Available online at www.sciencedirect.com
SCIENCE DIRECT®
Hearing Research 194 (2004) 31–46
HEARING RESEARCH
www.elsevier.com/locate/heares

Pitch is determined by naturally occurring periodic sounds

David A. Schwartz *, Dale Purves

Center for Cognitive Neuroscience and Department of Neurobiology, Duke University, Box 90999, Durham, NC 27708-0999, USA

Received 11 November 2003; accepted 23 January 2004
Available online 28 May 2004

Abstract

The phenomenology of pitch has been difficult to rationalize and remains the subject of much debate. Here we test the hypothesis that audition generates pitch percepts by relating inherently ambiguous sound stimuli to their probable sources in the human auditory environment. A database of speech sounds, the principal source of periodic sound energy for human listeners, was compiled and the dominant periodicity of each speech sound determined. A set of synthetic test stimuli were used to assess whether the major pitch phenomena described in the literature could be explained by the probabilistic relationship between the stimuli and their probable sources (i.e., speech sounds). The phenomena tested included the perception of the missing fundamental, the pitch-shift of the residue, spectral dominance and the perception of pitch strength. In each case, the conditional probability distribution of speech sound periodicities accurately predicted the pitches normally heard in response to the test stimuli. We conclude from these findings that pitch entails an auditory process that relates inevitably ambiguous sound stimuli to their probable natural sources.

Keywords: Pitch; Auditory; Perception; Probability; Speech; Psychoacoustics

1. Introduction

The perception of pitch is central to both language and music, two acoustically mediated forms of expression found in every human population. In language, the tonal quality of speech serves both lexical and indexical functions, conveying a speaker's emotional state, communicative intent (e.g., whether an utterance is a question or an assertion), and, in tone languages, which member of a set of homophones a speaker intends (Fromkin and Rodman, 1998). In music, the specific pitches elicited by tone-evoking stimuli played simultaneously or sequentially are the basis of harmony and melody, respectively.

Although the pitch a listener hears is roughly related to the repetition rate of an acoustical waveform, it has long been apparent that pitch is not a simple function of frequency, or indeed of any other physical parameter of the stimulus (Hall and Peters, 1981). The aspects of pitch that have proved particularly difficult to explain within a unified theoretical framework include the observations that: (1) the predominant pitch heard in response to a stimulus comprising harmonically related pure tones typically corresponds to the greatest common divisor of the tones (i.e., to the fundamental frequency, F_0), even when there is no spectral energy in the stimulus at that frequency (Seebeck, 1841; Schouten, 1938); moreover, for reasons that remain unclear, this phenomenon is apparent only for fundamental frequencies up to ~1000 Hz (Fletcher, 1924); (2) when the frequencies of a set of successive harmonics in a stimulus are increased by a constant value such that they lack a common divisor greater than ~70 Hz, the frequency associated with the predominant pitch no longer corresponds to the fundamental (a phenomenon called the 'pitch-shift of the residue') (de Boer, 1956; Schouten et al., 1962); (3) when the frequencies of a subset of the harmonics in a stimulus are altered such that the fundamental frequency of the subset differs from that of the rest of the harmonic series, the pitch corresponds to the fundamental of those harmonics that occupy a frequency band centered at ~600 Hz (a phenomenon called

* Corresponding author. Tel.: +1-919-684-3318; fax: +1-919-681-0815.
E-mail address: schwartz@neuro.duke.edu (D.A. Schwartz).

0378-5955/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.heares.2004.01.019

'spectral dominance') (Plomp, 1967; Ritsma, 1967, 1970; Moore et al., 1985; Dai, 2000; Jarveläinen et al., 2002); and (4) the relative strength or salience of the pitch heard in response to a stimulus is greatest for waveform repetition rates between ~200 and 500 Hz (Terhardt et al., 1986; Huron, 2001).

Here we test the hypothesis that these phenomena all reflect the way the human auditory system contends with the inherent ambiguity of sound stimuli. The relationship between a sound stimulus and its possible sources in the human auditory environment is by its nature uncertain because the mechanical forces acting on a resonating body, the physical properties of that body and the effects of the local environment on the transmittance of sound are conflated in the stimulus at the ear. In consequence, listeners cannot apprehend the relative contribution of these several physical factors (sources) by any analytical operation on the stimulus (Tarantola, 1987; Gordon et al., 1992). Nonetheless, reacting appropriately to the physical source(s) of the stimulus, rather than to the stimulus itself, is the objective of any behavior guided by audition. This "inverse acoustics problem" implies that auditory percepts are generated probabilistically, such that the listener hears the acoustical characteristics of the probable source of the stimulus and not the characteristics of the stimulus per se.

The naturally occurring sounds that evoke a perception of pitch comprise a repeating (i.e., periodic) pattern of pressure change at the ear combined with variations in sound pressure that have little or no repeating character. The hypothesis examined here is thus that the pitch heard in response to any stimulus will always correspond to the periodicity of the natural sound that has most often been the source of the periodic sound energy in that particular stimulus (see Section 4 for the broader background of this idea). As a result, the pitch heard will not necessarily correspond to the repetition rate of the stimulus as such (i.e., its fundamental frequency, F_0), but rather to the repetition rate of the naturally occurring periodic signals that have typically contributed periodic sound energy to the same or a similar stimulus. Thus, whereas most previous investigations of pitch have proceeded from an analysis of either the physical stimulus at the ear or the transformed signal at the auditory nerve, the present work proceeds from an analysis of the relevant auditory environment.

To examine the merits of this framework for understanding the phenomenology of pitch, we created a series of artificial sound stimuli and assessed their probabilistic relationship to a database of naturally occurring sounds representative of the pitch-evoking stimuli to which human listeners have normally been exposed (voiced speech sounds). According to the hypothesis under consideration here, the dominant pitch frequency heard by a listener in response to such artificial test stimuli (expressed in terms of the sinusoidal frequency listeners assign to it) should correspond to the periodicity of the probable source of the stimulus.

2. Materials and methods

2.1. Database

Human vocalization is the principal naturally-occurring source of the periodic sound energy to which human beings have been exposed over both evolutionary time and the lifetime of individuals; the only other natural sources of periodic stimuli are the calls of other animals and sounds generated by those relatively rare circumstances in which mechanical actions (e.g., of wind or water) generate periodic sound stimuli incidentally. Accordingly we used the Texas Instruments/Massachusetts Institute of Technology (TIMIT) Acoustic-Phonetic Continuous Speech Corpus as a representative database of the periodic sounds to which human listeners have typically been exposed. The corpus comprises 10 sentences uttered by each of 189 female and 441 male native English speakers representing eight dialect regions of the United States (Garofolo et al., 1990). The utterances were recorded at a rate of 16,000 samples/s. Other technical specifications regarding the selection of speakers, construction of the sentence list, recording conditions and signal processing can be found in Fisher et al. (1986) and Lamel et al. (1986) or obtained from the Linguistic Data Consortium at the University of Pennsylvania, http://www.ldc.upenn.edu/.

Fig. 1 illustrates the method by which we extracted a random (within gender) sample of individual speech sounds taken from 100 male and 100 female speakers in the TIMIT corpus. For each speaker, we randomly sampled 40 brief (50 ms) speech segments from each of 10 utterances, yielding a total of 400 speech segments per speaker (80,000 segments in total). Each segment was saved as a sequence of numbers representing the variation in sound pressure over time. An interval of 50 ms was chosen because it is the shortest period that allows reliable detection of the lowest speech sound frequencies (i.e., 3 cycles of a 60 Hz signal).

A standard autocorrelation algorithm (Boersma, 1993; Boersma and Weenink, 2003) was used to estimate both the waveform repetition rates in each segment (i.e., the periodicities), and the harmonics-to-noise ratio (a number between 0 and 1 equal to the normalized autocorrelation) of each periodicity, subsequently referred to as the "strength" of the periodicity (see Fig. 1(c)). Periodicities ranging from 60 to 8000 Hz were evaluated with a precision >0.01 Hz, thus sampling the full ~70–5000 Hz range of spectral energy in adult speech sounds (Lieberman and Blumstein, 1988). By default, the algorithm classifies a segment in which all estimated periodicities have a strength value < 0.45 as voiceless.

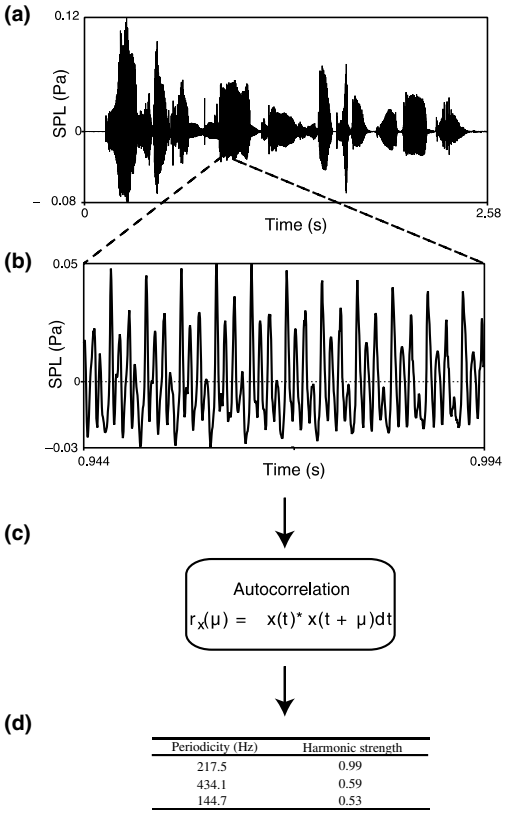

Fig. 1. Method by which speech segments were sampled from the speech sound database, and how the dominant periodicities and strengths of each segment were determined. (a) The time-domain representation of an utterance, made by a female speaker in this example. (The sentence spoken is, "The essay undeniably reflects our view ably".) (b) A blow-up of a 50 ms segment whose waveform has an average period of 4.6 ms (~217 Hz). (c) Schematic representation of the application of the autocorrelation algorithm (see text) to the speech sound depicted in (b). (d) This procedure yields the estimated periodicities in the waveform and their relative strengths, which were used subsequently to determine the relative likelihood of different periodicities in a given artificial test stimulus, as illustrated in Figs. 3 and 4.

Approximately 14% of the 80,000 speech sound segments were classified as voiceless by this method, consistent with the empirical finding that unvoiced phonation accounts for ~15% of speech (Cook, 1999). Because voiceless phonation is not a significant source of pitch-evoking stimuli, all speech segments classified as voiceless were removed from the database, leaving 68,836 segments (86% of the original sample). The algorithm was implemented in Praat speech analysis software (Boersma and Weenink, 2003) running on a Macintosh G4 computer.

