Biol Cybern (2007) 97:479–491
DOI 10.1007/s00422-007-0187-8

ORIGINAL PAPER

Biological Cybernetics

A model of perceptual segregation based on clustering the time series of the simulated auditory nerve firing probability

Emili Balaguer-Ballester · Martin Coath · Susan L. Denham

Received: 27 June 2007 / Accepted: 27 September 2007 / Published online: 10 November 2007

Abstract This paper introduces a model that accounts quantitatively for a phenomenon of perceptual segregation, the simultaneous perception of more than one pitch in a single complex sound. The method is based on a characterization of the time-varying spike probability generated by a model of cochlear responses to sounds. It demonstrates how the autocorrelation theories of pitch perception contain the necessary elements to define a specific measure in the phase space of the simulated auditory nerve probability of firing time series. This measure was motivated in the first instance by the correlation dimension of the attractor; however, it has been modified in several ways in order to increase the neurobiological plausibility. This quantity characterizes each of the cochlear frequency channels and gives rise to a channel clustering criterion. The model computes the clusters and the pitch estimates simultaneously using the same processing mechanisms of delay lines; therefore, it respects the biological constraints in a similar way to temporal theories of pitch. The model successfully explains a wide range of perceptual experiments.

1 Introduction

A harmonic complex sound is commonly heard as a single perceptual entity, with a pitch corresponding to the fundamental frequency (called the *fundamental*, *global* or *residual pitch*). The autocorrelation theories of pitch perception (Licklider 1951, 1959; Lyon 1984; Meddis and Hewitt 1991a, b; Meddis and O'Mard 1997; Yost et al. 1996; Bernstein and Oxenham 2005; Denham 2005) have proved successful in explaining the perceived pitch of harmonic and inharmonic sounds. On the other hand, a mixture of two complex tones with different fundamental pitches is typically heard as two entities. In particular, some manipulations of a harmonic complex can result in the perceptual segregation of one of the spectral components of the sound, which elicits a separate pitch that is different from the fundamental pitch. This phenomenon is known as perceptual segregation (Roberts 2005).

Several perceptual segregation studies, including (Hartman 1996; Roberts and Bailey 1996; Brunstrom and Roberts 1998, 2000; Li and Hartman 1998; Roberts and Brunstrom 2001), suggest that different mechanisms govern the computation of pitch and the perceptual fusion of the complex harmonics. These two mechanisms are the cross-channel comparison between periodicities, which govern perceptual segregation; and the aggregation of periodicities, which govern the computation of the fundamental pitch (Roberts 2005). Existing theories of pitch perception have recently addressed the qualitative aspects of this phenomenon (see review in de Cheveigné 2005; Roberts and Brunstrom 2001). Earlier, Meddis and Hewitt (1992) found that peripheral channel selection was a useful method for the qualitative identification of two concurrent vowels using autocorrelation theories of pitch. More recently, Roberts and Holmes (2006) provided a metric of the degree of fusion of the sound components into a single fundamental.

The novelty introduced in this paper is a plausible method for clustering cochlear frequency channels. The method is sensitive to the nonlinear dynamical behaviour of the responses in the frequency channels and accounts quantitatively for the perceptual segregation phenomenon. Most importantly, the model respects the main biological constraints of the autocorrelation theories of pitch. It shows

E. Balaguer-Ballester (✉) · M. Coath · S. L. Denham
Centre for Theoretical and Computational Neuroscience, University of Plymouth, Portland Square, Drake Circus, Plymouth, Devon PL4 8AA, UK
e-mail: emili.balaguer-ballester@plymouth.ac.uk

how these pitch theories contain the elements to generate a measure of the simulated auditory nerve (AN) firing probability time series inspired by the second generalized dimension of its attractor (correlation dimension, D_2). This parameter characterizes each frequency channel.

An interesting aspect is that the model simultaneously characterizes the dynamics of the cochlear frequency channels and estimates their autocorrelations. Remarkably, it uses neuronal delay lines (Licklider 1951, 1959) and low-pass filters just as the autocorrelation pitch theories. Therefore, the plausibility of the first stage of the model is similar to that of the autocorrelation models of pitch perception (Meddis and O'Mard 2006).

The resultant measure, $\tilde{D}$, preserves some similarities with the correlation dimension. However, it will be shown that the plausibility constraints imply that $\tilde{D}$ cannot be directly compared with the correlation dimension; because it has to be defined over a small subset of the AN firing probability phase space vectors.

In a second stage, the $\tilde{D}$ measure induces a clustering of the channels. The pitch extraction runs separately within each cluster, using cross-channel expected values of the *running* autocorrelations (i.e., an instantaneous estimate of the autocorrelations of the cochlear frequency channels). This simple model explains the most salient pitches reported by listeners in a wide range of perceptual segregation experiments.

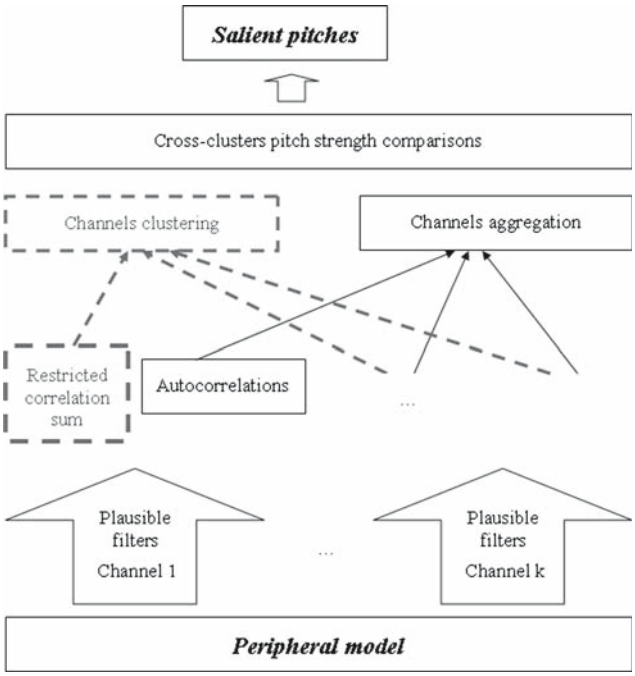


Fig. 1 Diagram of the model. The first stage consists of biologically plausible filters, which simulate the auditory nerve firing probability in each peripheral channel (see text). The filtered signals are used simultaneously to generate clusters of channels and to compute the pitch model responses within them. Finally, an across-clusters comparison of the pitch model responses predicts the saliency of the different pitches that can be heard out

2 Model description

The method provides temporal theories of pitch with a plausible procedure for channel selection synchronized with the pitch computation. The model consists of a characterization of peripheral channel firing probability followed by cross-channel comparison. The next sections describe these two stages.

