Virtual pitch in a computational physiological model

Ray Meddis[a] and Lowel P. O'Mard
Department of Psychology, Essex University, Colchester, CO4 3SQ, United Kingdom

(Received 31 May 2006; revised 28 September 2006; accepted 2 October 2006)

A computational model of nervous activity in the auditory nerve, cochlear nucleus, and inferior colliculus is presented and evaluated in terms of its ability to simulate psychophysically-measured pitch perception. The model has a similar architecture to previous autocorrelation models except that the mathematical operations of autocorrelation are replaced by the combined action of thousands of physiologically plausible neuronal components. The evaluation employs pitch stimuli including complex tones with a missing fundamental frequency, tones with alternating phase, inharmonic tones with equally spaced frequencies and iterated rippled noise. Particular attention is paid to differences in response to resolved and unresolved component harmonics. The results indicate that the model is able to simulate qualitatively the related pitch-perceptions. This physiological model is similar in many respects to autocorrelation models of pitch and the success of the evaluations suggests that autocorrelation models may, after all, be physiologically plausible. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2372595]

PACS number(s): 43.66.Ba, 43.66.Hg, 43.66.Nm [BLM] Pages: 3861–3869

I. INTRODUCTION

Virtual pitch is a sensation of tone height created by broadband sounds. This pitch can be compared to the tone height of a pure tone even though the broadband stimulus may not contain any spectral energy in the frequency region of that tone. There are many different explanations of the origin of virtual pitches and these are often categorized as either "spectral" or "temporal." The model described below addresses a group of hypotheses in the "temporal" category (for example, Licklider, 1951; Yost *et al.*, 1978; Lyon, 1984; Assmann and Summerfield, 1990; Meddis and Hewitt, 1991; Meddis and O'Mard, 1997). These theories assume that virtual pitch arises from temporal regularity or periodicities in sounds and that these regularities can be characterized using mathematical methods such as autocorrelation. This approach makes many qualitatively detailed and often successful predictions concerning the outcome of a wide range of virtual pitch experiments, particularly for stimuli where the periodicity is invariant.

The status of the autocorrelation models remain controversial, however, because to many it appears to be physiologically implausible and positive evidence in its favor is elusive. There are no structures in the auditory brainstem that look capable of carrying out the "delay-and-multiply" operations required by autocorrelation. This is a major impediment to a general acceptance of an autocorrelation-type model of virtual pitch perception.

This paper seeks to address the issue by proposing a signal-processing scheme based on anatomical structures and physiological processes known to exist in the auditory periphery and brainstem. The aim is to show that a system based on these components can behave in some important respects like autocorrelation and can simulate a number of the most important virtual pitch phenomena.

The model is an extended version of an older computational model that was originally developed to simulate the response of single units in the auditory brainstem to sinusoidally amplitude modulated (SAM) tones (Hewitt and Meddis, 1992, 1994). In those studies it was shown that the model is able to simulate appropriate modulation transfer functions (MTFs) in ventral cochlear nucleus (VCN) neurons when stimulated using sinusoidal amplitude modulated (SAM) tones. It can also simulate appropriate *rate* MTFs in single inferior colliculus (IC) neurons. In this report the model will be expanded and evaluated using broadband stimuli with a pitch quality rather than SAM tones. An earlier study has shown that model VCN units can successfully simulate response of their physiological equivalents to broadband pitch stimuli (Wiegrebe and Meddis, 2004).

SAM tones do not themselves produce a strong pitch percept. It may seem strange that a model designed to process such stimuli should be regarded as a candidate explanation for pitch perception: However, the model's sensitivity to particular rates of amplitude modulation is based on its ability to respond to particular periodicities in the stimulus. Therefore, it is possible that it could also prove useful in responding to the periodicities present in pitch-evoking stimuli. This was the motivation for extending and re-evaluating the model.

The original model simulated nervous activity in a single best-frequency (BF) channel, i.e., a single site along the basilar membrane (BM). The new model is an expanded version that includes multiple parallel BF channels and many IC units in each channel. An additional cross-channel processing stage has been added to allow information to be aggregated across channels. Nevertheless, the basic component model of IC unit functioning is unchanged from its previously published form in all material respects.

It was necessary to include the aggregation of information across BF-channels (across BM sites) because the strongest sensation of pitch is created by stimuli whose frequency

[a]Electronic mail:rmeddis@essex.ac.uk

components are resolved by the auditory periphery, i.e., each component is processed in a different channel. For example, most IC units have only narrow receptive fields, sometimes responding to only one harmonic of a multiharmonic stimulus. It follows that a complete account of the extraction of the pitch of a multitone complex must involve a stage where the activity of units with different BFs is somehow combined. As a consequence, it is essential to add another layer of units above the narrowly-tuned units of the IC. This does not mean that the location of this layer is known, only that its inclusion was unavoidable.

The overall architecture of the model is the same as that of an autocorrelation model (Meddis and O'Mard, 1997a). It consists of a cascade of the following stages: (1) peripheral segregation of sound into frequency bands, (2) extraction of periodicities on a within-channel basis, and (3) aggregation of periodicity information across BF-channels. The novelty in the model lies in the way in which periodicity is extracted; using physiologically plausible circuits rather than an artificial mathematical device.

The model will be evaluated below by assessing its response to a range of stimuli commonly used in psychophysical experiments. These pitch-evoking stimuli include (1) complex tones with a missing fundamental frequency, (2) tones with alternating phase, (3) inharmonic tones with equally spaced frequencies, and (4) iterated rippled noise. The range of possible tests is virtually infinite and these stimuli have been carefully chosen to represent stimuli that raise difficulties for many pitch theories. Particular attention will be paid, for example, to differences in response to resolved and unresolved component harmonics. These stimuli have also been used previously to evaluate the autocorrelation model of pitch perception. The aim here is to demonstrate that a physiologically realizable circuit can substitute for autocorrelation and, qualitatively at least, it can perform just as well.