Fig. 2 shows the relative frequency of occurrence (i.e., the empirical probability distribution) of strongest periodicities in the database of 68,836 speech sounds sampled from the TIMIT. The distribution is bimodal, with approximately half the segments distributed around a mean of ~110 Hz (arising primarily from the male speakers) and the other half distributed around a

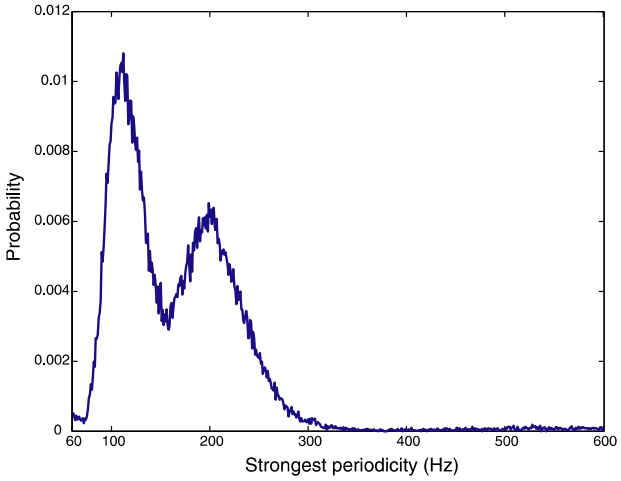

Fig. 2. Probability distribution of strongest periodicities in the speech sound database (see text for explanation).

mean of ~200 Hz (primarily female speakers). Ninety-seven percent of the speech sounds have their strongest periodicity at values less than 300 Hz. In the analysis that follows, therefore, predictions of pitch frequency are limited to frequencies in the 60–300 Hz range, where there is sufficient information in the database to obtain reliable estimates of mean cross-correlation.

2.2. Construction of test stimuli

A variety of artificial stimuli comprising two or more sine waves of equal amplitude, each having a specified frequency, were constructed to compare with the speech sounds from the database (see Section 3). For example, a stimulus comprising the third, fourth, and fifth harmonics of a 150 Hz fundamental was defined by $x = \sin(2\pi t * 450) + \sin(2\pi t * 600) + \sin(2\pi t * 750)$, where t is a 50 ms time vector sampled every 6.25×10^{-5} s. Each stimulus thus represented a 50 ms stationary complex tone sampled at 16,000 Hz, the sampling frequency of the TIMIT speech sounds.

2.3. Determining the relative likelihood of the possible sources of a test stimulus

In terms of the hypothesis being considered here, responding appropriately to the information carried by the periodic component of a pitch-evoking stimulus (e.g., the size or sex of the speaker) is predicated on *hearing* the periodicity of the probable source of the stimulus. In Bayesian terms, the probability of sound event X (a speech sound in the present case) being the source of the periodic sound energy in a test stimulus Y (i.e., $P[X|Y]$) is a function of the prior probability of the different possible sound event sources and of the likelihood of the stimulus given each possible source. That is,

$P(X|Y) = P(Y|X) * P(X)$.

Given the framework advanced here, the prior probability of different possible sources derives entirely from past experience (both evolutionary and developmental) with the natural sources of periodic sound in the human auditory environment. Thus, the prior probability distribution of possible sources is implicit in the composition of the speech sound database. To determine the relative likelihood of the stimulus given each possible source, we treated each stimulus Y as the sum of a speech sound X in the TIMIT database and Gaussian noise $(0, \partial^2)$. Each stimulus is thus probabilistically related to each speech sound, $P(Y|X)$ increasing with the physical similarity between X and Y. Under these conditions, the likelihood that Y has been generated by X (i.e., by a particular speech sound) increases with the squared cross-correlation between X and Y (see Zucker, 2003).

Accordingly, we computed the normalized cross-correlation of a given test stimulus with each of the 68,836 speech segments in the database using the built-in XCORR function of the MATLAB Signal Processing Toolbox (Figs. 3(a) and (b)). Given that likelihood increases with cross-correlation, we took the coefficient of the maximum normalized cross-correlation, which indicates the relative time-adjusted physical similarity of the test stimulus to a given speech sound (see Fig. 3(c)), as a measure of the likelihood of a test stimulus, given a particular speech sound (i.e., $P[Y|X]$). (Note that only the maximum cross-correlation coefficient relating two signals, and not the lag associated with this maximum, is used in the subsequent analysis; the lag can be considered a measure of the time adjustment needed to bring

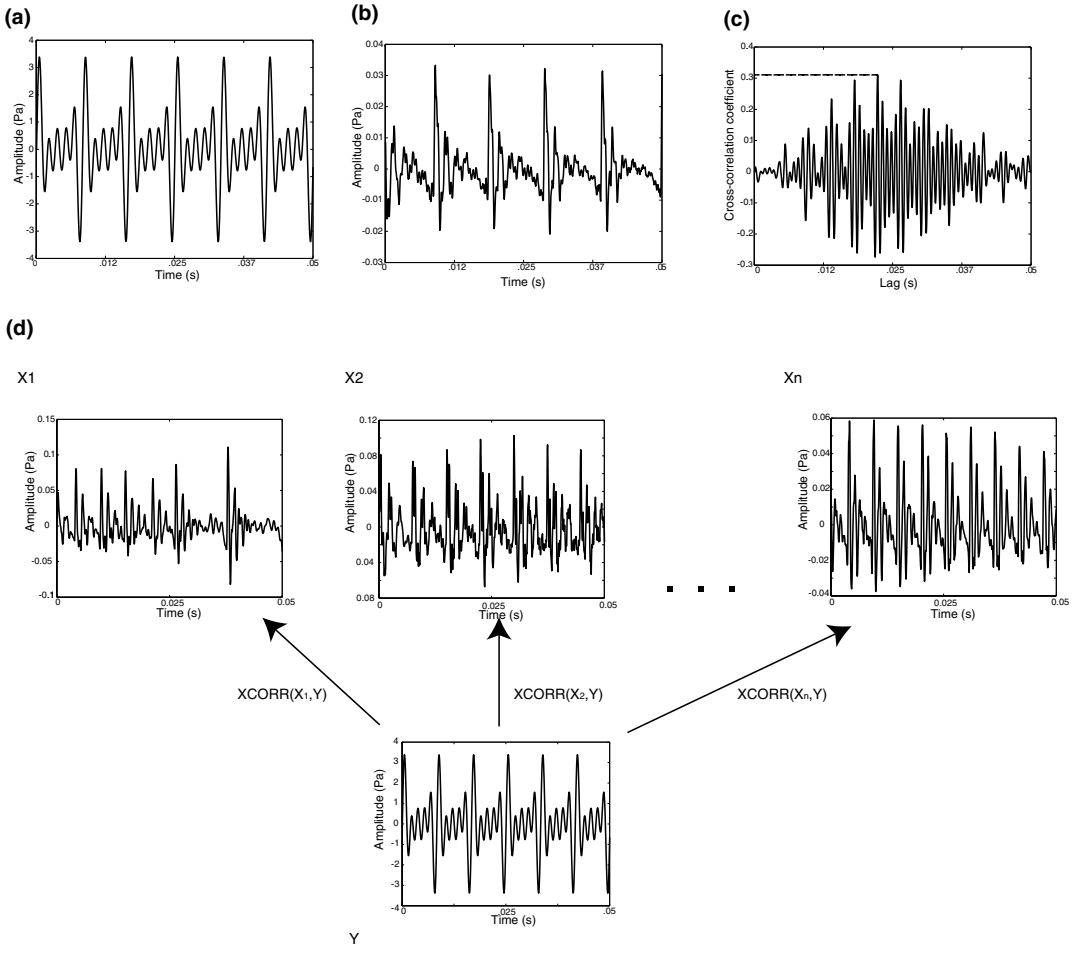

Fig. 3. Determining the relative likelihood of the possible sources of a test stimulus. (a) A time-domain representation of an artificially constructed test stimulus generated by the summation of four sinusoids (defined by $0.5 * \sin[2\pi t * 240] + \sin[2\pi t * 360 * t] + \sin[2\pi t * 480 * t] + 1.2 * \sin[2\pi t * 600 * t]$; see Section 2). (b) A 50 ms speech sound from the speech database. (c) The normalized cross-correlation function of the artificial stimulus in (a) and the speech segment in (b). The dashed line indicates the maximum cross-correlation coefficient in this comparison; this maximum was used in the subsequent analysis to estimate the likelihood that a given speech sound segment from the database could have contributed the periodic sound energy in a given artificial stimulus. (d) Schematic representation of this process applied to the entire TIMIT database to obtain a probability distribution of the possible natural sources of the periodic sound energy in stimulus Y. See text for details.

to the two signals into optimum alignment.) Since the prior probability distribution of possible stimulus sources is implicit in the speech sound database, the maximum normalized cross-correlation was taken as a measure of the likelihood of speech sound X being a source of the periodic sound energy in stimulus Y (i.e., the posterior probability $P(X|Y)$). Fig. 3(d) presents a schematic representation of this process applied to the entire TIMIT database to obtain a probability distribution of the possible natural sources of the periodic sound energy in stimulus Y.

2.4. Predicting pitch from the distribution of probable sources

The dominant periodicity value for each speech sound was rounded to the nearest integer and the mean cross-correlation for each integer periodicity bin computed. The periodicity value associated with the maximum mean cross-correlation was taken as the predictor of the frequency listeners would assign to the pitch of the stimulus (see Figs. 4(c) and (d) for illustration of this procedure). The relative height of the maximum was taken as the predictor of the relative strength of the pitch elicited by that stimulus.