2.1 Characterization of the dynamics of cochlear frequency channels

Figure 1 shows the model diagram. In the first stage, a nonlinear cochlear model (Lopez–Poveda and Meddis 2001) followed by an inner hair cell model (Sumner et al. 2003) is used to generate the AN firing probability in response to an acoustic stimulus. This peripheral model produces a separate representation for each frequency channel, which is characterized by a best frequency (BF) and a restricted frequency range of response. In the present implementation, 30 channels were used with centre frequencies logarithmically spaced between 100 and 8,000 Hz approximately.

The next stage consists of an integration of the AN signals. The aim of this stage is the simultaneous computation of the running autocorrelations (Meddis and Hewitt 1991a) and a parameter that characterizes the dynamics of each channel. Autocorrelation models typically use the spike probabilities for channel number k, $p(t,k)$. In Meddis and O'Mard (2006), these spike probabilities drive the activity of a layer of chopper cells in the cochlear nucleus, which generate the inputs to coincidence-detector neurons in the inferior colliculus.

The presented model assumes, for simplicity, that the auditory nerve firing probability provides direct input to the bank of coincidence-detector units, having a best frequency of $1/l$, as in Meddis and O'Mard (1997). Then a leaky integration computes the running autocorrelation $h(t,k)$ at each time step:

$$h(t,l,k) = p(t,k) \cdot p(t-l,k) + h(t-\Delta t,l,k) \cdot \mathrm{e}^{-\Delta t/\tau}, \quad (1)$$

where Δt is the sampling interval of the stimulus (corresponding to the sampling rate of 22,050 Hz for the stimuli tested), τ is a time constant and l is the autocorrelation lag. Equation 1 is similar (apart from constant factors) to a membrane low-pass filter equation representing a population of neurons receiving a synaptic input $p(t,k)p(t-l,k)$. However, this equation applies to firing probabilities and not to a membrane potential, therefore, the time constant is not necessarily

equal to the membrane constant of the coincidence-detector cells (Dayan and Abbot 2001). We used a time constant of $\tau = 400$ ms throughout the study. This value is motivated by perceptual studies (Halls and Peters 1981; Plack and White 2000; Wiegrebe 2001; Grose et al. 2002) and EEG studies (Krumbholz et al. 2003) which provide evidence for pitch integration times. However, the results presented in Sect. 3 are robust under other choices for the sampling rate and the time constant. The neural response $h(t, k, l)$ at different lags l contains the instantaneous running autocorrelations of the auditory nerve simulated firing probability in each cochlear frequency channel. However, in the context of pitch perception theories, this quantity is typically dimensionless. This will be the convention used throughout the paper.

The characterization of the time series of the AN firing probability for channel number k requires embedding it into a suitable phase space (right plot in Fig. 2a). A two-dimensional phase space consists of the set of vectors

$$\vec{x}(t, k) = \begin{bmatrix} p(t, k) \\ p(t - L(k), k) \end{bmatrix}; \tag{2}$$

thus, the x-axis in Fig. 2a (right plot) represents an L-delayed version of the frequency channel time series; where $L = L(k)$ is the embedding delay for the frequency channel number k. It corresponds approximately to the first zero of the autocorrelation, because it is convenient for an adequate embedding to minimize the correlation between axes (Kantz and Schreiber 1999). Therefore, the values of $L(k)$ depend on the frequency channel and on the stimulus. In this model, the embedding delay is the lag corresponding to the first positive minimum of Eq. 1 in each frequency channel. This value stabilizes after 80 ms in the stimuli tested and remains approximately constant throughout the stimulus duration. This simplification might not be adequate for other stimuli in which the pitch varies, such as the iterated ripple noises (Yost 1996; Wiegrebe 2001; Denham 2005). The $L(k)$ values found vary from 25 ms in the lowest channels to 1 ms in the higher channels.

The portrait of a time series in this space is a representation of the time series attractor (Takens 1981). For example, the representation stimulus waveform of a pure tone sinusoid is trivially a circle in phase space (Fig. 2a, left plot). However, the nonlinearities and stochasticity present in the peripheral model modify the waveform time series; therefore, the phase space plot of the AN simulated time series in each channel will be different, in general, to the stimulus phase space portrait (see Fig. 2b).

In summary, a possible strategy to characterize the dynamics of these time series consists of analysing the distribution of vectors in the phase space. A parameter set that summarizes this distribution of vectors are the generalized dimensions of the attractor (Kantz and Schreiber 1999). The second of these dimensions is the correlation dimension (D_2), which is relatively easy to estimate in a stationary section of the time series (Kantz and Schreiber 1999).

In order to calculate D_2, it is necessary first to compute the ratio of points in the phase space that are closer than ε to each other, this is called the correlation sum (Kantz and Schreiber 1999):

$$C(\varepsilon, t, k) = \frac{1}{\alpha} \cdot \sum_{i=i_{\min}}^{N} \sum_{j=j_{\min}}^{i+1} \Theta\left(\varepsilon - \|\vec{x}(i \cdot \Delta t, k) - \vec{x}(i \cdot \Delta t - j \cdot \Delta t, k)\|\right), \tag{3}$$

where $\Delta t I_{\min} = l(k)$, N is the number of samples, ε is the radius of the spheres into which the phase space is divided, Θ is the Heaviside function ($\Theta(\varepsilon - x) = 1$ if $\varepsilon > x$ and else is zero) and α is the total number of vector pairs. For example, Fig. 2a (right plot) shows that the distance between the two vectors exceeds ε, therefore this pair contributes zero to the summations in Eq. 3. The minimum number of samples separation ($j_{\min} < i_{\min}$) is chosen to avoid an excessive autocorrelation among the different vectors, and it is necessary to obtain the correct value of the correlations. An initial time of $j_{\min}\Delta t = 0.3$ ms is adequate according to the space-time separation plots method (Provenzale et al. 1992) for the stimuli tested.

The correlation dimension definition is

$$D_2(t, k) = \lim_{\varepsilon \to 0} {}^{\partial \ln(C(t, \varepsilon, k))}\!/_{\partial \ln \varepsilon}. \tag{4}$$

Therefore, the slope of $C(t, \varepsilon, k)$ with respect to the radius of the spheres on a double logarithmic axe provides an empirical estimate of the correlation dimension (Kantz and Schreiber 1999). For example, the value of D_2 of a pure tone waveform (Fig. 2a, left plot) is unity; and the D_2 of the simulated AN firing probability in one of the frequency channels (Fig. 2b) is approximately 0.8. In general, it has been found useful to characterize an approximately stationary time series by using standard algorithms for correlation dimension estimates (Kantz and Schreiber 1999; Balaguer et al. 2006). Our preliminary studies indicated that this method could be useful to analyze the simulated AN firing probability time series (Balaguer and Denham 2006).