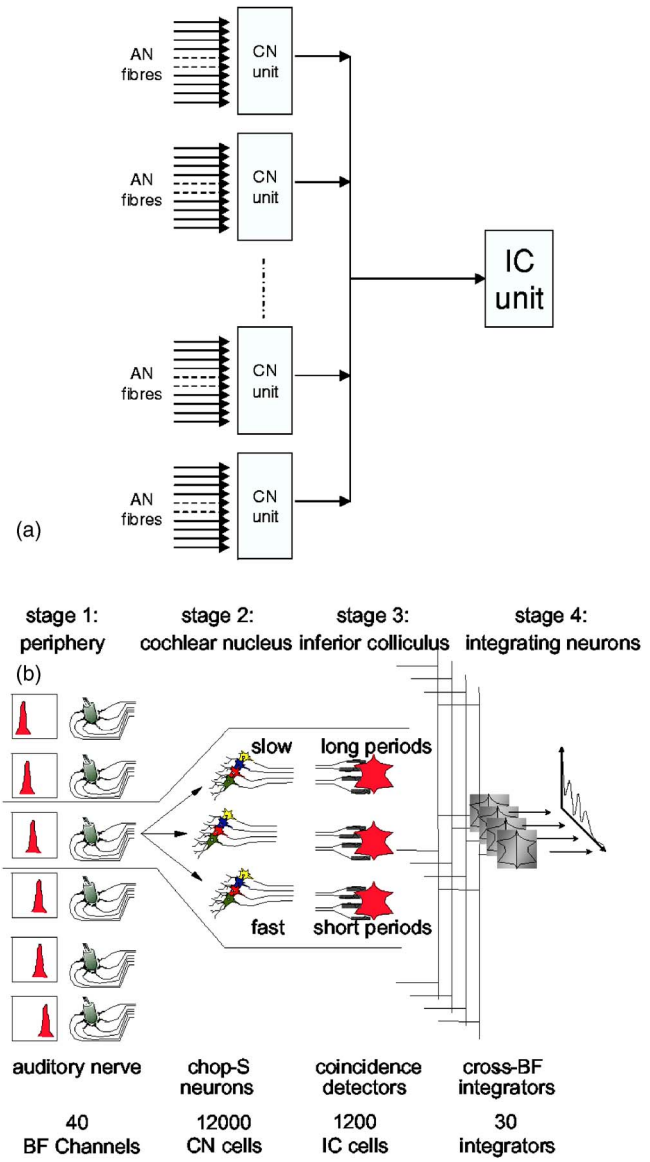

FIG. 1. (A) A single constituent module. A module consists of 10 VCN units feeding one IC unit. Each VCN unit receives input from 30 AN same-BF fibers. Within a module, all VCN units have the same saturated firing rate. (B) Arrangement of modules. There are 30 modules within each BF channel, each with a different characteristic firing rate. There are 40 channels with BFs ranging from 100 to 10000 Hz. A stage-4 unit receives one input from each channel from modules with the same saturated rate.

II. THE MODEL

The model contains thousands of individual components but is modular in structure [Fig. 1(A)]. The basic building block of the system is a module consisting of a cascade of three stages: auditory nerve (AN) fibers, VCN units, and an IC unit. Each module has a single IC cell receiving input from 10 VCN units *all with the same BF and saturated firing rate*. Each VCN unit receives input from 30 AN fibers *all with the same BF*. All modules are identical except for the BF and the saturated firing rate of the VCN units. Within a module, it is the saturated firing rate of the VCN units that determines the selectivity of the IC *rate* response to periodicity. The VCN units are modeled on VCN chopper units that fire at a fixed rate in response to moderately intense acoustic stimulation.

The VCN units are designed to imitate the behavior of sustained chopper units. These have low thresholds, a narrow dynamic range and a receptive field similar in width to that of an AN fiber (Blackburn and Sachs, 1989). The firing rate of VCN units is maximal about 20 dB above threshold and is largely independent of stimulus characteristics. As a consequence, the saturated chopping rate is referred to below as its *intrinsic* chopping rate. Different units have different intrinsic rates and these differences are fundamental to the ability of the model to detect and discriminate among different periodicities. The assumption that all units receive the same number of AN inputs is a simplifying assumption that may need to be revised at a future date. The assumption that all inputs to a VCN unit are from AN fibers with the same BF may also be a simplification. However, the narrowness of the receptive field of sustained chopper units is typically reported to be similar to that of a single auditory nerve fiber.

These modules are the same as those described in Hewitt and Meddis (1994) where it was shown that the VCN units

have appropriate modulation transfer functions (MTFs) and it was also shown that the module IC units have physiologically appropriate *rate*-MTFs. In other words the IC units show high firing rates for a restricted range of low-frequency modulations. These two MTFs are related in that the tuning of the IC unit depends on the chopping rate of VCN units supplying its input. If the intrinsic chopping rate of the VCN units is 100 spikes/s, the peak of the VCN MTF will be around 100 Hz and the peak of the IC *rate*-MTF is also at 100 Hz. A single IC unit can, therefore, be tuned to respond maximally to a particular modulation frequency by adjusting the chopping rate of its feeder VCN units.

The extended model replicates this core module many times within a single BF channel using different chopping rates in different modules. There are 10 VCN units in each block and, within a single channel, there are 30 blocks, each characterised by its chopping rate. This arrangement is replicated across 40 BF channels making a total of 12000 VCN units and 1200 IC units. The choice of 30 blocks and 40 BF channels is arbitrary and represents a trade-off against the considerable computational time involved.

This multirate, multi-BF architecture provides the basis for a further extension of the model; a fourth stage where periodicity information is aggregated *across* BF channels. This pitch-extraction stage is added to the model in the form of an array of hypothetical "stage 4" units called the "stage 4 profile." Each stage 4 unit receives input from one IC unit from each BF channel where all the contributing IC units have the same best modulation frequency and, hence, the same VCN chopping rate. Figure 1(B) illustrates the overall architecture. Each stage 4 unit has a unique pattern of sensitivities to periodicity.

