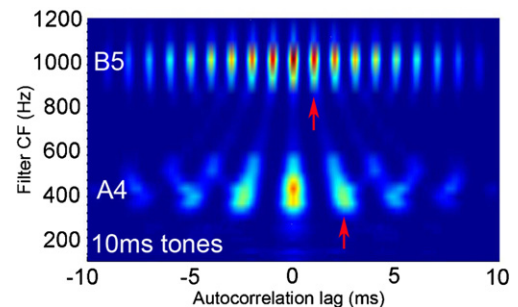

Fig. 4. Output of an autocorrelation model of pitch extraction prior to frequency integration. Input stimuli were 10 ms pure tones at 440 and 987.8 Hz (musical notes A4 and B5). The first non-zero autocorrelation peak corresponds to the model's pitch estimate (i.e., inverse of the periods marked by red arrows). Internal noise has been excluded from this image to facilitate visual comparison of autocorrelation peaks.

Top-right panel shows predictions plotted as a function of number of waveform cycles. Bottom panels show predictions sorted by stimulus octave (parameter). Bottom-left panel shows data corresponding to exact identification (associated with the filled circles of the top-left panel) as a function of duration, and bottom-right panel shows these predictions re-sorted by number of waveform cycles.

The autocorrelation model captures a number of trends in the data. First, there is a clear effect of duration on performance. While performance declines as stimulus duration is decreased, the model outperforms obtained data (Fig. 1) at the shortest-duration of 5 ms, suggesting that observers likely do not make optimal use of information derived from an autocorrelation process. Nonetheless, autocorrelation accurately predicts a decline in performance with
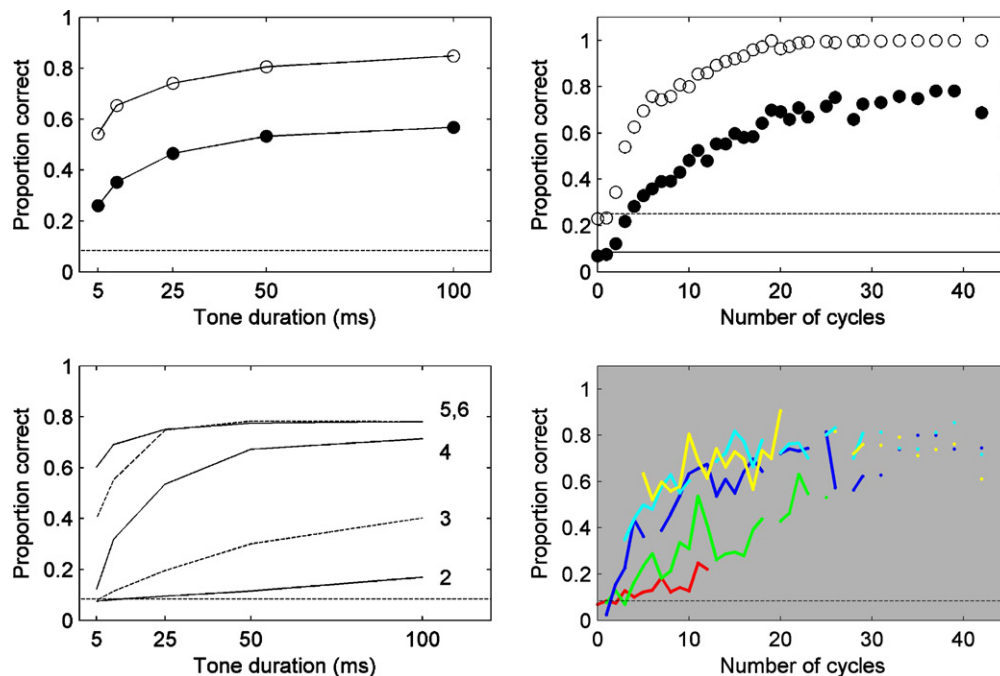
Fig. 5. Predictions of the autocorrelation model. Top left: proportion correct as a function of duration. Filled and open symbols represent exact note identification and identification to within a semitone, respectively. Bottom left: same as the top-left panel re-sorted by octave (cf. middle panel of Fig. 1). Top right: proportion correct predictions as a function of the number of waveform cycles (cf. top panel of Fig. 2). Bottom-right panel: predictions as a function of number of waveform cycles and octave number (cf. middle panel of Fig. 2). Octave numbers are 2 = red, 3 = green, 4 = blue, 5 = cyan, 6 = yellow.