3. Results

The experimental findings to be explained in terms of the hypothesis that the pitch of complex tones is determined by the probabilistic relationship between auditory stimuli and their natural sources are: (1) hearing pitches that correspond to the greatest common divisor of a set of successive harmonics for waveform repetition rates less than ~1000 Hz; (2) the pitches heard in response to

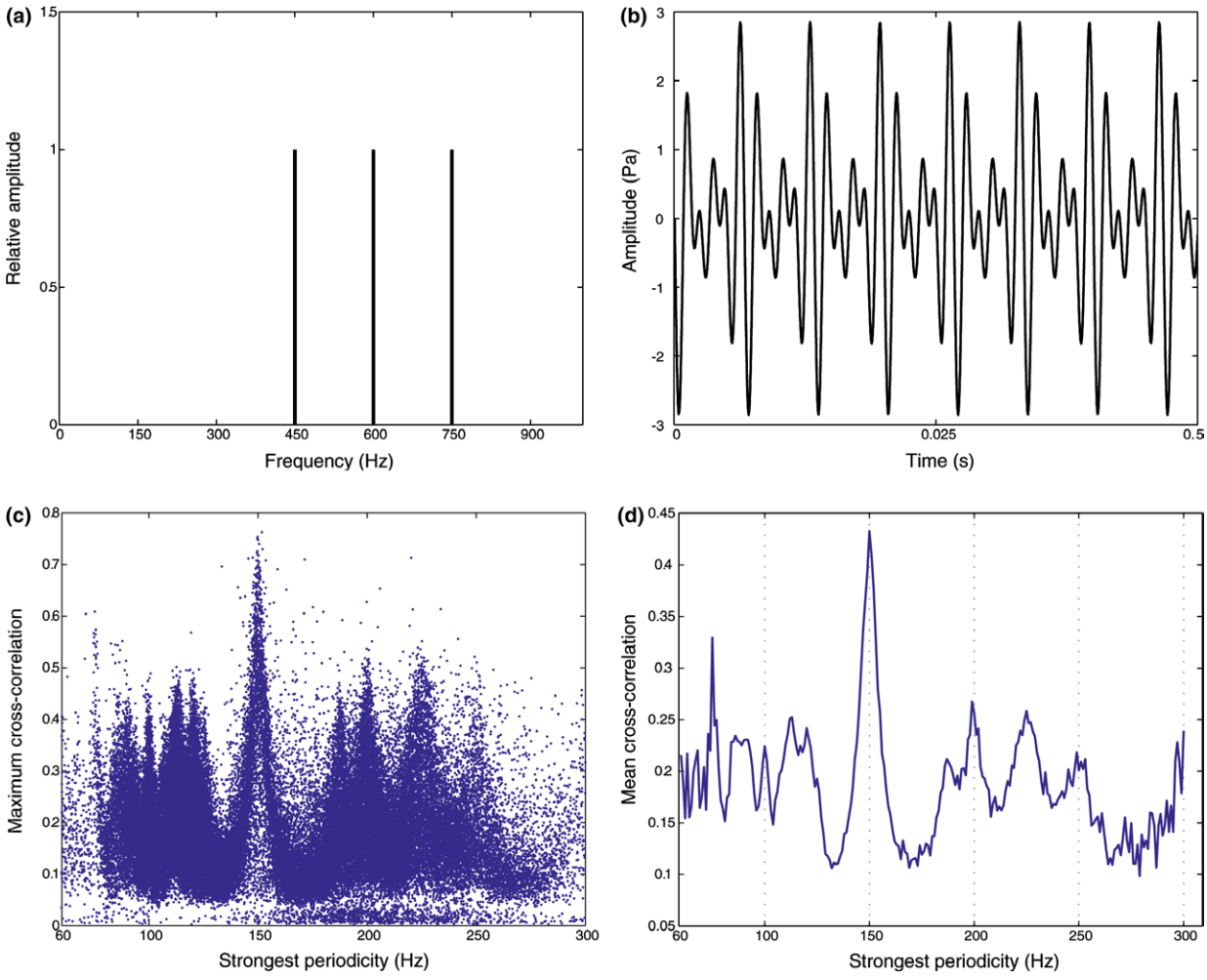

Fig. 4. Predicting the pitch of a stimulus that comprises three successive harmonics of a fundamental frequency. (a) Schematic frequency domain representation of a stimulus comprising the third, fourth and fifth harmonics of a 150 Hz fundamental. (b) Time-domain representation of the same stimulus. (c) The maximum cross-correlation coefficient of each speech sound in the database (see Fig. 3) with respect this stimulus, plotted against the strongest periodicity of each speech segment (see Fig. 1). (d) Average maximum cross-correlation for each integer frequency bin, derived from the data in (c). The periodicity associated with the maximum of the function is 150 Hz.

spectral components that lack a common divisor greater than 70 Hz (the 'pitch-shift of the residue'); (3) the dominant influence of a relatively narrow frequency band centered at ~600 Hz in determining pitch ('spectral dominance'); and (4) the greater strength and clarity of the pitches heard in response to stimuli with fundamental frequencies between ~200 and 500 Hz.

3.1. Hearing the common divisor of successive harmonics

Seebeck (1841) was the first to demonstrate that the frequency of the pitch heard in response to a set of two or more successive harmonics corresponds to the greatest common divisor of the harmonic set, even when there is no spectral energy at that frequency. This phenomenon (hearing the fundamental frequency of the harmonically-related spectral components) has been variously referred to as the pitch of the missing fundamental, periodicity pitch, low pitch, residue pitch and/or virtual pitch.

Figs. 4(a) and (b) show, respectively, a schematic frequency-domain and time-domain representation of a 50 ms artificial test stimulus comprising the third, fourth and fifth harmonics of a 150 Hz fundamental frequency. Figs. 4(c) and (d) present an illustration of the analytical procedure used to predict the pitch of this stimulus. Fig. 4(c) plots the strongest periodicity in each speech segment against the segment's normalized maximum cross-correlation with respect to the test stimulus (see Section 2). Fig. 4(d) shows the average maximum cross-correlation coefficient for each integer frequency bin, derived from the data in Fig. 4(c). The periodicity associated with the maximum of this function is 150 Hz, which matches both the absent 150 Hz F_0 of the sine tones making up the stimulus in Fig. 4(a) and the sinusoidal frequency listeners assign to the dominant pitch of this stimulus (Seebeck, 1841; Fletcher, 1924; Schouten, 1938; Licklider, 1954). Note that the calculation of the probability described in Section 2 pertains only to the periodicity associated with this global maximum of the probability distribution, whereas the distribution in Fig. 4(d) is multimodal (with, for example, local maxima at 75 and 225 Hz). The presence of these lesser peaks signify additional periodicities that have often been salient in the natural sources of the test stimulus, and is consistent with the observation that listeners may identify two or more distinct pitches in response to a tonal stimulus such as that illustrated in Fig. 4(a) (Schouten et al., 1962; see also Section 4).

Fig. 5 shows the results of using this same approach to predict the dominant pitch heard in response to a variety of other test stimuli comprising two or more successive harmonics. These include: (1) the first 12 harmonics of a complex tone having a fundamental frequency of 200 Hz (stimulus 1 in Fig. 5(a)); (2) the first 11 harmonics minus the fundamental (stimulus 2); (3) three successive upper harmonics covering different frequency ranges (stimuli 3,4 and 5); and (4) a stimulus comprising two upper harmonics only (stimulus 6). Fig. 5(b), like Fig. 4, presents an illustration of the analytical procedure used to estimate the pitch of these stimuli. In each case the periodicity associated with the maximum of the probability distribution accurately predicts the dominant pitch listeners hear in response to these stimuli (Fletcher, 1924; Schouten, 1938; Schouten et al., 1962; Smoorenburg, 1970; Rasch and Plomp, 1999).

3.2. Pitch-shift of the residue

When the frequencies of a set of successive harmonics such as the stimuli in Fig. 5(a) are altered by adding or subtracting a constant value such that they lack a common divisor greater than ~70 Hz, the pitch that listeners hear typically shifts in the direction of the frequency change, but no longer corresponds to the fundamental frequency of the set (de Boer, 1956). This phenomenon is called the 'pitch-shift of the residue'.

Figs. 6(a) and (b), for example, show a schematic frequency-domain and a time-domain representation, respectively, of a 50 ms test stimulus comprising the 9th, 10th and 11th harmonics of a series with a 200 Hz fundamental, each of which has been increased by 40 Hz. The sinusoidal components are thus 1840, 2040, and 2240 Hz, instead of 1800, 2000 and 2200 Hz. As a result, the fundamental frequency of the set is 40 Hz, even though the spacing between the components remains 200 Hz. The dominant pitch heard in response to this stimulus is ~204 Hz (Schouten, 1940). Figs. 6(c) and (d), like Fig. 5(b), illustrate the analytical procedure used to estimate the pitch of this stimulus. Fig. 6(c) plots the strongest periodicity of each speech segment against that segment's maximum cross-correlation with the test stimulus; Fig. 6(d) shows the mean of the maximum cross-correlation coefficients for each integer frequency bin, derived from the data in Fig. 6(c). Although the distribution is again multimodal, the periodicity associated with the maximum of the function is 204 Hz.

A further as yet unexplained aspect of the pitch-shift of the residue is that when the frequency increment (or decrement) applied to a set of successive harmonics is systematically varied, the slope of the best linear fit to the observed pitch-shift of the residue is typically ~20% steeper than the slope defined by the equation $\Delta P = \Delta F/n_c$, where P represents pitch, F represents frequency, and n_c represents the harmonic rank of the central spectral component in the stimulus (Schouten et al., 1962). Fig. 7 shows the change in pitch predicted on the basis of systematically varying the frequency increment (or decrement) of a stimulus comprising the fourth through the eighth harmonics of a 200 Hz F_0 (after de Boer, 1956). The filled circles indicate the