The within-module details are unchanged from the original modeling studies. One minor change from earlier work is a reduction from 60 to 30 in the number of AN fibers providing input to a single VCN unit. This was to improve the fit of the model VCN MTFs to published animal data (Meddis and O'Mard, 1997b). Improvements to the detail of the AN model have also been included in order to be consistent with recent modeling work from this laboratory. The AN section of the model has, for example, been recently updated and is fully described in Meddis (2006). This updated version includes a nonlinear cochlear response and improved forward masking behavior (Meddis and O'Mard, 2005) and first-spike latencies (Meddis, 2006). The parameters of the dual resonance nonlinear (DRNL) filterbank are based on human psychophysical data (Lopez-Poveda and Meddis, 2001).

Stage 1: Auditory periphery. This stage of the model is exactly as described in Meddis (2006) where all equations and parameters are given in full. The effect of the outer-middle ear was simulated using a Butterworth first-order band pass filter with lower and upper cutoffs of 450 and 5000 Hz, respectively. The basilar membrane was modeled as 40 channels whose best frequencies (BFs) were equally spaced on a log scale across the range 100–10000 Hz. This was implemented as an array of dual resonance nonlinear (DRNL; Meddis *et al.*, 2001) filters using the parameters for human listeners given in Lopez-Poveda and Meddis (2001). All fibers were modeled identically, except for stochasticity, as high spontaneous rate (HSR) fibers with thresholds at 0 dB SPL at 1 kHz. The output of the auditory periphery was a stochastic stream of independent spike events in each auditory nerve fiber.

Stage 2: VCN units. These are implemented as modified McGregor cells (MacGregor, 1987) and are fully described in Hewitt and Meddis (1992) and Meddis (2006). Brief computational details and parameters are given in the Appendix. Each unit receives 30 AN fibers all with the same BF.

The saturated chopping rate of the VCN units is determined in the model by the potassium recovery time constant (τGk). This time constant is varied systematically across the array of units in such a way as to produce 30 different chopping rates equally spaced between 60 and 350 spikes/s. These values were chosen to be realistic with respect to physiological observations. The appropriate values of τGk were determined with an empirically derived formula; tau $=\text{rate}^{-1.441}$ based on the pure tone response at 65 dB SPL. Time constants varied between 2.8 ms (50 Hz) and 0.22 ms (350 Hz).

Stage 3: IC units. These are described in full in Hewitt and Meddis (1994) and are implemented here using the same MacGregor algorithm as used for the VCN units and using the parameters given in the Appendix. A single IC unit receives input from 10 VCN units. It is a critical (and speculative) feature of the model that each IC unit receives input only from VCN units with the *same intrinsic chopping rate*. The thresholds of the IC units are set to require coincidental input from many VCN units.

Stage 4 units. These units receive input from 40 IC units (one per BF-channel). All inputs to a single unit have the same rate-MTF as determined by the intrinsic saturated rate of the VCN units feeding the IC unit. It is assumed that each spike input to the stage 4 unit provokes an action potential. Therefore, stage 4 units are not coincidence detectors but simply counters of all the spikes occurring in their feeder units. They are intended purely as an indicator of the across-channel activity in the IC units tuned to a particular periodicity irrespective of BF channel. There are 30 stage 4 units, one for each VCN rate. The output of the model is, therefore, an array of 30 spike counts called the "stage-4 profile." The stage-4 profile rates are based on spike counts across the whole stimulus.

This profile is illustrated in Fig. 2 where it shows the response of the model to a simple harmonic stimulus. As the fundamental frequency of the stimulus rises, the profile shifts to the right. The change in pitch is coded in the change in the overall location of the profile. All profiles are the result of a single presentation of the stimulus. It must be stressed that this profile cannot be read as a pitch-meter that indicates a specific pitch frequency (e.g., 200 Hz). This is not a requirement of models of human pitch perception. The profile reflects the response of the whole system to a pitch-evoking stimulus. All of the values in the profile change when the pitch changes. In this sense, the pitch information is distributed throughout the profile which is used to detect *changes* in pitch as measured in most psychophysical experiments.

The model was evaluated at a refresh rate of 44 100 kHz on a dedicated Appro-4144H 4U Quad Opteron Server, a

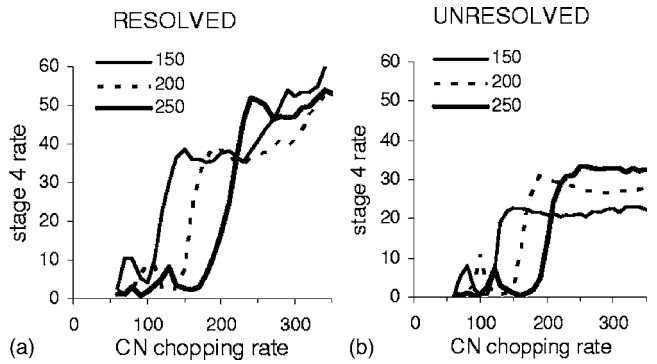

FIG. 2. Stage 4 rate profile for three 500-ms harmonic tones with F0=150, 200, and 250 Hz presented at 70 dB SPL. The *x* axis is the saturated rate of firing (spikes/s) of the VCN units at the base of the processing chain. The *y* axis is the rate of firing (spikes/s) of the stage 4 units. (A) Harmonics 3–8. (B) Harmonics 13–18.

quadruple parallel-processor computer. It was implemented using the Development System for Auditory Modeling (DSAM). The DSAM code is available from the authors at 'http://dsam.org.uk'

III. EVALUATION

The evaluation of the model follows a similar pattern and uses similar stimuli to an earlier evaluation of an autocorrelation model of pitch perception (Meddis and O'Mard, 1997a). The input consists of stimuli that are known from psychophysical studies to have a virtual pitch and whose pitch qualities vary with respect to the stimulus characteristics. The model's response is then compared qualitatively to published psychophysical observations.

A. Pitch of a harmonic tone

Figure 2(A) shows the stage 4 profile to three "missing fundamental" harmonic tones composed of harmonics 3–8 presented at 70 dB SPL for 500 ms. There is no spectral energy in the signals at their fundamental frequencies (F0 =150, 200, and 250 Hz). Figure 2(B) shows the profile for the same F0s where the tone is composed of harmonics 13–18 only. The effect of changing F0 is similar in that the profile shifts to the right as F0 increases. Despite differences in the overall shape, it is clear that the profiles in both figures discriminate easily among the three different pitches. The low harmonics in Fig. 2(A) are resolved by the human auditory periphery while the high harmonics in Fig. 2(B) are not (Plomp, 1964). Both are heard to have a clear pitch and listeners have little difficulty in detecting the changes in pitch despite the difference in peripheral resolvability.