However, the correlation dimension (Eq. 4) cannot be considered a biologically plausible calculation. In addition, it can only be defined if the time series is approximately "wide-sense stationary" (Kantz and Schreiber 1999), like the section plotted in Fig. 2b. Therefore, it was necessary to modify the correlation dimension algorithm in order to provide a biologically plausible model for the characterization of the peripheral channels. The main restriction that we impose on this model is that it has to be synchronized with the autocorrelation model of pitch extraction (Meddis and O'Mard 1997).

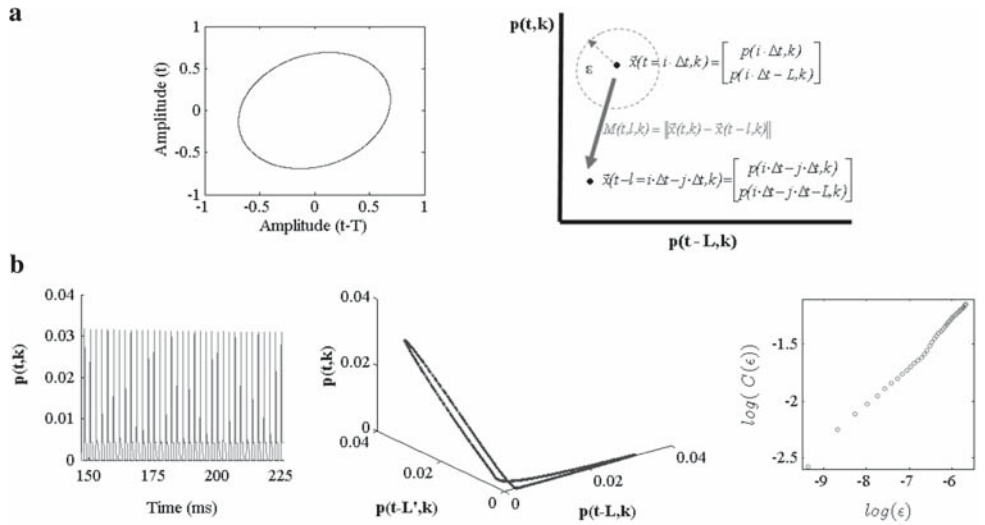


Fig. 2 **a** The left plot shows a two-dimensional embedding of a 440 Hz pure tone waveform (in units of normalized amplitude, sampled at 22 kHz, embedding delay $T = 0.6$ ms). The correlation dimension is one. The *right plot* shows an example of a two dimensional embedding of the firing probability time series $p(t, k)$ (frequency channel number k, L is the embedding delay). **b** The *left plot* shows the auditory nerve simulated firing probability in a single cochlear frequency channel (centre frequency 1183 Hz), using a nonlinear peripheral model response to the same 440 Hz pure tone (Lopez–Poveda and Meddis 2001, Sumner 2003), selected from a stationary part of the time series. The *centre plot* shows the corresponding 3D phase space embedding (delays $L = 4.7$ ms; $L' = 13.7$ ms). The *right plot* shows an approximate linear relationship; the slope of the linear section is $D_2 \approx 0.6 \pm 0.1$; the Takens estimator of $D_2 \approx 0.8$ (Theiler 1988)

The model starts by computing the distance between two points in phase (Fig. 2a, left plot),

$$M(t, l, k) = \|\vec{x}(t, k) - \vec{x}(t - l, k)\|; \tag{5}$$

it is easy to see (by using Eq. 2) that $M(t, l, k)$ consists of sums and differences of cross products of the AN firing probability (some of them used in Eq. 1). Equation 6 counts the number of distances smaller than some threshold ε. It is remarkable that this calculation can be done in parallel to the autocorrelation in Eq. 1.

$$\tilde{C}(t, \varepsilon, l, k) = \Theta(\varepsilon - M(t, k, l)) + \tilde{C}(t - \Delta t, \varepsilon, l, k); \tag{6}$$

Then, a sum across lags is

$$\tilde{C}(t, \varepsilon, k) = \sum_{l=n_{\min}\cdot\Delta t}^{\max(l)} \tilde{C}(t, \varepsilon, l, k). \tag{7}$$

This quantity is computed in parallel for a range of ε values (fixed throughout the stimulus), representing the precision with which the phase space is tiled.

A direct comparison with Eq. 3 shows that $\tilde{C}(t, \varepsilon, k)$ also counts the number of neighbours of some of the phase space vectors that fall into a sphere of radius ε. Equation 6 is equivalent to the i-sum in Eq. 3. Simultaneously, for a fixed t, Eq. 7 computes the j-sum in Eq. 3, because the j-index corresponds to the different lags within a single frequency channel k.

At this stage of the algorithm the question of the numerical difference between the correlation sum (Eq. 3) and the approximation $\tilde{C}(t, \varepsilon, k)$ is raised.

If the lag spacing were uniform and equal to the sampling period, the entire j-sum showed in Eq. 3 would be computed, and the approximate correlation sum $\tilde{C}(t, \varepsilon, k)$ and the theoretical value $C(t, \varepsilon, k)$ would correspond (except for a global factor $1/\alpha$).

However, there are three differences between these two quantities. Firstly, the pitch perception theories do not use such a lag resolution; the average lag resolution (0.5 ms) is an order of magnitude smaller than the sampling period ($\approx$0.05 ms). Secondly, the lag spacing is non-uniformly distributed (see Appendix). Thirdly, the maximum lag employed in the pitch models, $\max(l)$ is limited typically by the lowest frequency perceived (around 40 Hz). Therefore, the restricted set of phase space vectors used in the computation of $\tilde{C}(t, \varepsilon, k)$ are a consequence of a nonlinear sampling of the original phase space and contain typically 10% of the total number of phase space vectors. Consequently, it is expected that the values of the approximation $\tilde{C}(t, \varepsilon, k)$, will be different to the correlation sum values. Nevertheless, they are useful for detecting the changes in the dynamics between different frequency channels.

The slope of $\ln(\tilde{C}(t, \varepsilon, k))$ versus $\ln(\varepsilon)$ in a strongly linear region defines a quantity $\tilde{D}(t, k)$,

$$\ln(\tilde{C}(t, \varepsilon, k)) = \tilde{D}(t, k) \cdot \ln \varepsilon + \beta, \tag{8}$$

which is no longer a dimension of the attractor. The Appendix provides further details of the criterion for estimating $\tilde{D}$, specifically designed for this application.

These mathematical differences with the correlation dimension do not challenge the validity of this model. Moreover, the results shown in Sect. 3 demonstrate that the more neurobiologically plausible computation, $\tilde{D}(t,k)$ will be adequate to explain the perceptual segregation phenomena.