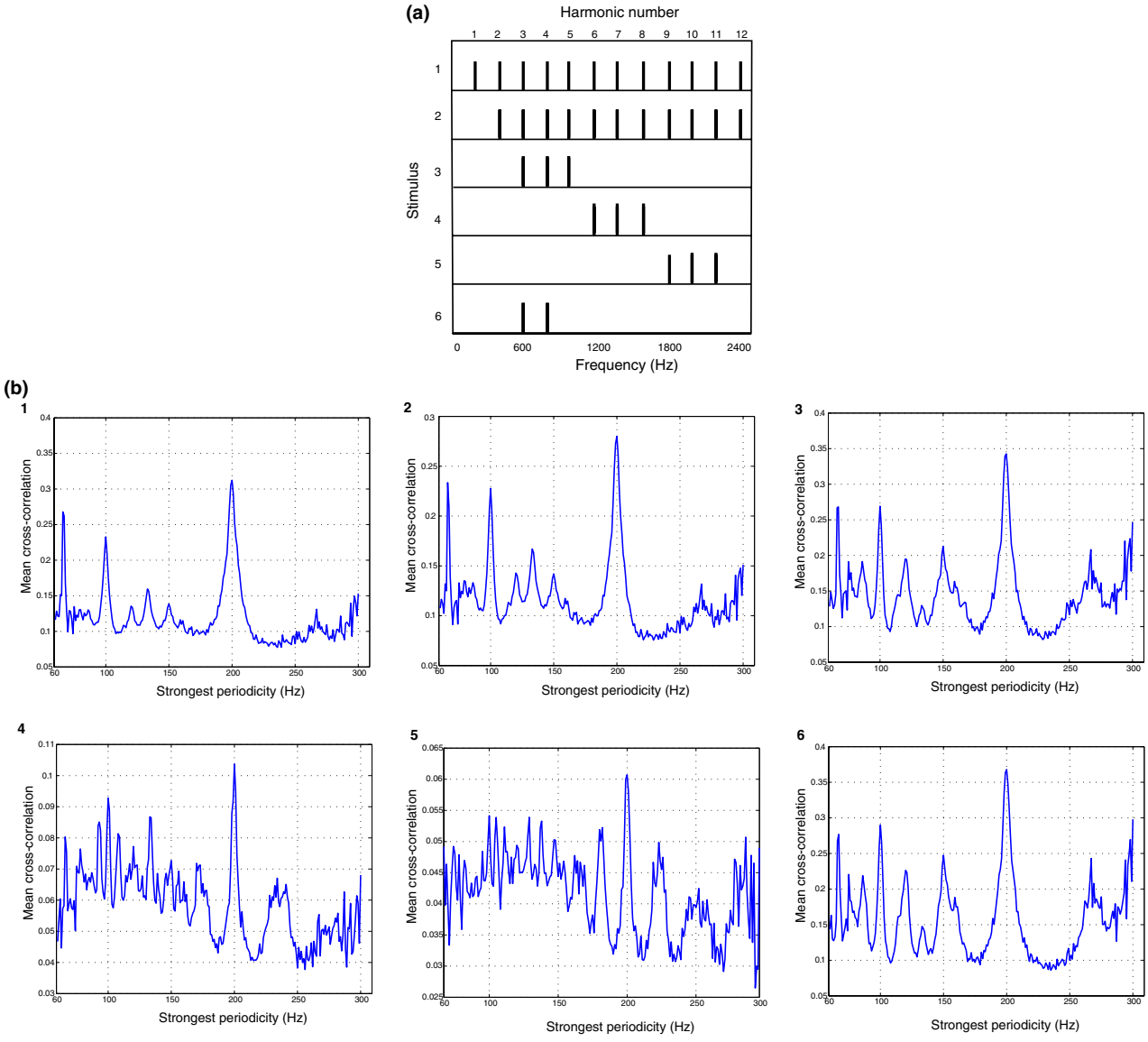

Fig. 5. Predicting the pitch of a variety of stimuli comprising successive harmonics, each analyzed in the manner illustrated in Fig. 4. (a) Schematic frequency-domain representation of six different stimulus types; all these examples have a fundamental frequency of 200 Hz. (b) Average maximum cross-correlation for each integer frequency bin. The periodicity associated with the maximum of each function is 200 Hz.

predicted pitches determined from the probabilistic analysis described above. The solid line indicates the best linear fit to the predicted pitch shifts, and the dashed line indicates the pitch shifts predicted by the function $\Delta P = \Delta F/n_c$. The slope of the solid line is 22% steeper than the slope of the dashed line, a difference consistent with the psychophysical findings reported by de Boer and Schouten et al. The accuracy of these predictions further supports the hypothesis that pitch is determined by the probabilistic relationship between a stimulus and its natural sources.

3.3. Spectral dominance

When the frequencies of only some of the harmonics of a complex tone stimulus are changed such that the fundamental of the lower harmonics (e.g., rank <7) differs from that of the higher harmonics, the pitch heard typically corresponds to the fundamental frequency of the lower harmonics (Plomp, 1967; Ritsma, 1967, 1970; Moore et al., 1985). More specifically, the percept corresponds to the F_0 of the three or four spectral components closest to ~600 Hz (Dai, 2000; Jarveläinen et al., 2002).

As an example, Figs. 8(a) and (b) show, respectively, a schematic frequency-domain and time-domain representation of a 50 ms stimulus comprising the first 10 harmonics of a series with a 200 Hz F_0, with the frequency of the second, third and fourth harmonics (i.e, the values 400, 600 and 800 Hz) augmented by 3% (indicated by the arrow in the figure). As a result, the F_0 of this altered subset of is 206 Hz. Despite the fact that the F_0 of the majority of spectral components in the stimulus

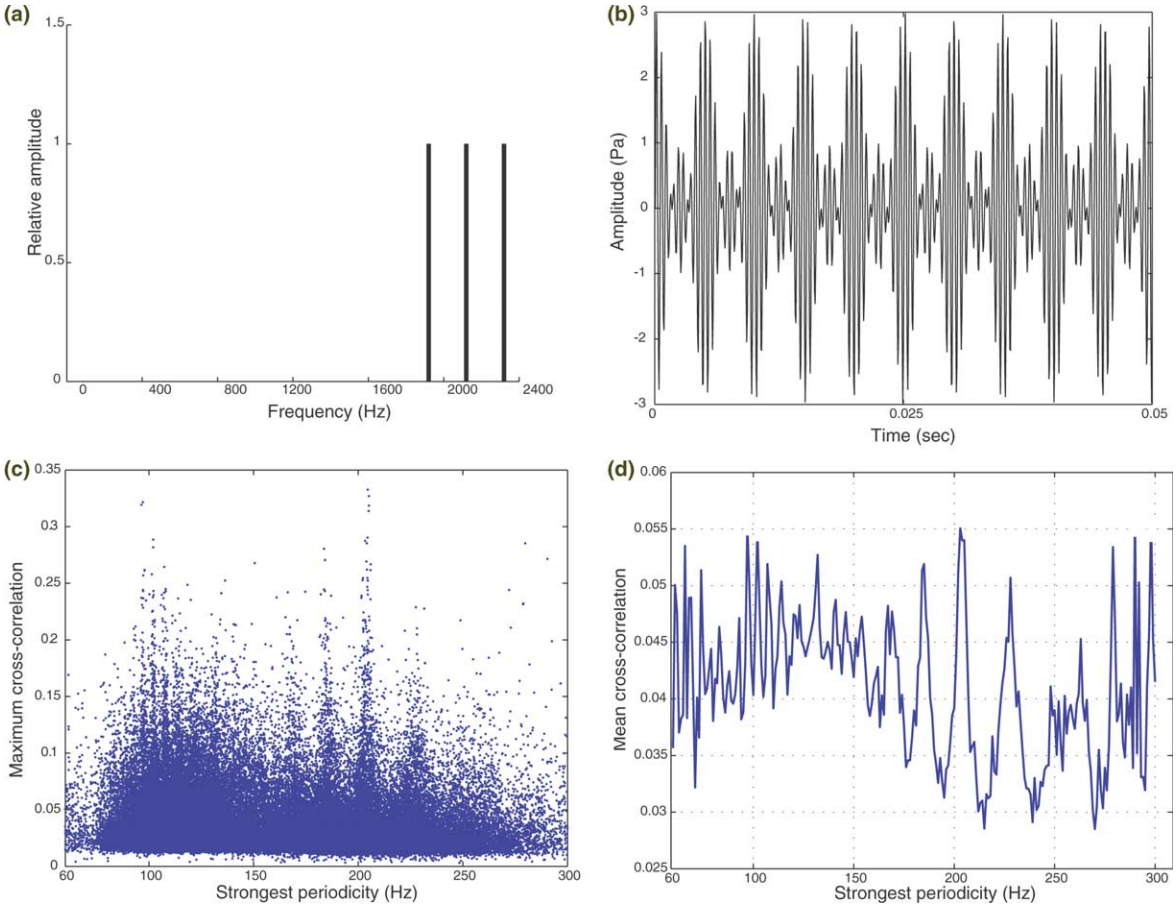

Fig. 6. Predicting the pitch-shift of the residue. (a) Schematic frequency-domain representation of a stimulus comprising the 9th, 10th and 11th harmonics of a 200 Hz fundamental; the frequency of each harmonic has been augmented by 40 Hz. (b) Time-domain representation of the same stimulus. (c) The maximum cross-correlation coefficient of each speech sound in the database with respect the stimulus in (a) and (b) plotted against the strongest periodicity in each speech segment. (d) Average maximum cross-correlation for each integer frequency bin, derived from the data in (c). Although the function has many local peaks, the periodicity associated with the maximum of the function in (d) is 204 Hz.

is still 200 Hz, the dominant pitch heard in response to this stimulus is ~206 Hz (Dai, 2000). Figs. 8(c) and (d) again illustrate of the analytical procedure used to predict the pitch of this stimulus. Fig. 8(c) shows the strongest periodicity of each speech segment in the database plotted against its maximum cross-correlation with respect to this stimulus, as in previous figures. Fig. 8(d) shows the average of the maximum cross-correlation coefficients for each integer frequency bin, derived from the data in Fig. 8(c). The periodicity associated with maximum of this function is 206 Hz, which matches the perceived pitch of the stimulus.

For the harmonic series in Fig. 8 with a fundamental frequency of 200 Hz the third harmonic corresponds to the center of the dominant frequency band at 600 Hz. However, for a series with a fundamental of 100 Hz it is the sixth harmonic that corresponds to this center frequency. As a result, the harmonic rank of the spectral components that determine the dominant pitch heard by listeners varies as a function of fundamental frequency, with the rank of the dominant components being inversely related to F_0 (Dai, 2000). Fig. 9(a) shows a schematic frequency-domain representation of the eight different stimuli we used to test whether the probabilistic relationship between a stimulus and its possible natural sources also predicts this observed variation in the rank of dominant harmonics as a function of F_0. The artificial test stimuli in this case comprised 10 harmonics with an F_0 of either 100 or 200 Hz; three harmonics of each stimulus were augmented by 3%, as indicated by the arrows in the figure. The frequency band of the augmented subset was systematically varied by incrementally increasing the harmonic rank of the center spectral component from $n = 2$ to $n = 9$. Fig. 9(b) shows the predicted change in pitch as a function of the rank of the middle harmonic in the subset for $F_0 = 100$ Hz; Fig. 9(c) shows the predicted change for $F_0 = 200$ Hz. In each case tested in this way the predictions accord with the psychophysical findings. As indicated in Figs. 9(b) and 9(c), when the augmented subset of harmonics includes the 600 Hz component, the predicted pitch increases. Conversely, when the augmented subset of harmonics does not include this component there is no predicted change in pitch.

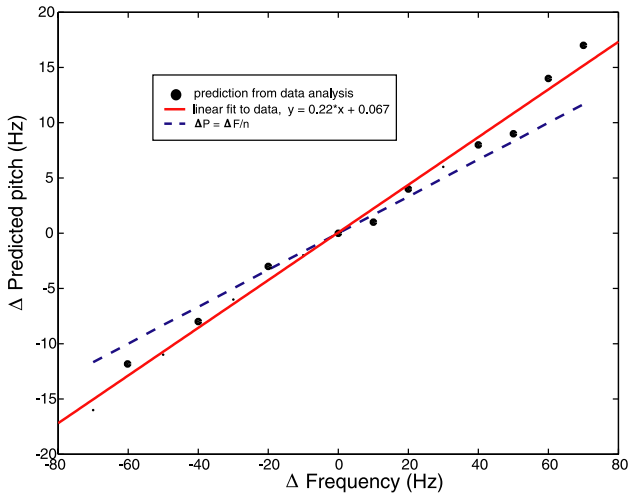

Fig. 7. Predicted pitch-shift of the residue as a function of frequency increments or decrements applied to a stimulus comprising the fourth, sixth, seventh, eighth harmonics of a 200 Hz F_0. Each stimulus was analyzed in the same manner as in the previous figures. In accord with psychophysical observation, the slope of the linear fit to the pitch shifts predicted from comparing each stimulus to the speech sound database is 22% greater than the slope of the line defined by $\Delta P = \Delta F/n_c$ (see text for details). Compare with Figs. 4 and 5 in Schouten et al. (1962).