The left to right upward trend in Fig. 2(A) is explained in terms of the intrinsic chopping rates of the different modules. High chopping rates in the VCN units give rise to more activity in those IC units that receive their input. As a consequence, the profile shows a higher stage 4 rate count for shorter intervals. It is significant, however, that the stimuli with unresolved harmonics [Fig. 2(B)] do not show this continued upward slope. The horizontal plateaux to the right of the functions suggests that the stage 4 rate is not reflecting the intrinsic chopping rate of the VCN units. On the contrary, for higher F0 the plateau is also higher. The stage 4 rate is reflecting the frequency of the envelope of the stimulus. In other words, some high-rate choppers are synchronizing to the envelope of the stimulus even though this has a longer period than the intrinsic chopping interval. The plateau is not seen for resolved harmonics because the signals in the individual channels do not have strongly modulated envelopes.

B. Alternating phase

Another difference between resolved and unresolved harmonics can be seen in the way phase affects the perceived pitch. An important characteristic of pitch perception is its insensitivity to the relative phase of the stimulus components among low, resolved harmonics. There is no change in perceived pitch when the phase relationship among the harmonics is altered. However, if the stimulus is composed of a group of high numbered (or "unresolved") harmonics, the pitch can be affected by the relative phases of the components. This is clearly demonstrated if the phases of successive harmonics are alternated (sin, cos, sin, cos, etc.) When this happens, the pitch of the stimulus is normally heard to rise by a whole octave. The pitch of unresolved harmonics in alternating phase reflects the period of the stimulus envelope which doubles when the phase is alternated.

Figure 3 shows the profiles for a harmonic complex when filtered to include either (a) low (resolved) harmonics (125–625 Hz) or (b) high (unresolved) harmonics (3900–5400 Hz). These stimuli are a subset of those used by Carlyon and Shackleton (1994) who found that the perceived pitch of HIGH harmonics reliably increased by an octave when the components were in alternating phase.

Figures 3(A) and 3(C) show the stage 4 profiles for SINE phase tones with F0=125 and 250 Hz. This should be compared with the profiles in Fig. 3(B) where F0 is 125 Hz and the components are in ALTERNATING (sine-cos) phase. The effect of changing the relative phase of the components is different for high and low numbered harmonics. The shape of the 125 Hz profile does not change much for low (resolved) harmonics when the phase is altered. However for high (resolved) harmonics, the shape changes substantially; the profile develops an extra upward slope at a chopping rate of 250 Hz and becomes similar to the stimulus with 250-Hz sine-phase tone. This result is qualitatively consistent with experimental observations of Shackleton and Carlyon.

It is unclear at this stage of the development of the model as to how this effect might be quantitatively assessed. However, a simple comparison of the profiles can be attempted using the mean Euclidean distance ($\sqrt{\Sigma d^2/n}$, where d is the difference between the values at corresponding points and n is the number of pairs of points). This statistic has the values 25 and 27 for *resolved* harmonics when the 125 Hz alternating phase stimulus is compared with 125 Hz and 250 Hz sine phase stimuli, respectively. This is a small difference indicating a preferred pitch of 125 Hz. However, corresponding mean Euclidean distances of 12 and 5 are obtained for *unresolved* harmonics indicating a clear preferred pitch match of 250 Hz for the unresolved 125 Hz alternating phase stimulus.