decreased duration and nearly perfectly predicts the difference in performance between exact pitch identification and identification to within a semitone (difference between filled and open circles). The decline in performance with decreased duration is predicted from autocorrelation because shorter duration stimuli produce a smaller secondary peak in the normalized autocorrelation function from which a pitch estimate is derived (i.e., relative to the main peak at lag of zero). The secondary peak is relatively larger for longer duration stimuli, resulting in a more accurate estimation of pitch given constant internal noise. The model also shows that improvement in performance with increasing number of cycles reaches asymptotic levels at a rate similar to that observed for behavioral data.

Second, the autocorrelation model predicts a strong effect of stimulus octave on performance. The octave effect exists even in the absence of a frequency-weighting function (representing outer- and middle-ear effects; Meddis and Hewitt, 1991) but is more accurately representative of observer performance when this function is present. The origins of the predicted octave effect, separate from that associated with the frequency weighting function, is the larger number of cycles at higher octaves given a fixed-duration stimulus, and sharper peaks of the autocorrelation function at higher frequencies which result in a more stable estimate of pitch and less pronounced effect of internal noise. In addition, the model predicts larger improvements in performance from the 2nd to 4th octaves, and smaller improvements from 5th to 6th octaves, an effect also observed in the data of Fig. 1. The model also

predicts an octave effect even when results are sorted as a function of number of cycles. This prediction is confirmed by results of Experiment II.

One relevant question is whether the auditory system of absolute-pitch listeners is different than that of the general population in a way that affects interpretation of the current data. This, we believe, is not likely the case. Absolute-pitch ability is generally considered to be related to long-term memory and not to enhanced sensory processing. Frequency discrimination studies, for example, have shown no difference in resolution between AP and non-AP listeners (Siegel, 1972; Sergeant, 1969; Fujisaki and Kashino, 2002). Other auditory abilities, for example gap detection, detection of tones in spectrally notched noise, and interaural-delay thresholds have been reported to be similar for AP and non-AP subjects (Fujisaki and Kashino, 2002).

The present findings also raise interesting questions as to mechanisms of AP encoding as well as for further research. That AP listeners can extract pitch from a 4-cycle waveform, equivalent to the best reported results for pitch discrimination by non-AP subjects, suggests that the putative cognitive mechanisms which associate linguistic representations with pitch in accessing long-term memory for musical notes do not require a sustained stimulus. This suggests that as soon as a pitch percept is invoked, higher-level AP processes can instantaneously associate the appropriate semantic label to that musical pitch. The current results of course apply to "musically tuned" frequencies. An interesting question to address in future work is whether durational

limits observed here also apply to assimilation of off-pitch frequencies to their proximal standards? Would a similar minimum number of cycles be necessary in discrimination or matching tasks by AP subjects, especially given that AP musicians appear to compare pitches using different mechanisms (e.g., via more symbolic or lexical representations) than those employed by non-AP subjects (e.g., more direct judgments of frequency ratios or relative pitch)?

In summary, direct estimation of pitch of short-duration tones shows that a minimum of 4 cycles is sufficient to generate an identifiable percept of pitch. A temporal mechanism of pitch encoding at frequencies below 2 kHz, implemented as an autocorrelation model predicts several trends in the data including a decline in pitch-identification ability with decreased stimulus duration, improved performance with increasing stimulus frequency, as well as diminishing improvements at higher octaves (octaves 2–3 vs. 5–6). An autocorrelation mechanism, however, predicts better pitch *identification* than expected for very brief stimuli, consistent with Siebert's (1970) findings on pitch *discrimination* that observers do not make ideal use of available temporal information in very short tones. Finally, consistent with model predictions, a brief tone burst with a constant number of cycles produces a more salient pitch at higher frequencies as reflected in the monotonic improvement in pitch-identification performance as a function of stimulus octave.