These findings indicate that the probabilistic relationship between a stimulus and its natural sources accurately predicts the phenomenon of spectral dominance.

3.4. Pitch strength

The pitches listeners hear in response to complex tone stimuli of equal power differ not only in the sinusoidal frequency assigned to them but in the relative strength (sometimes called clarity or salience) of the pitch heard, a perception that varies as a function of both the spectral profile of the stimulus and its fundamental frequency (Fastl and Stroll, 1979; Terhardt et al., 1986). For example, stimuli 3 and 5 in Fig. 5(a) have the same fundamental frequency but differ in their spectral profiles. Stimuli comprising low harmonics (e.g., stimulus 3) typically evoke a stronger sense of pitch than do stimuli comprising higher harmonics (e.g., stimulus 5; Fastl and Stroll, 1979). Consistent with the hypothesis that pitch strength reflects the probabilistic relationship between a stimulus and its possible sources in the human auditory environment, the functions corresponding to stimuli 3 and 5 in Fig. 5(b) show that the mean cross-correlation corresponding to the maximum of the function (i.e., the height of the maximum) is greater for stimulus 3 than stimulus 5.

When the spectral profile is held constant and F_0 is varied, the change in pitch strength as a function of F_0 approximates an inverted U when frequency is plotted on a log scale, with maximum pitch strength occurring at frequencies in a broad region centered near 300 Hz, declining steeply with F_0 values both above and below this bandwidth (Terhardt et al., 1986; Huron, 2001). Fig. 10 shows F_0 plotted against mean cross-correlation with respect to the speech sound database for a set of stimuli comprising the first 10 harmonics of a given F_0. The fundamental frequencies plotted correspond to the musical notes C_1, G_1, C_2, G_2 . . .,C_8, and thus encompass almost the full the range of musical pitch. The mean cross-correlation values are greatest for F_0 values between ~200 and 500 Hz and decline steeply with F_0 both above and below this frequency region. (Note that the mean cross-correlation values are taken as a measure of relative, not absolute, pitch strength.)

Thus pitch strength is also well predicted by the probabilistic relationship between a stimulus and its possible natural sources.

3.5. Analysis based on sawtooth waves

A potential objection to this series of results is that they are simply a consequence of the fact that speech sounds are approximately periodic, the implication being that the same analysis applied to any database of periodic sound stimuli would do as well. It is important to note in this regard that the amplitude spectrum of a voiced speech sound does not decrease monotonically with frequency, as does the spectrum of artificial harmonic stimuli such as sawtooth or square waves. On the contrary, amplitude increases with frequency up to the first formant, which very rarely corresponds to F_0, and then decreases, with additional local amplitude maxima (formants) at higher frequencies (see, for example, Stevens, 1999). If pitch reflects the auditory system's experience with not only the characteristic frequency relationships in natural sources of periodic stimuli but, as expected, the characteristic amplitude relations among the spectral components as well, then accurate pitch estimates should *not* be derivable from analyses comparing test stimuli to complex periodic sounds whose spectral components lack these characteristic amplitude features.

We thus repeated the analyses using a set of sawtooth waves instead of speech sounds. As shown in Fig. 11, sawtooth waves comprise a complete harmonic series of spectral components in which amplitude decreases exponentially as a function of frequency. If the observed relationship between speech sounds and pitch is simply a consequence of the periodic character of speech sounds, then the phenomenology of pitch should be predicted just as accurately from the cross-correlation of the test stimuli and these non-speech periodic sounds. Note that this control does not involve comparing the cross-correlations derived from analyses involving speech sounds with the cross-correlations derived from analyses of sawtooth waves. Rather, it entails comparing the

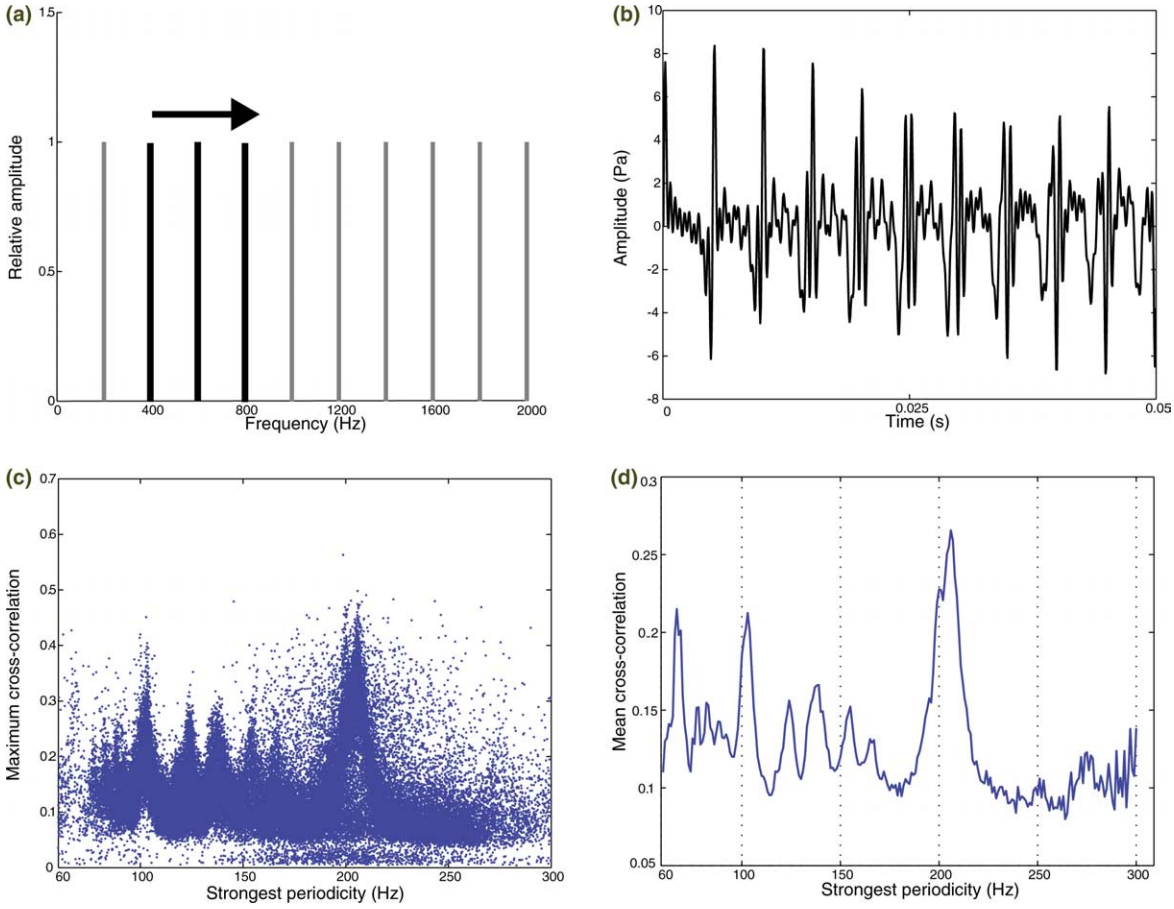

Fig. 8. Predicting the dominant influence of spectral components near 600 Hz in determining pitch. (a) Schematic frequency-domain representation of a stimulus comprising the first 10 harmonics of a stimulus with a 200 Hz F_0, in which the second, third, and fourth harmonics have been augmented by 3%. (b) Time-domain representation of the same stimulus. (c) The maximum cross-correlation coefficient of each speech sound in the database with respect the stimulus in (a) and (b), plotted against the strongest periodicity in each speech segment. (d) Average maximum cross-correlation for each integer frequency bin, derived from the data in (c). The periodicity associated with the maximum of the function in (d) is 206 Hz.

accuracy of the pitch predicted by the most likely speech sound, given the stimulus, with the accuracy of the pitch predicted by the most likely sawtooth wave, given the stimulus. Thus, the comparison of cross-correlation values occurs within and not between databases.

Fig. 12 shows the pitch frequency and strength predictions derived from a set of 241 50 ms sawtooth waves encompassing F_0 values from 60–300 Hz. An analysis comparing the test stimuli used earlier with sawtooth waves yields accurate pitch predictions for the missing fundamental and pitch shifted stimuli in Fig. 5 because the periodicity of the probable source of such stimuli happens to correspond to the F_0 of the stimulus itself. The more challenging comparisons are with test stimuli comprising spectral components that have no common divisor >70 Hz. We therefore restricted our analysis to the phenomena of spectral dominance and pitch strength, both of which pose particular difficulties for other pitch theories, but are accurately predicted by the probabilistic relationship between a stimulus and its natural sources.

In Fig. 12(a) the solid line shows the predicted changes in pitch as a function of the rank of the middle harmonic in the augmented subset for the spectral dominance stimuli depicted in Fig. 9(a) ($F_0 = 100$ Hz) based on comparing the stimuli to the set of sawtooth waves; the dashed line indicates the predicted changes in pitch for the same stimuli based on the analysis of the speech sound database (see Fig. 9(b)). The predictions derived from the sawtooth waves for stimuli comprising harmonics of a 100 Hz F_0 deviate by nearly a full octave from the psychophysical data; moreover the pertinent curve shows no clearly dominant spectral region. In Fig. 12(b) the solid line again shows predicted change in pitch as a function of the rank of the middle harmonic in the augmented subset for the spectral dominance stimuli depicted in Fig. 9(a), where $F_0 = 200$ Hz. The predictions derived from the sawtooth waves suggest a dominant spectral region centered near 900 Hz, which is again inconsistent with the psychophysical findings. In addition, the results are at odds with the observation that the harmonic rank of the dominant spectral com-