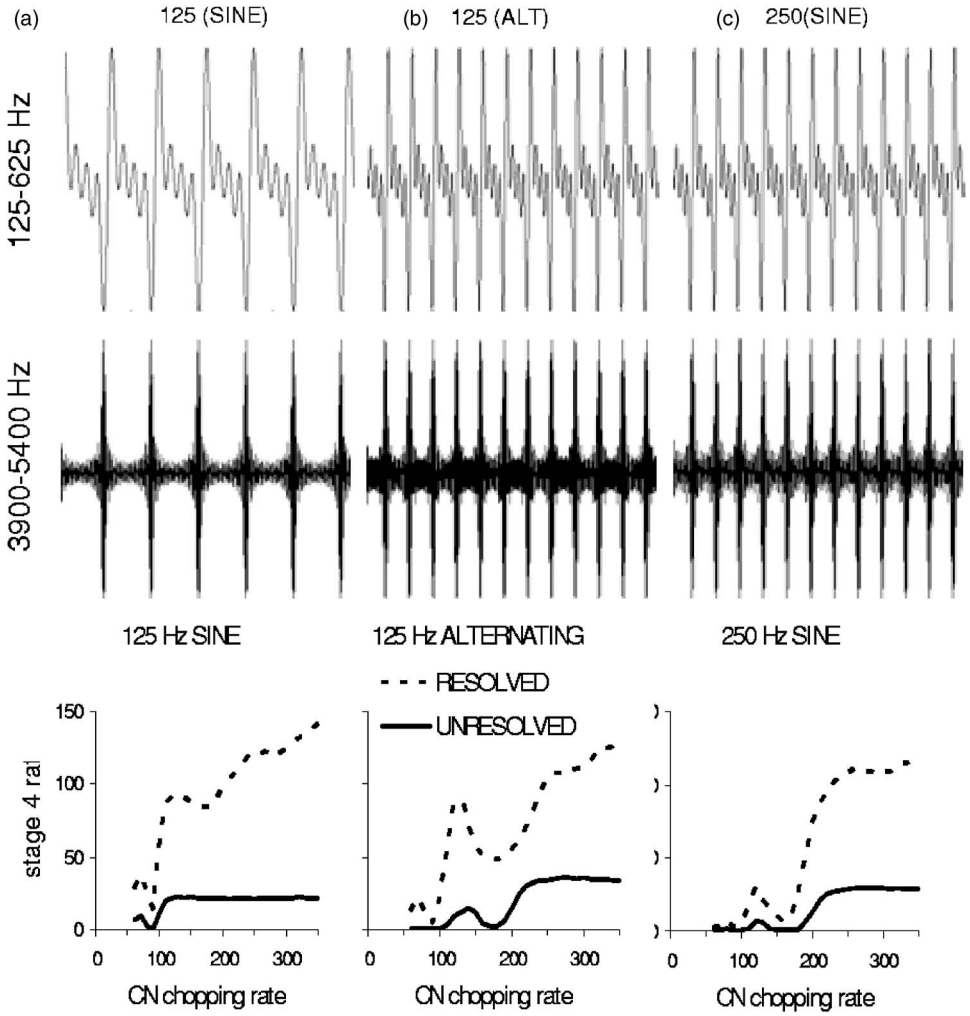

FIG. 3. Stage 4 rate profiles in response to stimuli consisting of low/resolved (125–625 Hz) and high/unresolved harmonics (3900–5400 Hz). (A) F0=125, sine phase. (B) F0=125 Hz alternating phase. (C) F0=250 Hz sine phase. Stimuli were 500 ms duration and presented at 70 dB SPL.

C. Inharmonic tones

De Boer (1956) showed that the pitch of a harmonic complex remained strong when the complex was made inharmonic by shifting the frequency of all components by an equal amount. The heard pitch of the complex shifts by a fraction of the shift of the individual components. This pitch shift is relatively large when the stimulus is composed of low harmonics but small for complexes composed of only high harmonics. For example, Moore and Moore (2003) showed that a complex composed of 3 resolved harmonics would show a pitch shift of 8% when the individual harmonics were all shifted by 24%. On the other hand, complexes of 3 unresolved harmonics showed much less pitch shift. The stimulus consisting of the least resolved (i.e., highest frequency) components showed little or no pitch shift. This was true for F0s of 50, 100, 200, and 400 Hz.

Two extreme examples of their stimuli were employed in the following demonstration which used 70-dB SPL, 500-ms tones consisting of either three resolved harmonics (3, 4, 5) or three unresolved harmonics (13, 14, 15) with F0=200 Hz. Pitch shifts were produced by shifting all component tones by either 0, 24%, or 48% and generating a stage 4 profile for each.

The resulting stage 4 profiles are shown in Fig. 4. The profiles for the *resolved* harmonics change with the shift. Shifting the frequencies of the *unresolved* harmonics [Fig. 4(B)] had little effect. Qualitatively at least, this replicates the results of Moore and Moore at least for their two extreme stimuli. When all harmonics are shifted by the same amount, the envelope of the signal is unchanged because it depends on the spacing between the harmonics which remains constant. The model reflects the unchanged periodicity of the envelope of the stimulus with unresolved harmonics in Fig. 4(B). On the other hand, the model reflects the changing

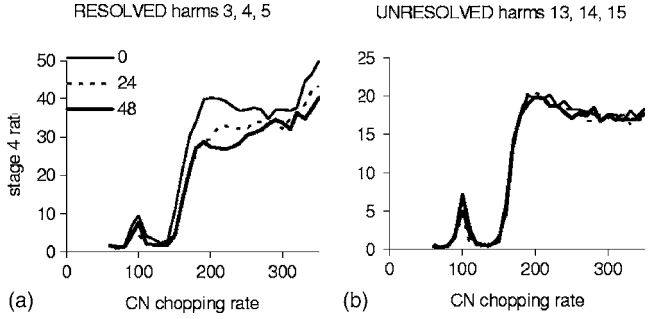

FIG. 4. Stage 4 rate profiles for shifted harmonic stimuli (F0=200 Hz). Shifts applied equally to all harmonics are 0, 24, or 48 Hz. (A) Harmonics 4, 5, and 6 (resolved). (B) Harmonics 13, 14, and 15 (unresolved). Shifting the harmonics has a smaller effect when the harmonics are unresolved.

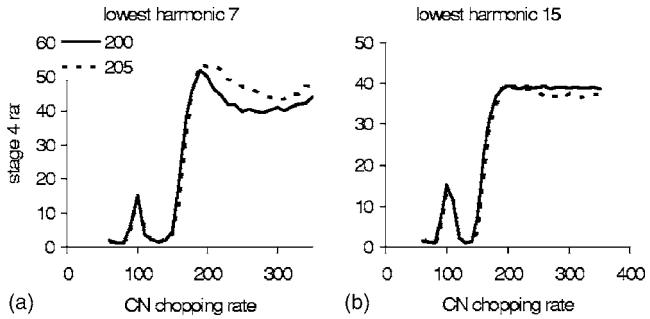

FIG. 5. Effects of pitch differences on the stage 4 rate profile for stimuli consisting of 11 successive harmonics. A shift of 5 Hz (F0=200, 205 Hz) is applied. Stimuli differ in terms of the number of the lowest harmonic in the series. (A) Lowest harmonic=7. (B) Lowest harmonics=15.

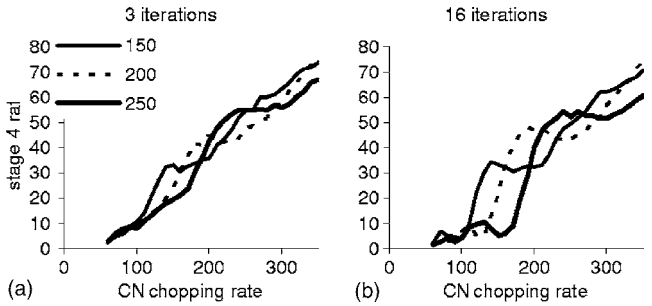

FIG. 6. Stage 4 rate profiles produced by the model in response to iterated rippled noise stimuli with iteration delays corresponding to pitches of 150, 200, and 250 Hz. (A) 3 iterations. (B) 16 iterations.

values of the resolved components when resolved stimuli are used [Fig. 4(A)].

This demonstration contrasts with the alternating phase demonstration in the previous section. The alternating phase effect occurs only when high harmonics are involved. On the other hand, when the pitch change is caused by inharmonic component frequency shifts, the effect is strongest when the shifts are applied to the low (resolved) harmonics.