2.2 Cochlear frequency channels selection

Large changes in $\tilde{D}$ values across the channels give rise to an empirical criterion for establishing boundaries and the grouping of adjacent channels into clusters. Channels are grouped together into a cluster if the relative discrepancy in their $\tilde{D}(t,k)$ values is less than some threshold. A new cluster begins wherever a channel has $\tilde{D}(t,k)$ values that differ by more than this threshold from the $\tilde{D}(t,k)$ values of any of the channels in the previous cluster (see Fig. 3b). A maximum within-cluster discrepancy of 50% gives satisfactory results

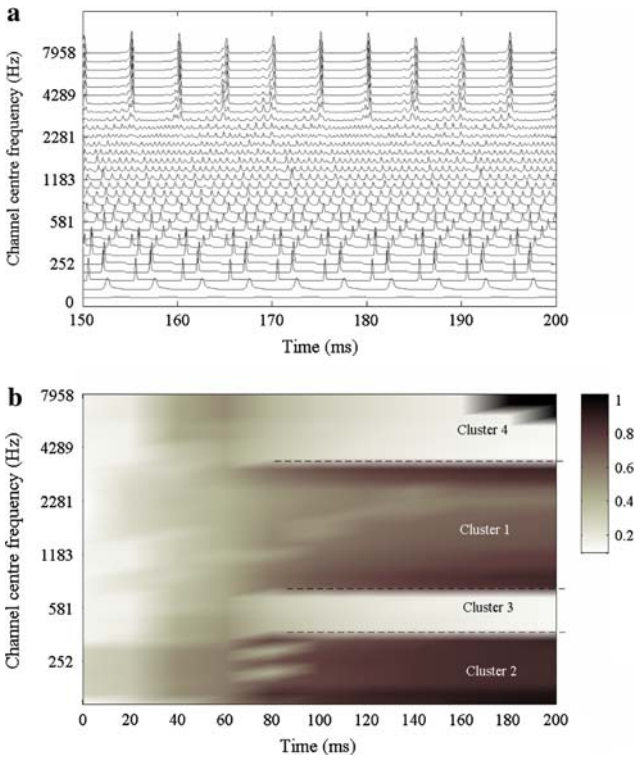


Fig. 3 **a** Probability of firing in different frequency channels during a short stationary section taken from the peripheral model response to stimulus 1. **b** $\tilde{D}$ values of the auditory nerve simulated firing probability of stimulus 1 (using the biologically plausible algorithm described in the text). The *dotted horizontal lines* show the cluster boundaries generated. The convention used in this model is that the clusters should contain more than three channels; therefore, the higher channels in the figure do not form a separate cluster. *Black* $\tilde{D}=1$; *White* $\tilde{D}=0$

(Sect. 3). A further restriction is that clusters should consist of more than three peripheral channels. This restriction is reasonable; because such a small cluster of channels gives a running autocorrelation response (Eq. 1) typically biased towards the centre frequency of their filters (and therefore less related to the stimulus periodicities). The drawback of this restriction is that having only 30 peripheral channels, it might be impossible, in principle, to separate some pitches having closely related harmonics. An increase in the number of channels of the peripheral model when needed would presumably avoid this problem. However, this increase in channel resolution is not necessary for the stimuli tested in this paper.

The next stage consists of computing an expected value of Eq. 1 across frequency channels and normalizing,

$$S_{\text{cluster}}(t,l)=\frac{\sum_{k=1}^{N}P(k|\text{cluster};t)\cdot h(t,l,k)}{\max_l(\sum_{k=1}^{N}P(k|\text{cluster};t)\cdot h(t,l,k))}, \quad (9)$$

where $P(k|\text{cluster};t)$ is the conditional probability at time t that the peripheral channel k belongs to a given cluster. Using the criterion described above, one finds that a channel can only belong to a single cluster, therefore for a cluster containing $N_$cluster channels, $\sum_k P(k|\text{cluster};t)=1$; where $P(k|\text{cluster};t)=1/N_$cluster for the channels grouped together into a single cluster and zero elsewhere. Equation 9 is a normalized summarized autocorrelogram (*SACF* in Meddis and O'Mard 1997), where the normalization occurs within each cluster individually. However, as a novelty, it includes the channel selection probabilities before the sum.

Although the binary probabilities provide successful results, $P(k|\text{cluster};t)$ could, in principle, take small non-zero values for all of the channels. This formulation accounts for the observation that even distant channels may have some small contribution to perceptual segregation (Roberts and Holmes 2006).

The final step of the model is a nonlinear transformation of Eq. 9, perceptually associated with the strength of the pitch (Yost 1996; Roberts and Holmes 2006),

$$PS(\text{Cluster})=\frac{10^{2\cdot S_{\text{cluster}}(t,l)}}{100}. \quad (10)$$

The fundamental pitch is typically the dominant percept, and its saliency corresponds to the maximum of the pitch strength in all of the clusters. The existing models (Meddis and Hewitt 1991a, b; Meddis and O'Mard 1997) already account for this pitch.

The novelty of the method presented here is the prediction of more than a single perceptual entity. In this paper, the maximum difference in pitch strength profiles between clusters defines the perceptual saliency of other pitches apart from the fundamental.

In the Appendix we show that this approach is consistent with the results of previous models using a pure tone stimulus.

3 Evaluation of the model

The first evaluation uses the harmonic complex stimuli introduced by Brunstrom and Roberts (1998) and Roberts and Brunstrom (2001); these stimuli elicit the sensation of a fundamental pitch as well as a salient segregated component. Stimulus 1 consists of partials 1–12 of 200 Hz in which the 6th harmonic (1,200 Hz) is mistuned downwards by 4%. Stimulus 2 comprises 14 harmonics of a 200 Hz fundamental; then harmonics 6, 7, 8 are removed and a 1,300 Hz probe tone is inserted. The model has also been benchmarked using other stimuli that elicit typically weaker perceptual segregation, like the ones used by Roberts and Bailey (1996), and other stimuli not shown in this paper. In stimulus 3, a 400 Hz even harmonic is inserted in an odd complex (harmonics 1, 3, 5,... 15) of a 100 Hz fundamental. The duration of these stimuli is 400 ms (20 ms sine on/off ramp, sampling rate 22 kHz). The final test uses a more realistic sound: Stimulus 4 consists of two synthesized vowels played simultaneously (Culling and Darwin 1993). The vowels are 'a' (100 Hz fundamental) and 'e' (4 semitones higher fundamental). The duration is 200 ms and the sampling rate is 22 kHz. In all of the figures that follow, we compare the predictions of the model at the cessation time of the stimuli with the perceptual data, because the listeners were required to provide pitch matching only after hearing the whole stimulus.