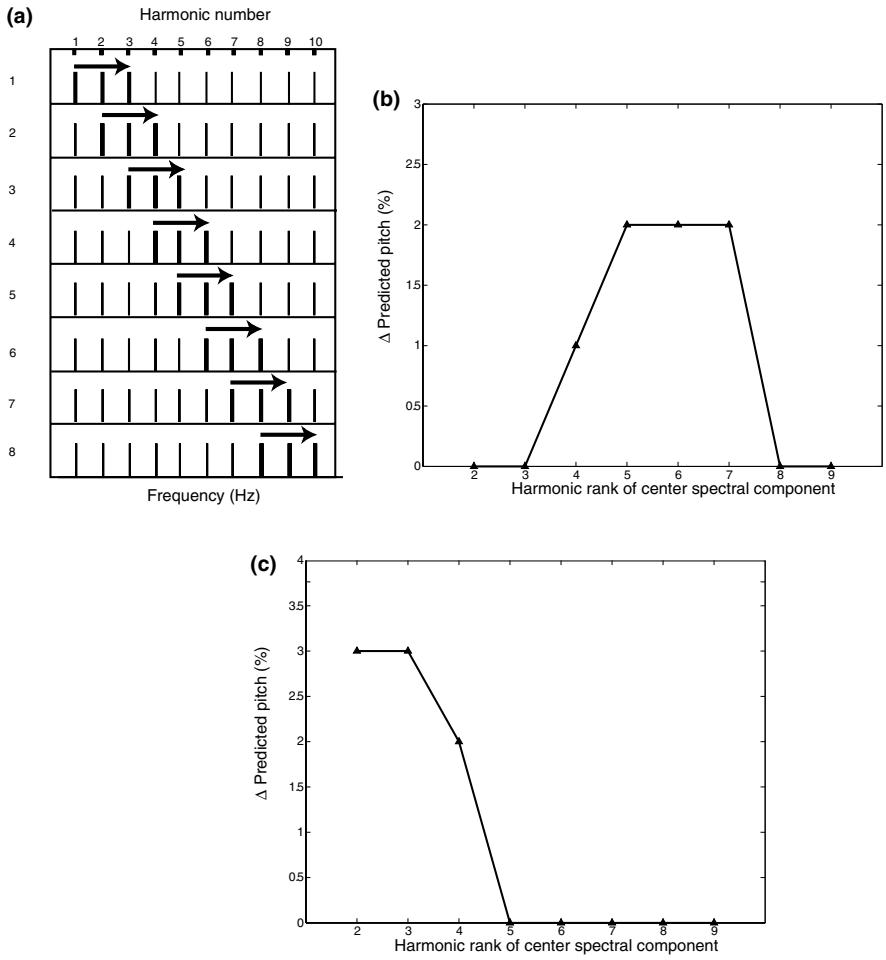

Fig. 9. Predicting the inverse relationship between the fundamental frequency of a stimulus and the harmonic rank of the dominant spectral components. (a) Schematic frequency-domain representation of the stimuli used. In each case, the frequency values of a different set of three successive harmonics (indicated by the dark lines beneath the arrows) were augmented by 3%. (b) Predicted pitch change plotted as a function of the harmonic number of the center component of the frequency-shifted set, for stimuli with a fundamental frequency of 100 Hz. (c) Predicted pitch change plotted as a function of the harmonic number of the center component of the frequency-shifted set, for stimuli with a fundamental frequency of 200 Hz. In accord with psychophysical observations (cf. Fig. 2 in Dai (2000)), the harmonic rank of the dominant spectral components (i.e., the components corresponding to the upward shift in predicted pitch) is inversely related to F_0, being higher for $F_0 = 100$ Hz than for $F_0 = 200$ Hz. For both fundamentals, predicted pitch changes in the direction of the frequency increment only when the augmented harmonics include the 600 Hz spectral component (i.e., the sixth harmonic for $F_0 = 100$ Hz and the third harmonic for $F_0 = 200$ Hz).

ponents in a stimulus varies inversely with F_0 (see above).

Finally, Fig. 12(c) compares the predictions derived from the set of sawtooth waves to the predictions shown in Fig. 10. The solid line indicates predicted relative pitch strength based on the sawtooth waves, and the dashed line indicates the predictions based on the speech sound database. The pitch strength predictions from the analysis of the sawtooth waves are again inconsistent with the perceptual responses to such stimuli (Terhardt et al., 1986; Huron, 2001).

4. Discussion

The phenomenon of pitch has traditionally been considered "the heart of hearing theory" (Plomp, 2002, p. 28), and has been the focus of much auditory research during the last century. Despite this effort, a unified and parsimonious theory that accounts for the complex phenomenology of pitch has not been forthcoming (see, for example, Yost, 2000, p. 199; Rossing et al., 2002, p. 131; Bernstein and Oxenham, 2003). The various results we report here all support the hypothesis that the pitch elicited by a stimulus is determined according to the conditional probability distribution of the different possible sources of the periodic sound energy in the stimulus. This probabilistic relationship between periodic stimuli and their possible sources accurately predicts a wide variety of the pitch phenomena that have been described, including the predominant pitch of stimuli comprising successive harmonics of fundamental frequencies in the ~60–300 Hz range, the pitch-shift of the residue, spectral dominance, and pitch strength.

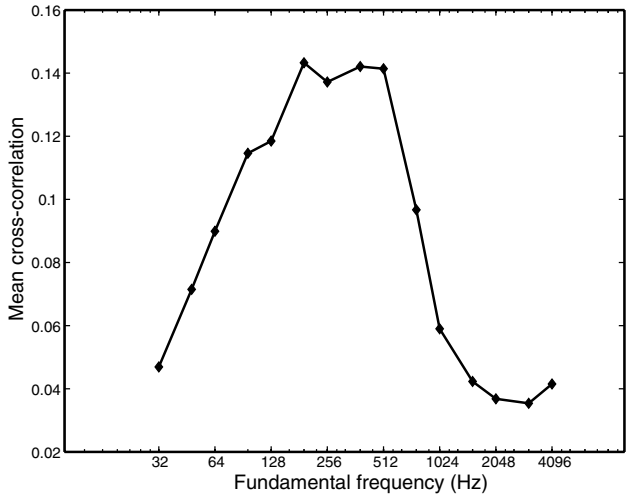

Fig. 10. Predicting the inverted-U function of pitch strength vs. F_0. The stimuli used in this analysis were the first 10 harmonics of periodic complex tones with fundamentals over the range of 32–4096 Hz. Graph shows the mean cross-correlation coefficient of the speech sounds in the database with respect to this series of stimuli, plotted against the fundamental frequency of the stimulus (note log scale). In accord with psychophysical observations (cf. Fig. 1 in Terhardt et al. (1982a) and Fig. 2 in Terhardt et al. (1986)) the mean cross-correlation, which in the present framework is the predictor of relative pitch strength, is greatest for F_0 values between 200 and 500 Hz, with cross-correlation values declining steeply both above and below this frequency region.

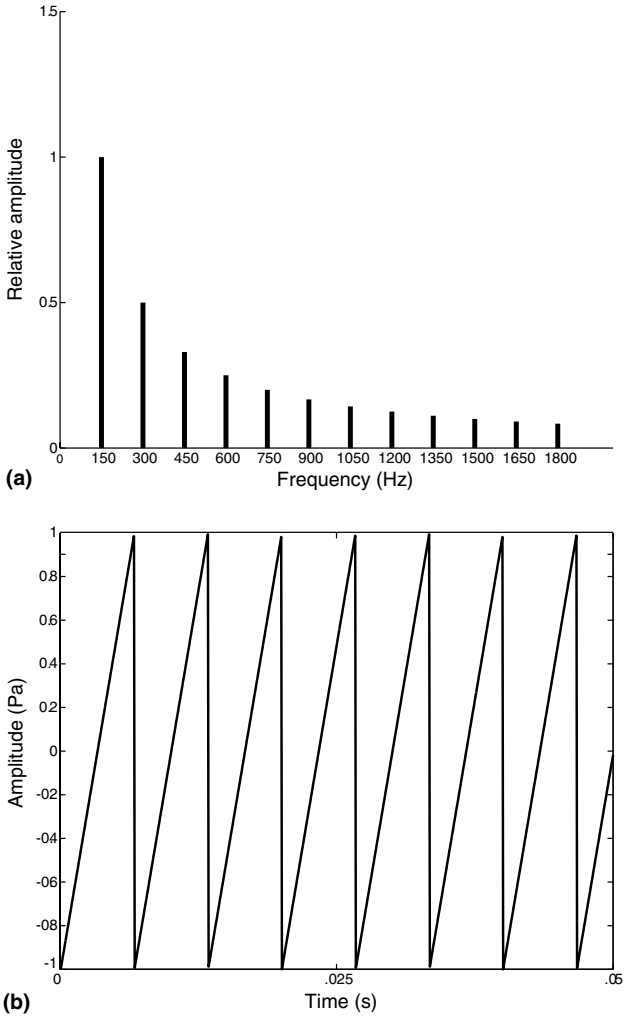

Fig. 11. Example of the signals used to assess whether accurate pitch predictions can be made by comparing the test stimuli we used to any set of periodic sounds. (a) Schematic frequency-domain representation of a 150 Hz sawtooth wave. The first 12 harmonics are shown. (b) Time-domain representation of the same stimulus. See text for further explanation.

Accurate predictions of spectral dominance and pitch strength in particular derive specifically from an analysis of the natural sources of periodic stimuli for human listeners (i.e., speech sounds), and could not be derived by a similar analysis of unnatural periodic signals (sawtooth waves). Nor are the speech segments that contributed to the maxima in each mean cross-correlation function simply 'copies' of the relevant artificial test stimuli, since the mean cross-correlation values associated with the maxima of these functions were all <0.5.

4.1. Previous theories of pitch

Earlier theories that can account for some aspects of the phenomenology of pitch can be grouped into three broad categories: spectral theories (e.g., Goldstein, 1973; Terhardt, 1974; Plomp, 1976), temporal theories (e.g., Schouten et al., 1962; Moore, 1973; Houtsma and Smyrzynski, 1990), and hybrid spectro-temporal theories (Moore, 1982; Meddis and Hewitt, 1991).

Spectral theories propose that pitch perception involves frequency analysis of a complex auditory stimulus into its sinusoidal components, followed by a pattern recognition process that estimates the fundamental frequency of the harmonic series that best fits the pattern of spectral energy in the stimulus. The pitch corresponding to this frequency is then taken to be what is heard. Some spectral theories (e.g., Goldstein, 1973; Gerson and Goldstein, 1978) share with the hypothesis advanced here an emphasis on the probabilistic nature of the auditory process that determines pitch. This framework successfully predicts the pitch of stimuli comprising successive harmonics. A major weakness of such theories, however, is the need to use empirically-derived spectral weighting functions to accurately predict the pitch listeners hear in response to pitch shifted stimuli. Weighting factors are also need to explain the phenomena of spectral dominance and pitch strength.