D. Threshold for pitch difference

Harmonic number is also involved in determining the sensitivity to pitch differences. Houtsma and Smurzynski (1990) showed that the frequency difference threshold for the pitch of an 11 component harmonic complex was greater for high than for low harmonics. The model was therefore tested using tones consisting of either 11 successive low harmonics (lowest=7) or high harmonics (lowest=15), all in sine phase. F0 was 200 Hz and stimuli were presented for 500 ms. In both cases a second stimulus was presented with F0 increased to 205 Hz. A 5-Hz F0 difference was found by Houtsma and Smyrzinski to be close to the pitch difference threshold for stimuli when the lowest harmonic was 13 (see their Fig. 3).

The stage 4 rate profiles for each of these stimuli are given in Fig. 5. The 5-Hz shift in F0 has a greater effect for the stimulus composed of low harmonics [Fig. 5(A)] than for the stimulus consisting of high harmonics [Fig. 5(B)]. This is qualitatively consistent with Houtsma and Smyrzinski's result.

E. Iterated rippled noise

Iterated rippled noise (IRN) is of particular interest in the study of pitch because it produces a clear pitch percept but does not have the pronounced periodic envelope that is typical of harmonic and inharmonic tone complexes. When IRN is created by adding white noise to itself after a delay (d), the perceived pitch is typically matched to a pure tone or harmonic complex whose fundamental frequency is $1/d$. The strength of the pitch percept is also typically proportional to the number of times the delay-and-add process is repeated. The model was evaluated using stimuli constructed using a delays of 6.67, 5, and 4 ms reciprocals and a gain of 1. These stimuli have pitches around 150, 200, and 250 Hz. Figure 6 shows the rate profiles for these three values of IRN delay at two levels of iteration. Figure 6(A) shows only small differences in the profiles for the different delays when only three iterations are used. When 16 iterations are used [Fig. 6(B)] the differences between the profiles are more marked and this is consistent with what is heard.

F. Operation of the model

It is difficult to give a rapid intuitive insight into the operation of the model because of its complexity but the following may be helpful. While sustained chopping units fire regularly, their firing times are not necessarily in phase with one another even when the units have exactly the same chopping rate. However, the units will synchronize their firing to the phase of a pulsating stimulus if the pulse rate is the same as the intrinsic firing rate of the VCN unit. As a result all units with the same intrinsic firing rate will become synchronized with each other because they are synchronized to the same driving source.

An important consequence of this synchrony is that groups of VCN units will send synchronous volleys of spikes towards their target IC unit. The IC unit functions in the model as a coincidence detector. It produces the greatest response when the input from the VCN units is synchronized. This is most likely to occur when the period of the stimulus has some simple numerical relationship to the firing period of the VCN units. This gives the model its sensitivity to periodicity within BF-channels. Figure 7 illustrates the model response to a harmonic tone consisting of harmonics 3–8 of a 200-Hz F0. The IC response is shown for two different τGk values. The IC response across all channels is more vigorous and more consistent for $\tau Gk=0.52$ when the chopping rate ($\approx$200 Hz) is similar to the fundamental (Fig. 7, middle row). When $\tau Gk=0.85$ (chop rate $\approx$150 Hz) the pattern is less coherent across time (Fig. 7 bottom row). The best stimulus to recruit a chopper unit is a stimulus whose period is the same as that of the intrinsic firing of the chopper.

The model is sensitive to the period of a pure tone as well as the period of the envelope of a complex tone. This is important when the harmonics are resolved and the BM response in a single channel is a sinusoidal oscillation. It is also the case that the VCN units will synchronize to resolved low-frequency tones when the frequency of the tone is a whole multiple of the chopping rate of the VCN unit. For

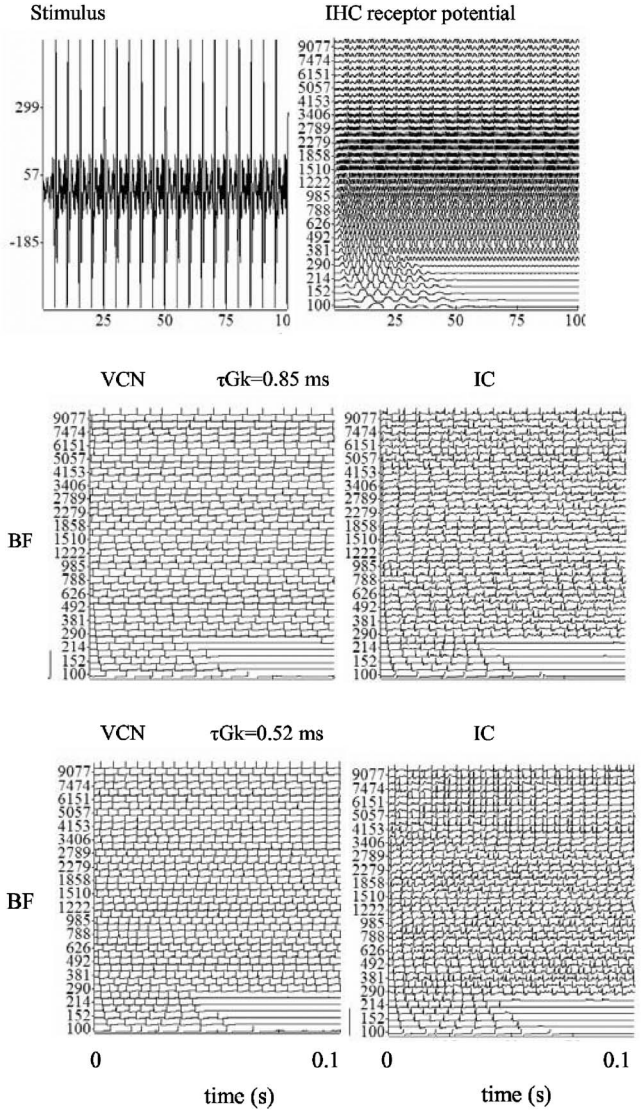

FIG. 7. Illustration of the model response to harmonic tone (F0=200 Hz, harmonics 3–8). The IC response is shown for two different τGk values. The IC response across all channels is more vigorous and more consistent when $\tau Gk=0.52$ (middle row, chop rate ≈200 Hz approx) compared to $\tau Gk=0.85$ (bottom row, chop rate ≈150 Hz).

example, a chopper with a spike rate of 150 spikes/s will synchronize with a 300-Hz pure tone or any multiple of 150 Hz although this effect becomes weaker as the tone frequency rises towards higher multiples of the unit's intrinsic rate. As a consequence, a complex tone consisting of the resolved harmonics of 150 Hz will stimulate choppers firing at a rate of 150 Hz in all of the different BM sites stimulated by the individual harmonic components.