Acknowledgements

References

Baharloo, S., Johnston, P.A., Service, S.K., Gitschier, J., 1998. Absolute pitch: an approach for identification of genetic and nongenetic components. Am. J. Human Genet. 62, 224–231.

Beerends, J.G., 1989. The influence of duration on the perception of pitch in single and simultaneous complex tones. J. Acoust. Soc. Am. 86, 1835–1844.

Beutelmann, R., Brand, T., 2006. Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am. 120, 331–342.

Bilsen, F.A., Goldstein, J.L., 1974. Pitch of dichotically delayed noise and its possible spectral basis. J. Acoust. Soc. Am. 55, 292–296.

Campbell, R.A., 1963. Frequency discrimination of pulsed tones. J. Acoust. Soc. Am. 35, 1193–1200.

Deutsch, D., Henthorn, T., Marvin, E., Xu, H.-S., 2006. Absolute pitch among American and Chinese conservatory students: prevalence differences, and evidence for a speech-related critical period. J. Acoust. Soc. Am. 119, 719–722.

Doughty, J.M., Garner, W.M., 1947. Pitch characteristics of short tones: I. Two kinds of pitch threshold. J. Exp. Psychol. 37, 351–365.

Doughty, J.M., Garner, W.M., 1948. Pitch characteristics of short tones: II. Pitch as a function of duration. J. Exp. Psychol. 38, 478–494.

Durlach, N.I., Mason, C.R., Gallun, F.J., Shinn-Cunningham, B., Colburn, H.S., Kidd, G., 2005. Informational masking for simulta-neous nonspeech stimuli: psychometric functions for fixed and randomly mixed maskers. J. Acoust. Soc. Am. 118, 2482–2497.

Freyman, R.L., Nelson, D.A., 1986. Frequency discrimination as a function of tonal duration and excitation-pattern slopes in normal and hearing-impaired listeners. J. Acoust. Soc. Am. 79, 1034–1044.

Fujisaki, W., Kashino, M., 2002. The basic hearing abilities of absolute pitch possessors. Acoust. Sci. Tech. 23, 77–83.

Gabor, D., 1947. Acoustical quanta and the theory of hearing. Nature 159, 591–594.

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

Gockel, H., Carlyon, R.P., Plack, C.J., 2005. Dominance region for pitch: effects of duration and dichotic presentation. J. Acoust. Soc. Am. 117, 1326–1336.

Golay, M.J.E., 1961. Complementary series. IRE Trans. Inform. Theory 7, 82–87.

Henning, G.B., 1970. Effects of duration on frequency and amplitude discrimination. In: Plomp, R., Smoorenburg, G.F. (Eds.), Frequency Analysis and Periodicity Detection in Hearing. A.W. Sijthoff, Leiden.

Holdsworth, J., Nimmo-Smith, I., Patterson, R., Rice, P., 1988. Implementing a Gammatone filterband. In: Annex C of the SVOS Final Report (Part A: The auditory Filter Bank), MRC (Medical Research Council), APU (Applied Psychology Unit) Report 2341, University of Cambridge, Cambridge, United Kingdom.

Levitin, D.J., 1994. Absolute memory for musical pitch: evidence from the production of learned melodies. Percept. Psychophys. 56, 414–423.

Mark, H.E., Rattay, F., 1990. Frequency discrimination of single-cycle, double-cycle, and triple-cycle sinusoidal acoustic signals. J. Acoust. Soc. Am. 88, 560–563.

Meddis, R., Hewitt, M.J., 1991. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: pitch identification. J. Acoust. Soc. Am. 89, 2866–2882.

Metters, P.J., Williams, R.P., 1973. Experiments on tonal residues of short duration. J. Sound Vib. 26, 432–436.

Miyazaki, K., 1989. Absolute pitch identification: effects of timbre and pitch region. Music Percept. 7, 1–14.

Moore, B.C.J., 1973. Frequency difference limens for short-duration tones. J. Acoust. Soc. Am. 54, 610–619.