Temporal theories propose that pitch involves phase locking of auditory nerve fiber firing to the fine structure of the peripherally transduced stimulus waveform, pitch being determined by the time interval separating instances of coincident firing (see references above). Although such theories can account for both the pitch of successive harmonics and the pitch-shift of the residue,

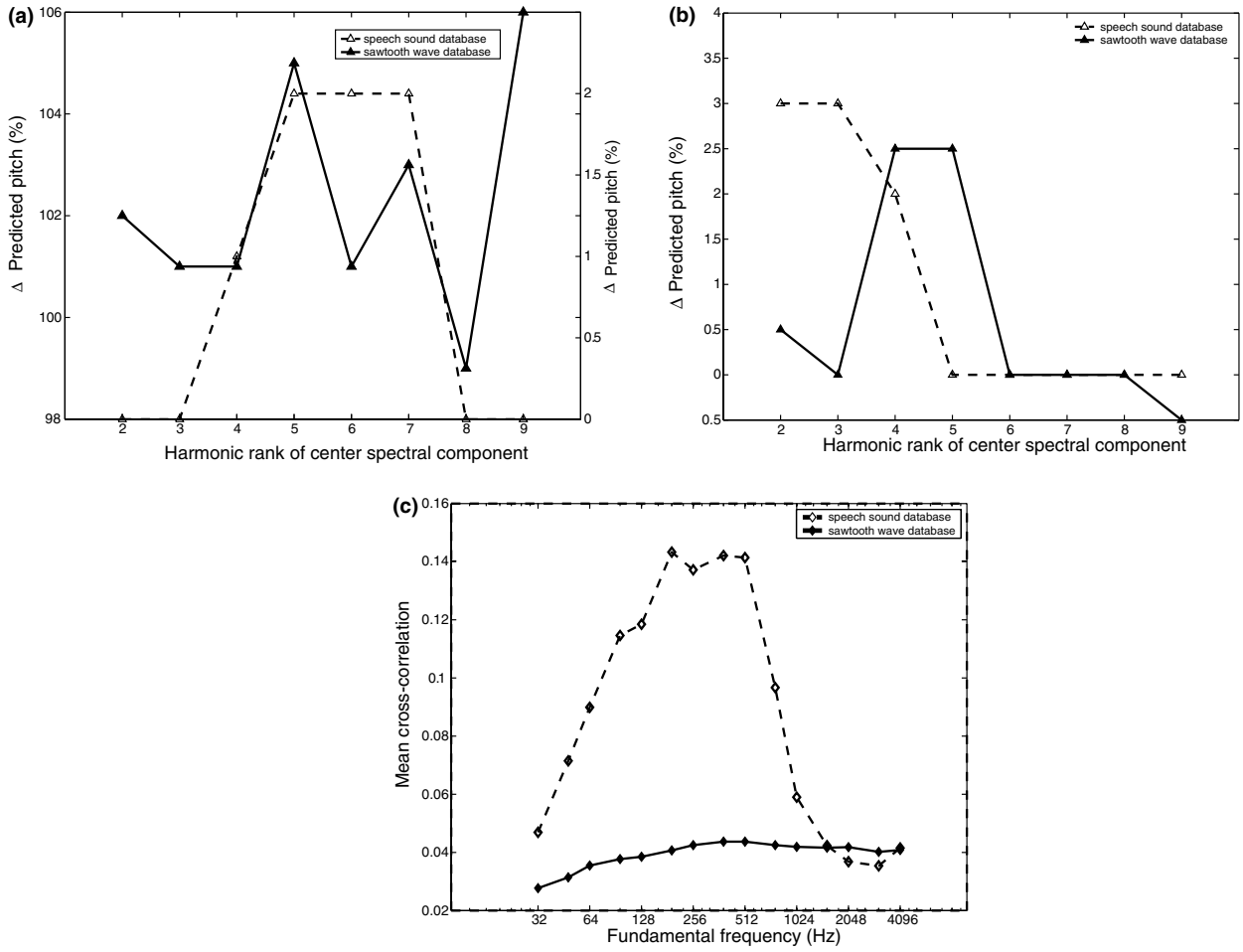

Fig. 12. The pitch predictions derived from a set of non-speech periodic sound segments (see Fig. 11) compared with the predictions derived from the speech sound database. In both (a) and (b), the solid line shows predicted change in pitch as a function of the rank of the middle harmonic in the augmented subset for the spectral dominance stimuli depicted in Fig. 9(a), based on comparing the stimuli with sawtooth waves. The dashed line indicates the predicted change in pitch for the same stimuli based on the analysis of the speech sound database (see Fig. 9(b)). (a) Results for $F_0 = 100$ Hz. (b) Results for $F_0 = 200$ Hz. (c) Solid line shows mean cross-correlation coefficient of the sawtooth waves in the database with respect to a given stimulus plotted against the fundamental frequency of the stimulus (note log scale). The stimuli used in this analysis were the first 10 harmonics of periodic complex tones with fundamentals over the range of 32–4096 Hz. The dashed line indicates the prediction based on the speech sound database (see Fig. 10).

they fail to explain spectral dominance and pitch strength (Moore, 1993). Finally, spectro-temporal theories, like spectral theories, need empirically derived spectral weighting functions (modeled, for example, as auditory filters with different output functions for low and high frequency harmonics; Moore, 1993) to explain spectral dominance and pitch strength.

Attempts to provide physiological support for current psychophysical models of pitch have also enjoyed only limited success. For example, Cariani and Delgutte (1996a,b) directly tested the 'predominant interval hypothesis' proposed in several spectro-temporal models of pitch (e.g., Meddis and Hewitt, 1991) but were unable to predict accurately the pitch-shift of the residue, the pitch strength of low F_0 stimuli or the rate pitch of low frequency alternating polarity pulse trains (see below). The predominant interval hypothesis, like most other pitch models, assumes that pitch is fundamentally a determination of stimulus periodicity, an assumption that is inconsistent with present evidence.

4.2. An empirical (probabilistic) approach

The different approach that we used here to examine the basis of pitch was suggested by the inevitably uncertain nature of the auditory stimulus-source relationship. As outlined in the Introduction, auditory stimuli, like visual stimuli, are inherently ambiguous: the physical characteristics of the stimulus at the ear do not, and cannot, specify the physical properties of the generative source. Nevertheless it is toward the stimulus sources that behavior must be directed if percepts are to be biologically useful. A wide range of recent work in vision is consistent with the hypothesis that sensory systems

meet the challenge of stimulus ambiguity by relating stimuli to their possible natural sources probabilistically (reviewed in Rao et al., 2002; Purves and Lotto, 2003; Purves et al., 2004). By generating percepts determined according to empirical stimulus-source associations rather than according to the physical properties of the stimuli as such, the listener brings past experience – incorporated in the nervous system by both natural selection and individual development – to bear on the quandary of stimulus ambiguity.

The fact that the varied phenomenology of pitch can be successfully predicted from the probabilistic relationship between tone-evoking stimuli and their sources in the human auditory environment suggests that the same statistical process that underlies the perception of brightness, color, motion, and spatial relationships in vision (see references above) and tonal consonance in music (Schwartz et al., 2003) also underlies the perception of pitch. Whereas theories of tone perception that proceed from the assumption that pitch is the perception of stimulus characteristics per se find it necessary to include a variety of additional free parameters and/or theoretical constructs (e.g., harmonic templates, spectral weighting functions, neural coincidence detectors) to generate accurate predictions of pitch, the hypothesis that pitch is determined by the probabilistic relationship between a stimulus and its possible natural sources provides a complete and parsimonious rationale for this complex variety of perceptual effects.

4.3. Relevance to additional aspects of pitch

Conceiving pitch in these terms may also explain several more general aspects of tone perception not specifically examined here. For example, the question of whether the auditory system uses one or multiple mechanisms to determine pitch remains a matter of debate (Goldstein, 2000). The pitches heard in response to alternating polarity click trains is often cited as evidence for the existence of two qualitatively different pitch mechanisms (Flanagan and Gutman, 1960; Pierce, 1991, 2001). Specifically, when listeners are presented with alternating polarity click trains, pitch frequency typically corresponds to stimulus click rate for rates $<$100 Hz, but to the F_0 of the stimulus when the click rate exceeds $\sim$500 Hz. That is, pitch corresponds to click rate at low frequencies and to F_0 at high frequencies. Under the assumption that pitch is determined by the extraction of stimulus periodicity or F_0, no single mechanism can account for these data.

If, however, pitch is the perception of the periodicity of the probable natural sources of any stimulus rather than the extraction of stimulus periodicity as such, these findings can be readily rationalized. Recall that the 60–300 Hz bandwidth includes the dominant periodicities of $\sim$97% of voiced speech sounds (see Fig. 2). If the prior probability of pitch frequencies corresponds to the empirical distribution of periodicities in these natural periodic sounds, then the pitches that listeners hear will be strongly biased toward frequencies in the 60–300 Hz range. As a result, when click rate and F_0 differ, the pitch heard will correspond to the frequency that lies closest to the F_0 range of speech sounds. At low click rates, the fundamental lies far outside the range of speech sound F_0s; consequently, a source periodicity corresponding to the click rate is more likely. Conversely, at high click rates, the stimulus F_0 lies within the F_0 range of speech sounds, whereas click rate lies outside the range of speech sound F_0s; consequently a source periodicity corresponding to the fundamental of the stimulus is more probable.

Other aspects of pitch phenomenology that can be explained in this probabilistic framework are the absence of residue pitch for stimuli with $F_0 >\sim 1000$ Hz (Fletcher, 1924) and the phenomenon of pitch ambiguity (i.e., the identification of two or more distinct pitches in response to the same complex tone; Schouten et al., 1962; Terhardt et al., 1982b). The statistical property of natural periodic sounds relevant to explaining the first of these phenomena is that 1000 Hz is roughly the upper limit of the fundamental frequencies that the human vocal apparatus produces (i.e., C_6 on the musical scale). Applying the statistical analysis here to a speech sound database that included the full range of fundamental frequencies produced by human vocalization would presumably predict the residue or virtual pitches $>$300 Hz that humans hear (see also Terhardt et al., 1982b).

With respect to pitch ambiguity, the mean cross-correlation functions we derived exhibit a variety of peaks with different heights. Whereas the highest peak in the mean cross-correlation vs. periodicity function predicts the dominant pitch that listeners typically hear, the periodicities associated with the subsidiary peaks in each plot would presumably predict other pitches that listeners identify less frequently in response to a given stimulus. For example, although the pitch most often reported in response to the stimulus shown in Figs. 7(a) and (b) is $\sim$204 Hz, some listeners (or the same listener on a different occasion) instead report hearing pitches corresponding to sinusoidal frequencies of 185 Hz and 227 Hz (Moore, 1993). The periodicities associated with the peaks to the immediate left and right of the maximum at 204 Hz are 185 Hz and 227 Hz, respectively (see Fig. 7(d)).

4.4. Physiological implications

Several recent studies of the neural basis of pitch perception in humans have suggested that the parameters of an acoustical signal are spatially mapped onto regions of the auditory cortex, such that adjacent cortical locations represent adjacent values of the parameter

(Langner et al., 1997; Cansino et al., 2003; Fujioka et al., 2003). There is no consensus, however, as to what acoustical parameter(s) this mapping signifies. Langner et al. (1997), for example, proposed that the auditory cortex is organized according to both frequency (i.e., "tonotopy") and periodicity (i.e., "periodotopy"), and that tonotopic and periodotopic gradients are arranged orthogonally. Cansino et al. (2003), however, found little support for the existence of a periodotopic organization and proposed instead that the topographical organization of auditory cortex signifies the spectral content of acoustical signals, while stimulus periodicity is coded temporally. Fujioka et al. (2003) also failed to find support for an independent spatial mapping of stimulus periodicity. These latter authors suggested that cortical representation of the pitch of complex tones is likely to entail multiple spatial representations of stimulus parameters, as well as temporal coding mechanisms.