Figure 3a illustrates the simulated AN probability of firing, $p(t, k)$, in each frequency channel in response to stimulus 1 (Roberts and Brunstrom 2001). The plots show an approximately stationary part of the time series of 50 ms duration. In the figure, the clusters of channels identified by the algorithm can be visually appreciated. This preliminary observation motivated the development of the algorithm described in Sect. 2.

Table 1 shows the values of the correlation dimension, D_2, corresponding to the short time series shown in Fig. 3a (and to the rest of the stimuli used in this report, see Appendix for further details). Figure 3b shows the corresponding $\tilde{D}(t, k)$ values during the first 200 ms of stimulus 1. $\tilde{D}(t, k)$ and $D_2(t, k)$ are not directly comparable, because the computation of $\tilde{D}(t, k)$ uses time series values starting at $t = 0$, but D_2 uses only a stationary section of the time series (Fig. 3a). In addition, the $\tilde{D}(t, k)$ values are typically smaller, because they are computed in a transformed phase space which is emptier than the original one, as it was explained in Sect. 2.1.

Figure 3b illustrates how the $\tilde{D}$ values fluctuate up to approximately 80–100 ms and then stabilize progressively throughout the stimulus duration. The plot shows three clear boundaries between adjacent channels (see dashed lines in Fig 3). As indicated in Sect. 2.2, the high $\tilde{D}$ values at 200 ms in channels having centre frequencies over 6,200 Hz do not form a cluster, because it would contain only three channels. Therefore, four clusters of channels emerge after approximately 80–100 ms. The convention used throughout the study is to label cluster 1 as the one that contains mainly the segregated component (as further explained below), and the rest of the clusters are labelled starting from the lower frequency channel (Fig. 3b).

Table 1 Approximate D_2 values using the Takens estimator (Theiler 1988) of a stationary section of the time series studied in this report (values in parentheses are computed using the slope of a linear region)

Channel centre frequency (Hz)	Stimulus			
	1	2	3	4
100	0.9	0.9	0.5(0.2)	0.5
131	1.0	1.0	0.5(0.2)	0.5
167	0.4	0.4	0.7(0.3)	0.4
207	0.4	0.4	0.4(0.1)	0.5(0.2)
252	0.4	0.4	0.7(0.4)	0.7(0.2)
303	0.7	0.5	0.6(0.2)	0.7(0.2)
361	0.6	0.6	0.6(0.3)	0.7(0.3)
426	0.6	0.6	0.5(0.4)	0.8(0.4)
499	0.8	0.9	0.7(0.5)	0.9(0.5)
581	0.8	0.8	0.8	0.8(0.5)
674	0.8	1.0	1.0	0.7(0.5)
779	1.2	1.1	1.3	1.0(0.5)
897	1.5	1.6	1.3	0.8(0.5)
1032	1.7	1.4	1.7	1.0
1183	1.7	1.7	1.9	0.9
1353	1.8	1.6	1.7	1.2
1545	1.9	1.8	1.8	1.3
1761	2.0	1.9	1.8	1.5
2006	2.1	2.0	1.8	1.6
2281	2.2	2.1	1.7	1.4
2592	2.1	2.1	1.4	1.5
2943	1.7	2.0	0.8	1.5
3339	1.1	1.4	0.8	1.2(0.9)
3785	1.1	1.0	1.5(0.8)	1.4(1.1)
4289	1.1	1.2	1.5(0.8)	1.3(1.0)
4857	1.4	1.4	1.7(0.9)	1.6(1.0)
5498	1.3	0.8	1.7(1.0)	1.5(1.0)
6222	1.3	0.9	1.5(0.9)	1.3(0.9)
7038	1.2	1.4	1.5(0.9)	1.3(0.9)
7958	1.1	1.3	1.5(0.9)	1.5(0.9)

The listeners report two salient pitches in stimulus 1: the 200 Hz fundamental and another pitch at 1,152 Hz (average over all listeners). For the corresponding in-tune complex (partials 1–12 of 200 Hz), the listeners report only the 200 Hz pitch.

When all of the channels sum together, the maximum occurs at 200 Hz in both the in-tune and the mistuned complex (Meddis and Hewitt 1991a, b). Figure 4a shows the final S-function responses (at 400 ms) of both stimuli within each