IV. DISCUSSION

The aim of this study was to demonstrate that a model using physiological components could share some of the properties previously shown to characterize cross-channel autocorrelation models. The mathematical model and the new physiological model have a great deal in common. Both use an auditory first stage to represent auditory nerve spike activity. Both extract periodicity information on a within-channel basis. Both models accumulate information across channels to produce a periodicity profile. They differ only in the mechanism used to determine periodicities. In the physiological model, periodicity detection uses VCN units working together with their target IC units. This replaces the delay-and-multiply operations of the autocorrelation method.

The physiological mechanism depends on the synchronization properties of the model VCN chopper units. When the choppers are driven by a stimulus periodicity that coincides with their intrinsic driven firing rate, all the VCN units with the same firing rate will begin to fire in synchrony with the stimulus and with each other. This periodicity may originate either from low frequency pure tones or from modulations of a carrier frequency. It is this synchrony that drives the receiving IC units. This is not, of course, exactly the same process as autocorrelation. But it is similar in the sense that they both systematically extract periodicity information on a temporal basis.

At first sight, the rate profile does not look like an autocorrelation function. It progresses from long to short intervals and has a steep slope across VCN chopping rate. On the other hand, the summary autocorrelation function progresses from short to long lags and is normally presented as an approximately horizontal function across lags. However, Pressnitzer *et al.* (2001) have recommended that autocorrelation functions can give a better representation of some psychophysical results (e.g., lower limit of pitch) when they are weighted so that the contribution of longer time intervals is progressively reduced by a linear weighting function (see their Fig. 7). If their suggestion were applied to summary autocorrelation functions and the function reversed to go from long to short intervals, the parallel would become more visually apparent.

A simplifying assumption in the present model is that all IC units receive input from VCN units with the same intrinsic firing rate. This has recently been questioned by another modeling study showing that more realistic IC MTFs can be obtained using a spread of inputs across different chopping rates and a second converging layer within the IC (Guerin *et al.*, 2006). This surprising result was published too late to influence the study reported here. However, it does suggest that it might be possible to proceed successfully with fewer restrictive assumptions than those implemented above.

It would be premature to claim that the model described above is a complete pitch model. There are more pitch phenomena than those considered here and the new model needs to be tested on a much wider wide range of stimuli. Indeed some stimuli produce a pitch that is claimed not to be predicted by existing autocorrelation models (e.g. Carlyon *et al.*, 2002; Plack and White, 2000). These are matters for further investigation. Nevertheless, this study has demonstrated that a cross-channel autocorrelation model can be simulated, to a first approximation, by a physiological model and is worthy of further consideration.

This report has restricted itself to a qualitative assessment of the model to show that the stage 4 rate profiles demonstrate changes that parallel psychophysical changes. A quantitative study will clearly be necessary. However, this will require the solution of problems that lie beyond the scope of this study. One important problem is that the stage 4

TABLE I. MacGregor point neuron parameters for VCN units.

Dendritic low-pass filer cutoff (Hz)	400
tau$M$, membrane time constant (s)	5×10^{-4}
tauGk, potassium time constant (s)	3.7×10^{-3} to 0.25×10^{-3}
b, increment in Gk (Siemens/s)	0.1
Ek potassium reversal potential (V)	−10
Th0 threshold (mV)	2

TABLE II. MacGregor point neuron parameters for IC units.

Dendritic low-pass filer cutoff (Hz)	4000
tau$M$, membrane time constant (s)	1×10^{-3}
tauGk, potassium time constant (s)	1×10^{-4}
b, increment in Gk (Siemens/s)	0.1
Ek potassium reversal potential (V)	−10
Th0 threshold (mV)	25

profile reflects stimulus level as well as stimulus pitch. As the signal level is increased, the spread of activity across the channels is also increased. More IC units are recruited and the profile shows a higher overall activity count. Pitch perception is largely (but not completely) independent of level and two pitches can be matched even when levels are slightly different. These two variables must be disentangled before a fully quantitative comparison can be attempted.

It is likely that other models of IC functioning might do equally as well (Nelson and Carney, 2004; Langner, 1981) although this has not been tested. In particular, the Nelson and Carney model has been shown to simulate the response of IC units to SAM tones. This property is an important pre-requisite. The Hewitt and Meddis (1992, 1994) models were used only because they were more familiar to the authors. It is possible that any model of IC functioning would work so long as it reproduces rate MTFs as measured physiologically (e.g., Rees and Palmer, 1988; Krishna and Semple, 2000). However, these alternative IC models would need to be adapted to integrate periodicity information across BF channels in an additional layer of neuronal processing.

de Cheveigne (1998) has also developed a physiological model of periodicity extraction at the level of the cochlear nucleus using delay-and-inhibit circuits. This mechanism has been shown to be formally similar to autocorrelation. It is possible that delay-and-inhibit "building blocks" would serve the present purpose equally well although this again has not been tested. However, it remains to be seen whether the delay-and-inhibit idea is physiologically plausible given that it relies on an extensive range of graded inhibitory delays.

There is abundant evidence for VCN units that chop at the rates featured in this study (60–350 spikes/s). However, the physiological plausibility of the new model might itself be questioned in the light of the observations of Krishna and Semple (2000). Their results indicated that the rate MTF functions observed in the IC are more complex than was remarked on in earlier studies. They found not only rate-enhancement but also rate-suppression. Some IC units showed suppression of firing at some modulation frequencies and enhancement at others. Also rate MTFs [Krishna and Semple, 2000, Figs. 3(A), 3(C), and 3(D)] show enhancement peaks at 500, 400, and 100 Hz, respectively, but substantial suppression at lower modulation frequencies. Our IC units show none of these complexities and a more sophisticated model than that presented here should take this into account.