Parncutt, R., 1994. Template-matching models of pitch and rhythm perception. J. New Music Res. 23, 145–167.

Patterson, R.D., Peters, R.W., Milroy, R., 1983. Threshold duration for melodic pitch. In: Klinke, R., Hartmann, R. (Eds.), . In: Hearing: Physiological Bases and Psychophysics. Springer Verlag, Berlin.

Pitman, J., 1993. Probability. Springer-Verlag, New York.

Pollack, I., 1967. Number of pulses required for minimal pitch. J. Acoust. Soc. Am. 42, 895.

Ritsma, R.J., Cardozo, B.L., Domburg, G., Neelen, J.J.M., 1966. The build-up of the pitch percept. IPO Ann. Prog. Rep. 1, 12–15.

Robinson, K., Patterson, R.D., 1995. The stimulus duration required to identify vowels, their octave, and their pitch chroma. J. Acoust. Soc. Am. 98, 1858–1865.

Ronken, D.A., 1971. Some effects of bandwidth-duration constraints on frequency discrimination. J. Acoust. Soc. Am. 49, 1232–1242.

Ross, D.A., Olson, I.R., Marks, L.E., Gore, J.C., 2004. A nonmusical paradigm for identifying absolute pitch possessors. J. Acoust. Soc. Am. 116, 1793–1799.

Saberi, K., Green, D.M., 1996. Adaptive psychophysical procedures and imbalance in the psychometric function. J. Acoust. Soc. Am. 100, 528–536.

Saberi, K., Green, D.M., 1997. Evaluation of maximum-likelihood estimators in non-intensive auditory psychophysics. Percept. Psychophys. 59, 867–876.

Saberi, K., Petrosyan, A., 2005. Neural cross-correlation and signal decorrelation: insights into coding of auditory space. J. Theor. Biol. 235, 45–56.

Sekey, A., 1963. Short-term auditory frequency discrimination. J. Acoust. Soc. Am. 35, 682–690.

Sergeant, D.C., 1969. Experimental investigation of absolute pitch. J. Res. Music Educ. 17, 135.

Siebert, W.M., 1970. Frequency discrimination in the auditory system: place or periodicity mechanism? Proc. IEEE 58, 723–730.

Siegel, J.A., 1972. The nature of absolute pitch. In: Gordon, I.E. (Ed.), . In: Studies in Psychology of Music, vol. 8. University of Iowa Press, Iowa City, p. 65.

Sipovsky, A.V., Gershundi, G.V., Gorelik, B.M., Korotkin, I.I., Lubinsky, J.A., 1972. The determination of differential frequency thresholds for short tone signals. Biofizika 17, 495–502.

Spiegel, M.F., Green, D.M., 1981. Two procedures for estimating internal noise. J. Acoust. Soc. Am. 70, 69–73.

Terhardt, E., 1979. Calculating virtual pitch. Hear. Res. 1, 155–182.

Terhardt, E., Stoll, G., Schermbach, R., Parncutt, R., 1986. Pitch ambiguity, tone affinity, and identification of successive intervals. Acustica 61, 57–66.

Wightman, F.L., 1973. The pattern-transformation model of pitch. J. Acoust. Soc. Am. 54, 407–416.

Yost, W.A., 1996a. Pitch of iterated rippled noise. J. Acoust. Soc. Am. 100, 511–518.

Yost, W.A., 1996b. Pitch strength of iterated rippled noise. J. Acoust. Soc. Am. 100, 3329–3335.

Zatorre, R.J., 2003. Absolute pitch: a model for understanding the influence of genes and development on neural and cognitive function. Nature Neurosci. 6, 692–695.

Zatorre, R.J., Perry, D.W., Beckett, C.A., Westbury, C.F., Evans, A.C., 1998. Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. Proc. Natl. Acad. Sci. USA 95, 3172–3177.

Zhou, B., Green, D.M., Middlebrooks, J.C., 1992. Characterization of external ear impulse responses using Golay codes. J. Acoust. Soc. Am. 92, 1169–1171.