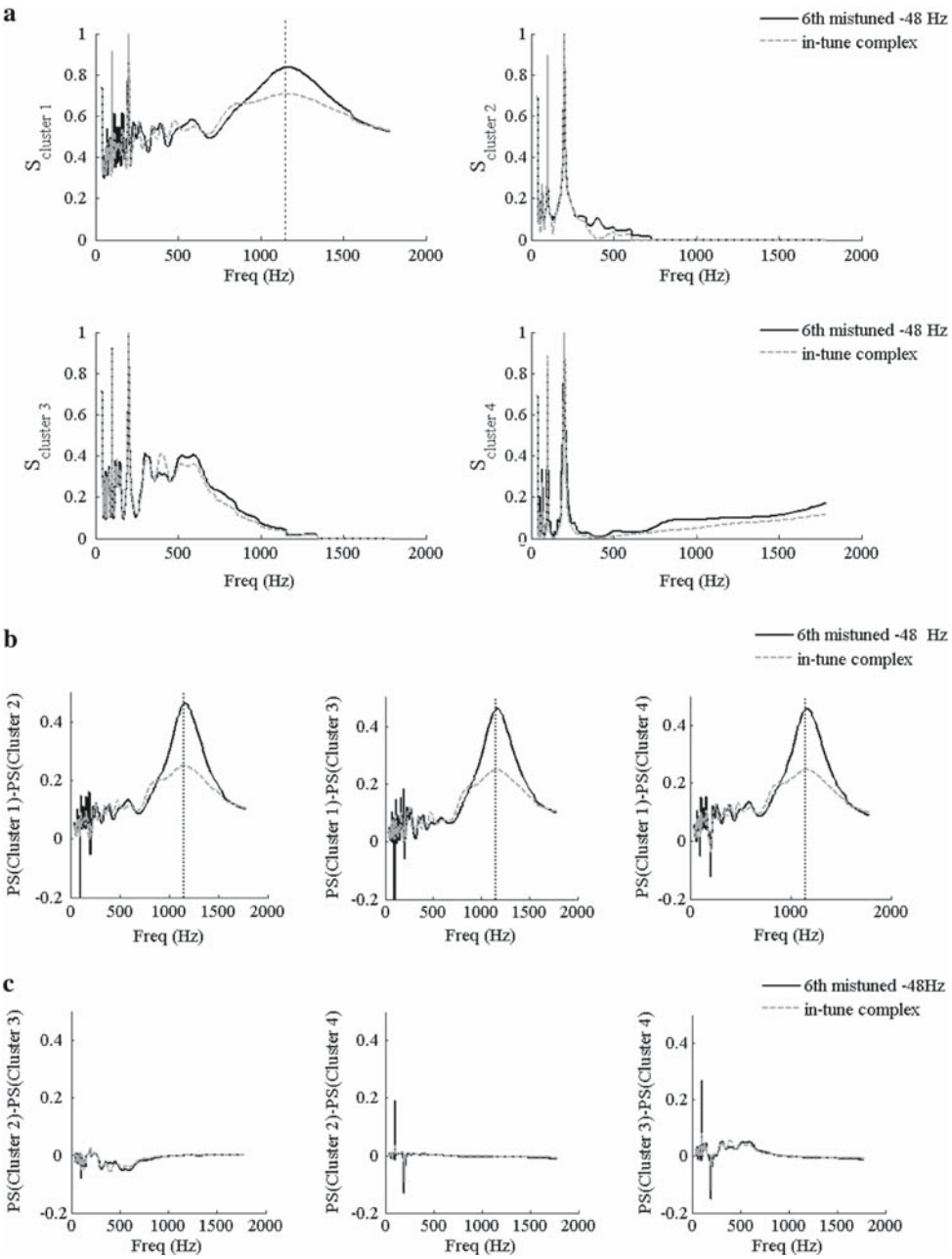


Fig. 4 **a** S_{cluster} of the cochlear model responses to stimulus 1 in the four channel clusters indicated in Fig. 3b. **b** Difference between pitch strength of stimulus 1 in cluster 1 and the pitch strength in clusters 2, 3 and 4. **c** Pitch strength differences between clusters 2, 3 and 4. The *dashed line* shows the S_{cluster} responses and the pitch strength subtractions for the base (in-tune) harmonic complex (1–12 components of a 200 Hz fundamental). The *vertical dotted line* corresponds to the perceptually segregated component (Roberts and Brunstrom 2001); which corresponds also to the maximum of the pitch strength subtractions for stimulus 1 (*solid line* in plot **b**). Therefore, the model correctly predicts the saliency of this segregated pitch (see text)

of the four clusters (Eq. 9). The four responses of the in-tune and the four responses of the mistuned complex are similar, and show a maximum at 200 Hz as expected. However, the *S* response in cluster number 1 presents a more prominent peak around 1,152 Hz in the mistuned complex (indicated by a vertical line in Fig. 4a). This response suggests that cluster number 1 is the one that contains the information about the segregated pitch.

The next step is the transformation of this information into a saliency rating. Figure 4b and 4c show the pitch strength subtraction functions of the four *S*-responses (Eq. 9) corresponding to stimulus 1. The dashed line also shows the pitch strength computations for the corresponding in-tune harmonic complex (1–12 harmonics of 200Hz). The six subtractions between clusters 2, 3 and 4 approximately cancel each other except at 200 and 100 Hz (Fig. 4c). Subtraction of the pitch strength functions with respect to cluster 1 gives a more interesting result for stimulus 1 (Fig. 4b): the maximum pitch strength difference corresponds to the segregated component pitch (1,152 Hz), indicated by the dotted vertical line. Figure 4b also shows how the in-tune harmonic complex presents significantly lower pitch strength in the frequency of the segregated component, as reported in Brunstrom and Roberts (2001).

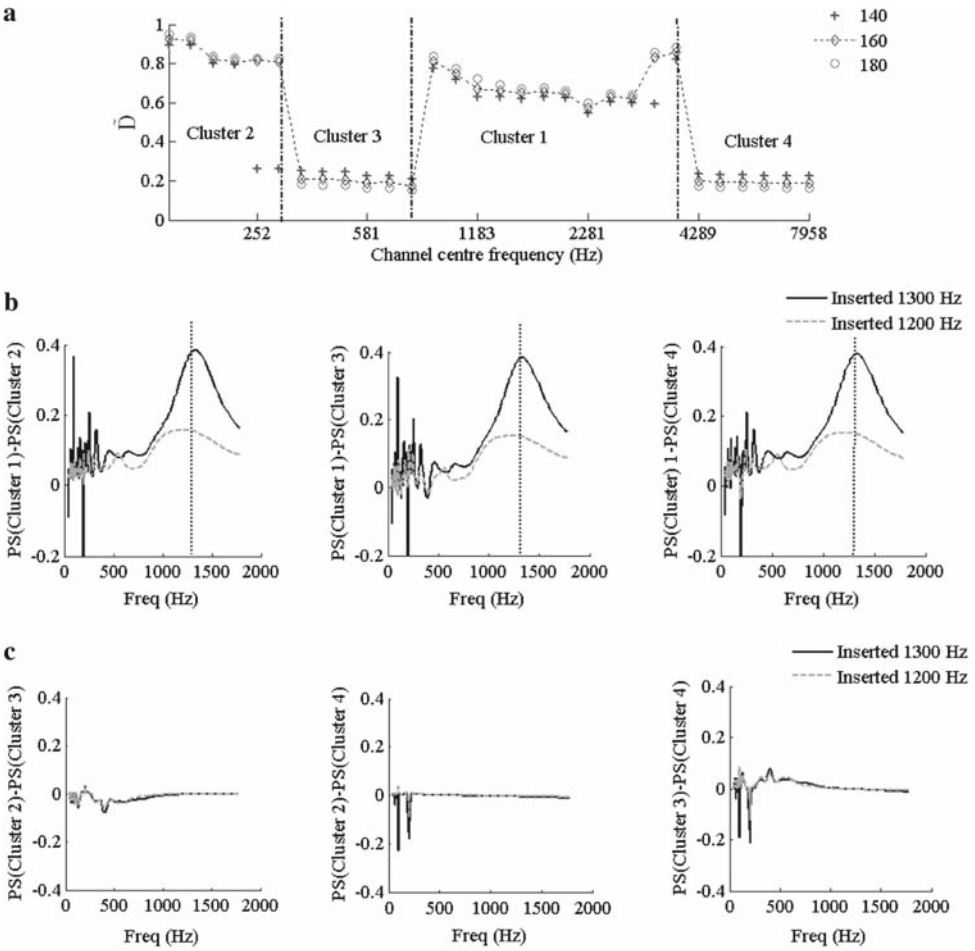


Fig. 5 **a** $\tilde{D}$ values of the auditory nerve simulated firing probability of stimulus 2. The three curves represent the $\tilde{D}$ values computed at 140, 160 and 180 ms. **b** Difference between pitch strength of stimulus 2 in cluster 1 and pitch strength in clusters 2, 3 and 4. **c** Pitch strength differences between clusters 2, 3 and 4. The *dashed line* in plots **b** and **c** shows the same subtractions for the base (in-tune) harmonic complex (harmonics 1 to 5 and 8 to 14 of a 200 Hz fundamental). The vertical dotted line corresponds to the perceptually segregated component (Brunstrom and Roberts 1998), which corresponds also to the maximum of the pitch strength subtractions for stimulus 2 in plot **b** (*solid line* in plot **b**). Therefore, the model predicts the saliency of this segregated pitch (see text)

In summary, considering the twelve possible pitch strength subtractions, the 1,152 Hz component is the most salient pitch (rated 0.47), apart from the 200 Hz fundamental. The next pitch in importance is 100 Hz, but its saliency is considerably smaller (0.27, see right plot in Fig. 4c). The saliency of the 1,152 Hz pitch for the in-tune complex is much lower (0.24). As indicated above, this is consistent with the perceptual experiments (Roberts and Brunstrom 2001).