ACKNOWLEDGMENTS

This work was supported by a research grant from the Biotechnology and Biological Sciences Research Council. We thank Susan Shore, Ian Winter, Brian Moore, and an anonymous reviewer who provided helpful comments on the manuscript.

APPENDIX: SUSTAINED CHOPPER MODEL

The VCN unit model is based on MacGregor (1987) point neuron model. The present implementation consists of two stages: input at the dendrites and spike generation at the soma, respectively. The original MacGregor neuron provides for the possibility that the spiking threshold can vary over time but this has been omitted here. This is achieved by setting his parameter c to 0.

Each auditory nerve spike was represented as a 6 mV postsynaptic pulse lasting 50 μs. The dendritic input stage applies a first-order low pass filter (3 dB cutoff at 400 Hz, 0 dB gain) to the AN peri stimulus time histogram (PSTH) to produce a voltage change $I(t)$ at the soma.

The *trans*-membrane potential at the soma, E, is represented as a deviation from resting potential, Er, and tracked using the equation,

$$dE/dt = -E/\tau_m + I(t) + Gk(t)(Ek - E), \tag{A1}$$

where τ_m is the membrane time constant, Ek is the potassium reversal potential (relative to Er) and $Gk(t)$ is the cell potassium conductance. Conductance changes are tracked using

$$dGk = -Gk^* dt/\tau_{Gk} + b^* s, \tag{A2}$$

where τ_{Gk} is the potassium time constant, b is the increase in Gk following an action potential (AP) indicated when $s=1$. An AP is initiated when the membrane potential exceeds a threshold $E(t) > \mathrm{Th0}$. This threshold was fixed throughout. See Table I for parameter values.

The IC neuron used the same equations but some parameters were changed to produce tighter coincidence criteria (see Table II).

Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–796.

Blackburn, C. C., and Sachs, M. B. (**1989**). "Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis," J. Neurophysiol. **62**, 1303–1329.

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

Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?" J. Acoust. Soc. Am. **95**, 3541–3554.

De Boer, E. (**1956**). "On the residue in hearing," Ph.D. thesis, University of Amsterdam.

Guérin, A., Le Bouquin Jeannès, R., Bès, J., Faucon, G., and Lorenzi, C. (**2006**). "Evaluation of two computational models of amplitude modulation coding in the inferior colliculus," Hear. Res. **211**, 54–62.

de Cheveigné, A. (**1998**). "Cancellation model of pitch perception," J. Acoust. Soc. Am. **103**, 1261–1271.
Hewitt, M. J., and Meddis, R. (**1992**). "A computer model of a cochlear-nucleus stellate cell: Responses to amplitude-modulated and pure-tone stimuli," J. Acoust. Soc. Am. **91**, 2096–2109.
Hewitt, M. J., and Meddis, R. (**1994**). "A Computer Model of Amplitude-Modulation Sensitivity of Single Units in the Inferior Colliculus," J. Acoust. Soc. Am. **95**, 2145–2159.
Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304–310.
Krishna, B. S., and Semple, M. N. (**2000**). "Auditory temporal processing: Responses to sinusoidally amplitude-modulated tones in the inferior colliculus," J. Neurophysiol. **84**, 255–273.
Langner, G. (**1981**). "Neuronal mechanisms for pitch analysis in the time domain," Exp. Brain Res. **44**, 450–454.
Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128–133.
Lopez-Poveda, E. A., and Meddis, R. (**2001**). "A human nonlinear cochlear filterbank," J. Acoust. Soc. Am. **110**, 3107–3118.
Lyon, R. F. (**1984**). "Computational models of neural auditory processing," IEEE ICASSP **84**, 3.
MacGregor, R. J. (**1987**). *Neural and Brain Modeling* (Academic, San Diego).
Meddis, R., and O'Mard, L. P. (**2005**). "A computer model of the auditory nerve response to forward masking stimuli," J. Acoust. Soc. Am. **117**, 3787–3798.
Meddis, R., and O'Mard, L. P. (**1997a**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.
Meddis, R., and O'Mard, L. P. (**1997b**). "Modeled chop-S and pauser/build-up responses in the cochlear nucleus: Further studies," ARO midwinter meeting, abstract 454.
Meddis, R. (**2006**). "Auditory-nerve first-spike latency and auditory absolute threshold: A computer model," J. Acoust. Soc. Am. **119**, 406–417.
Meddis, R., and Hewitt, M. J. (**1991**). "Virtual pitch and phase-sensitivity studied using a computer model of the auditory periphery: I pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.
Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**). "A computational algorithm for computing nonlinear auditory frequency selectivity," J. Acoust. Soc. Am. **109**, 2852–2861.
Meddis, R., and O'Mard, (**1997**). "A unitary theory of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.
Moore, G. A., and Moore, B. C. J. (**2003**). "Perception of the low pitch of frequency-shifted complexes," J. Acoust. Soc. Am. **113**, 977–985.
Nelson, P. C., and Carney, L. H. (**2004**). "A phenomenological model of peripheral and central neural responses to amplitude-modulated tones," J. Acoust. Soc. Am. **116**, 2173–2186.
Plack, C. J., and White, L. (**2000**). "Pitch matches between unresolved complex tones differing by a single interpulse interval," J. Acoust. Soc. Am. **108**, 696–705.
Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.
Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.
Rees, A., and Palmer, A. R. (**1988**). "Neuronal responses to amplitude-modulated and pure tone stimuli in the guinea-pig inferior colliculus, and their modification by broadband noise," J. Acoust. Soc. Am. **85**, 1978–1994.
Wiegrebe, L., and Meddis, R. (**2004**). "The representation of periodic sounds in simulated sustained chopper units of the ventral cochlear nucleus," J. Acoust. Soc. Am. **115**, 1207–1218.
Yost, W. A., Hill, R., and Perez-Falcon, T. (**1978**). "Pitch and pitch discrimination of broadband signals with rippled power spectra," J. Acoust. Soc. Am. **63**, 1166–1173.