In general, physiological studies of the neural basis of pitch perception have been guided by the spectral and temporal theories of pitch described above. Thus, studies have been undertaken to find the neural correlates of the frequency analysis and periodicity coding processes that various psychoacoustical theories have proposed, and experimental phenomena have been interpreted in terms of the hypothetical constructs such theories advance. If, as we have argued, pitch can be more fully understood in terms of the probabilistic relationship between auditory stimuli and their natural sources, then these same physiological findings will need to be interpreted in terms of this new conceptual framework. To speculate about the means by which the auditory system embodies the statistical characteristics of the human auditory environment would be premature; given the present results, however, it is certainly of interest to now ask how the known facts of auditory system anatomy and physiology can be rationalized in these terms.

4.5. Conclusion

The inherent ambiguity of any sound pressure change at the ear precludes, in principle, a determination of source signal periodicity based on any analytical process applied to features that might be abstracted from the stimulus. The biological solution to this quandary must therefore be a statistical one. We have shown here that much of the known phenomenology of pitch can be accurately predicted from the conditional probability distribution of the possible sources of the periodic sound energy in a given stimulus. Thus the perception of pitch, like the perception of brightness, color and form in vision, is best understood in terms of the probabilistic relationship between a stimulus and its possible sources in nature. The information about the statistical characteristics of the human auditory environment derives from the past behavioral experience of both the species and the individual, and must be substantiated in auditory processing circuitry.

Acknowledgements

We are grateful to Catharine Howe, Fuhui Long, Rich Mooney, Surajit Nundy, Bob Peters, Liwei Sha, Zhiyong Yang and two anonymous reviewers for helpful criticism, Nikos Pitsianis for programming assistance, and the Duke Visualization Analysis Lab for the use of computing facilities.

References

Bernstein, J.G., Oxenham, A.J., 2003. Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? J. Acoust. Soc. Am. 113, 3323–3334.

de Boer, E., 1956. On the "residue" in hearing. Unpublished doctoral dissertation. University of Amsterdam.

Boersma, P., 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proceedings of the Institute of Phonetic Sciences 17, 97–110.

Boersma, P., Weenink, D., 2003. PRAAT 4.1: Doing phonetics by computer. Department of Phonetic Sciences, University of Amsterdam. [There is no print version; download is available at http://fonsg3.let.uva.nl/praat/].

Cansino, S., Ducorps, A., Ragot, R., 2003. Tonotopic cortical representation of periodic complex sounds. Hum. Brain Mapp. 20, 71–81.

Cariani, P.A., Delgutte, B., 1996a. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J. Neurophys. 76, 1698–1716.

Cariani, P.A., Delgutte, B., 1996b. Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. J. Neurophys. 76, 1717–1734.

Cook, P., 1999. Pitch, periodicity, and noise in the voice. In: Cook, P. (Ed.), Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics. MIT Press, Cambridge, MA, pp. 195–208.

Dai, H., 2000. On the relative influence of individual harmonics on pitch judgment. J. Acoust. Soc. Am. 107, 953–959.

Fastl, H., Stoll, G., 1979. Scaling of pitch strength. Hear. Res. 1, 293–301.

Fisher, W.M., Doddington, G.R., Goudie-Marshall, K.M., 1986. The DARPA speech recognition research database: specifications and status. Proceedings of the DARPA Speech Recognition Workshop, Report No. SAIC-86/1546.

Flanagan, J.L., Gutman, N., 1960. On the pitch of periodic pulses. J. Acoust. Soc. Am. 32, 1308–1319.

Fletcher, H., 1924. The physical criterion for determining the pitch of a tone. Phys. Rev. 23, 427–437.

Fromkin, V., Rodman, R., 1998. An Introduction to Language, sixth ed. Harcourt Brace, Fort Worth.

Fujioka, T., Ross, B., Okamoto, H., Takeshima, Y., Kakigi, R., Pantev, C., 2003. Tonotopic representation of missing fundamental complex sounds in the human auditory cortex. Eur. J. Neurosci. 18, 432–440.

Garofolo, J.S., Lamel, L.F., Fisher, W.M., Fiscus, J.G., Pallett, D.S., Dahlgren, N.L., 1990. DARPA-TIMIT Acoustic-phonetic contin-

uous speech corpus [CD-ROM]. US Department of Commerce, Gaithersburg, MD.
Gerson, A., Goldstein, J.L., 1978. Evidence for a general template in central optimal processing for pitch of complex tones. J. Acoust. Soc. Am. 63, 498–510.
Goldstein, J.L., 1973. An optimum processor theory for the central formation of the pitch of complex tones. J. Acoust. Soc. Am. 54, 1496–1516.
Goldstein, J.L., 2000. Pitch perception. In: Kazdin, A. (Ed.), Encyclopedia of Psychology. Oxford University Press, Oxford.
Gordon, C., Webb, D., Wolpert, S., 1992. One cannot hear the shape of a drum. Bull. Am. Math. Soc. 27, 134–138.
Hall, J.W., Peters, R.W., 1981. Pitch for nonsimultaneous successive harmonics in quiet and noise. J. Acoust. Soc. Am. 69, 509–513.
Houtsma, A.J.M., Smyrzynski, J., 1990. Pitch identification and discrimination for complex tones with many harmonics. J. Acoust. Soc. Am. 87, 304–310.
Huron, D., 2001. Tone and voice: a derivation of the rules of voice leading from perceptual principles. Music Percept. 19, 1–64.
Jarveläinen, H., Verma, T., Välimäki, V., 2002. Perception and adjustment of pitch in inharmonic string instrument tones. J. New Music Res. 31, 311–319.
Langner, G., Sams, M., Heil, P., Schulze, H., 1997. Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. J. Comp. Physiol. 181, 665–676.
Lamel, L.F., Kassel, R.H., Seneff, S., 1986. Speech database development: design and analysis of the Acoustic-Phonetic Corpus. Proceedings of the DARPA Speech Recognition Workshop, Report No. SAIC-86/1546.
Licklider, J.C.R., 1954. "Periodicity" pitch and "place" pitch. J. Acoust. Soc. Am. 26, 945.
Lieberman, P., Blumstein, S.E., 1988. Speech Physiology, Speech Perception, and Acoustic Phonetics. Cambridge University Press, New York.
Meddis, R., Hewitt, M., 1991. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. J. Acoust. Soc. Am. 89, 2866–2882.
Moore, B.C.J., 1973. Some experiments relating to the perception of complex tones. Quart. J. Exp. Psychol. 25, 451–475.
Moore, B.C.J., 1982. An Introduction to the Psychology of Hearing, second ed. Academic Press, London.
Moore, B.C.J., 1993. Frequency analysis and pitch perception. In: Yost, W.A., Popper, A.N., Foy, R. (Eds.), Human Psychophysics. Springer, New York.
Moore, B.C.J., Glasberg, B.R., Peters, R.W., 1985. Relative dominance of individual partials in determining the pitch of complex tones. J. Acoust. Soc. Am. 77, 1853–1860.
Pierce, J., 1991. Periodicity and pitch perception. J. Acoust. Soc. Am. 90, 1889–1893.
Pierce, J., 2001. Introduction to pitch perception. In: Cook, P. (Ed.), Music, Cognition and Computerized Sound. MIT Press, Cambridge, MA.
Plomp, R., 1967. Pitch of complex tones. J. Acoust. Soc. Am. 41, 1526–1533.
Plomp, R., 1976. Aspects of Tone Sensation. Academic Press, London.
Plomp, R., 2002. The Intelligent Ear: On the Nature of Sound Perception. Erlbaum, Mahwah, NJ.
Purves, D., Lotto, R.B., 2003. Why We See What We Do: Evidence For An Empirical Theory of Vision. Sinauer Associates, Sunderland, MA.
Purves, D., Lotto, R.B., Nundy, S., Williams, S.M., 2004. Perceiving brightness. Psych. Rev. 111, 142–158.
Rao, R.P.N., Olshausen, B.A., Lewicki, M.S. (Eds.), 2002. Probabilistic Models of the Brain: Perception and Neural Function. MIT Press, Cambridge, MA.
Rasch, R., Plomp, R., 1999. The perception of musical tones. In: Deutsch, D. (Ed.), The Psychology of Music, second edition. Academic Press, New York.
Ritsma, R.J., 1967. Frequencies dominant in the perception of the pitch of complex sounds. J. Acoust. Soc. Am. 42, 191–198.
Ritsma, R.J., 1970. Periodicity detection. In: Plomp, R., Smoorenburg, G.F. (Eds.), Frequency Analysis and Periodicity Detection in Hearing. AW Sijthoff, Leiden, The Netherlands, pp. 250–263.
Rossing, T.D., Moore, F.R., Wheeler, P.A., 2002. The Science of Sound, third ed. Addison-Wesley, Reading, MA.
Schouten, J.F., 1938. The perception of subjective tones. Proc. K. Ned. Akad. Wetensc. 34, 1086–1093.
Schouten, J.F., 1940. The residue, a new component in subjective sound analysis. Proc. K. Ned. Akad. Wetensc. 43, 356–365.
Schouten, J.F., Ritsma, R.J., Cardozo, B.I., 1962. Pitch of the residue. J. Acoust. Soc. Am. 34, 1418–1424.
Schwartz, D.A., Howe, C.Q., Purves, D., 2003. The statistical structure of human speech sounds predicts musical universals. J. Neurosci. 23, 7160–7168.
Seebeck, A., 1841. Beobachtungenuber über einige Bedingungen der Entschung von Tönen. Ann. Phys. Chem. 53, 417–436.
Smoorenburg, G.F., 1970. Pitch perception of two-frequency stimuli. J. Acoust. Soc. Am. 48, 924–942.
Stevens, K.N., 1999. Acoustic Phonetics. MIT Press, Cambridge, MA.
Tarantola, A., 1987. Inverse Problem Theory. Elsevier, Leiden, The Netherlands.
Terhardt, E., 1974. Pitch, consonance, and harmony. J. Acoust. Soc. Am., 55,1061–1069.
Terhardt, E., Stoll, G., Schermbach, R., Parncutt, R., 1986. Tonhöhenmehrdeutigkeit, Tonverwandtschaft und Identifikation von Sukzessivintervallen. Acustica 61, 57–66.
Terhardt, E., Stoll, G., Seewann, M., 1982a. Pitch of complex signals according to virtual-pitch theory: tests, examples, and predictions. J. Acoust. Soc. Am. 71, 671–678.
Terhardt, E., Stoll, G., Seewann, M., 1982b. Algorithm for extraction of pitch and pitch salience from complex tonal signals. J. Acoust. Soc. Am. 71, 679–688.
Yost, W.A., 2000. Fundamentals of Hearing: An Introduction. Academic Press, San Diego.
Zucker, S., 2003. Cross-correlation and maximum likelihood analysis: a new approach to combine cross-correlation functions. Mon. Not. R. Astron. Soc., 342,1291–1298.