Other manipulations of the spectrum can produce perceptual segregation of a stimulus frequency. Brunstrom and Roberts (1998) removed the central part of the spectrum of a base harmonic complex, and afterwards they studied the saliency of a component when inserted in different positions within the base complex. When the inserted component is out of tune (stimulus 2), it is more salient. When the inserted component is one of the harmonics of the fundamental (*in-tune*), it is least salient. Figure 5a illustrates the corresponding $\tilde{D}$ values of the cochlear frequency channels computed at 140, 160 and 180 ms. It shows how the values stabilize after 140 ms. The values at the stimulus cessation time are similar to the values shown in Fig. 3b. The maximum *S*-function response in the four clusters corresponds to the 200 Hz fundamental, as is predicted by the standard pitch models (not shown). The pitch strength subtractions with respect to cluster 1 show again a sharp peak in the perceptually segregated component (1,300 Hz, dotted vertical line in Fig. 5b). Other pitch strength subtractions have a significant maximum at the fundamental pitch of 200 Hz and a smaller value at 100 Hz (Fig. 5b, c). Moreover, the pitch strength subtraction values when an in-tune component is inserted (1,200 Hz) are much lower (dashed lines in Fig. 5b). Therefore, the model response predicts that 1,300 Hz is the second most salient pitch (rated 0.4) after the 200 Hz fundamental, and the saliency of in-tune components inserted is much lower (0.15). This result again matches the perceptual experiments (Brunstrom and Roberts 1998).

Stimulus 3 presents a perceptual segregation phenomenon that challenged previous pitch models (Roberts and Bailey 1996). The standard autocorrelation model (Meddis and O'Mard 1997) predicts a maximum pitch strength response for the 100 Hz fundamental, as expected. As in the stimuli presented before, the insertion of an even component distorts the regularity of an otherwise odd harmonic complex. However, in this stimulus, all components are still harmonics of the same fundamental frequency; therefore, it is likely to be more difficult to explain the reported segregation of the

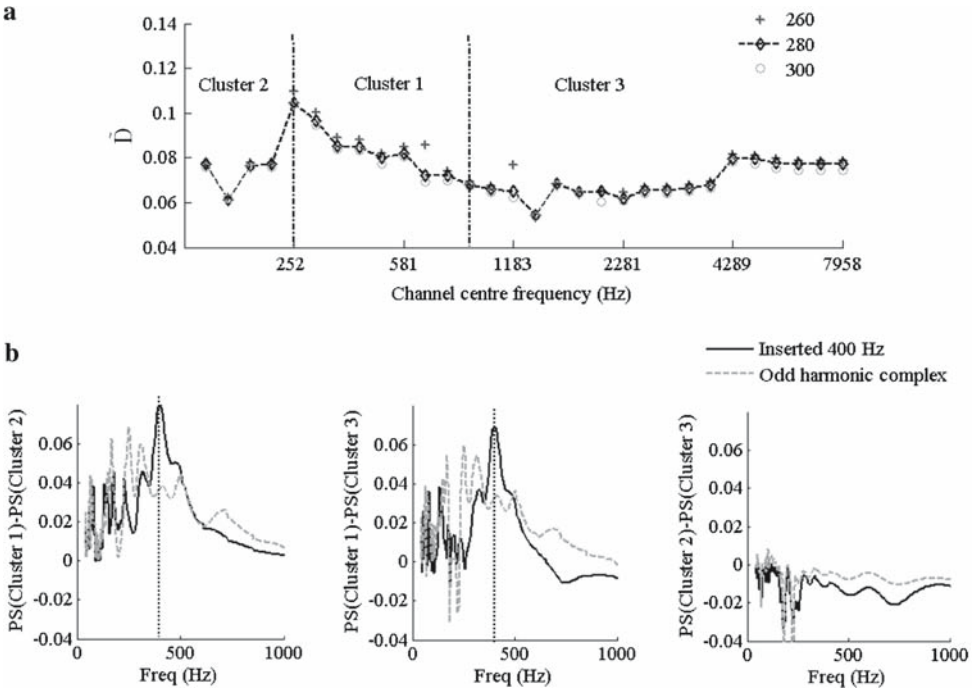


Fig. 6 **a** $\tilde{D}$ values of the auditory nerve simulated firing probability of stimulus 3. The three curves represent the $\tilde{D}$ values computed at 260, 280 and 300 ms. **b** Pitch strength subtractions of stimulus 3 in clusters 1, 2 and 3. The *dashed line* shows the same subtractions for the base harmonic complex (1–15 odd components of 100 Hz fundamental). The vertical line corresponds to the perceptually segregated component (Roberts and Bailey 1996) and to the maximum of the pitch strength subtractions for stimulus 3 (*solid line*). Therefore, the model predicts the saliency of this segregated component (see text)

even component. In this stimulus, the nonlinear downsampling of the original phase space (Sect. 2.1) produces values of $\tilde{D}(t,k)$ much lower than one (Fig. 6a).

The $\tilde{D}(t,k)$ values shows that only three clusters emerge after 280 ms. The boundary that separates clusters 1 and 3 is not easily visible. It is worth remembering from Sect. 2.2 that the cluster 3 begins whenever a channel has $\tilde{D}(t,k)$ values that differ by more than 50% from any of the channels of the previous cluster (in this case, cluster number 1).

The pitch strength subtractions between the clusters (Fig. 6) predict that the perceptually segregated component (400 Hz) is the most salient pitch after the 100 Hz fundamental. This segregated pitch shows low pitch strength differences for an odd harmonic complex (dashed lines in Fig. 6). The overall values in Fig. 6 are much smaller than for stimuli 1 and 2, consistent with the reportedly lower perceptual saliency of the segregated 400 Hz component.

To conclude, Fig. 7 shows the model behaviour for a synthesized speech sound (stimulus 4). This stimulus reported high identification rates of the presence of two vowels (Culling and Darwin 1993). The clustering shown in the left plot can be understood in similar terms to the clustering shown in Fig. 6a. In addition, the three initial channels do not form a cluster, because clusters are restricted to having four or more channels (Sect. 2.2).

The relative pitch strengths between the two clusters predict the saliency of the 'a' vowel pitch (100 Hz); indicated by dotted vertical lines. The maximum saliency associated with the 'e' vowel is located at 135.1 Hz (approximately one semitone higher than its 126 Hz fundamental). Therefore, the model predicts approximately the saliency of the two vowel pitches.

4 Discussion

This study shows how the delay lines used in temporal models of pitch perception can be employed to generate clusters of the auditory peripheral channels. Existing autocorrelation models at the level of the AN have difficulties in explaining perceptual segregation, and it was necessary to incorporate these further computations to provide a remedy. However, there exist very successful functional approaches. Kalpuri's (2003) signal processing algorithm outperforms the ability of trained musicians to separate concurrent sounds; this method produces an iterative estimation of the fundamental also using principles of harmonicity and spectral filtering at the level of the stimulus waveform. The present model, however, has a different aim. It focuses on preserving some biological constraints while explaining the listener's perceptions. Firstly, it uses a physiological peripheral model. Then, it computes simultaneously a characteristic magnitude of the simulated AN spike probabilities time series, $\tilde{D}$, and the pitch. The computations require the availability of the delayed AN signals and a bank of coincidence-detector units as in autocorrelation models of pitch perception (Meddis and O'Mard 1997). Tonotopy and the existence of delay lines are the key factors in the aggregation of a subset of distances in phase space (Eqs. 6, 7) simultaneously with the running autocorrelations (Eq. 1). Both integrations are parallelized for the lags

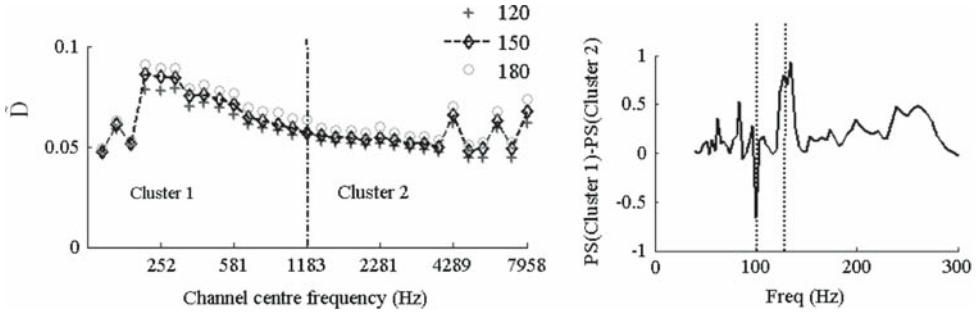
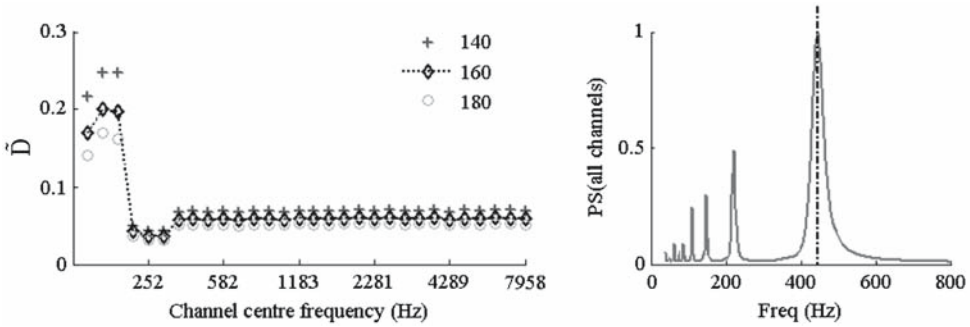


Fig. 7 $\tilde{D}$ values of the auditory nerve simulated firing probability of stimulus 4 (*left plot*). The three curves represent the $\tilde{D}$ values computed at 120, 150 and 180 ms. The *right plot* shows the pitch strength subtractions of stimulus 4 between the two clusters of channels emerged. The *vertical lines* correspond to the fundamental of the vowels 'a' (100 Hz) and 'e' (126 Hz). The maximum saliency associated to the 'e' vowel is located at 135.1 Hz (approximately one semitone higher than its 126 Hz fundamental). The predictions of the model are reasonably accurate (see text)

Fig. 8 $\tilde{D}$ values of the auditory nerve simulated firing probability of a 440 Hz pure tone (*left plot*). The three curves represent the $\tilde{D}$ values computed at 140, 160 and 180 ms. The *right plot* shows the pitch strength considering all of the frequency channels (the *vertical line* indicates the pitch percept)



and for each cochlear frequency channel. The sum in Eq. 7 considers only the subset of phase space vectors corresponding to the typical lag spacing that are used in temporal pitch models. Therefore, the algorithm employs the same best frequency sensitivity range as the pitch models.

In summary, the model architecture (Fig. 1) computes simultaneously a restricted correlation sum of the available region of the phase space (Eqs. 6, 7) and the autocorrelations (Eq. 1) of the frequency channels. For both operations, it employs biologically plausible leaky integrators. The plausibility of autocorrelation models of pitch has been controversial (de Cheveigné and Pressnitzer 2006). However, recently Meddis and O'Mard (2006) implemented a standard autocorrelation model (Eqs. 1, 9), using a leaky integrate-and-fire network in the cochlear nucleus and inferior colliculus.

An advantage of Eqs. 6 and 7 is the mathematical interpretation: the earlier stages of the auditory system are sufficiently equipped to compute a measure of the cochlear frequency channels inspired, but not constrained, by the correlation dimension algorithm. Moreover, the $\tilde{D}$ definition does not require, in principal, the stationarity of the time series. Clearly, Sect. 3 results indicate that $\tilde{D}$ is a plausible alternative to the correlation dimension to characterize the temporal responses in different frequency channels even when they cannot be considered "wide-sense stationary". The Appendix shows an interpretation of the algorithm in terms of more biologically meaningful quantities.

In the next stage, the $\tilde{D}$ values forms clusters of frequency channels; suggesting a stochastic model of lateral interactions between units arranged in clusters. An across-channel interaction has been observed experimentally (Hancock et al. 1997; Winter 2005). Therefore, the peripheral clustering method is local in frequency, but it is not restricted to adjacent channels. This is consistent with Roberts and Holmes (2006) perceptual results. They reported that the adjacent frequency channels are the most influential in the perceptual segregation of a component from the complex. The contribution of distant channels reduces with the centre frequency difference, but cannot be completely neglected.

The pitch strength subtractions between clusters provide a measure of relative degree of saliency of the components of the sound. This difference produces a reduction or cancellation of the pitch strength of the fundamental, and imitates the effect of the harmonic sieve-like models of pitch perception (Goldstein 1973; de Cheveigné 1998, 2005; de Cheveigné and Kawahara 2002). This relative pitch strength predicts correctly the second most important perceptual entity reported by the listeners. Roberts and Holmes (2006) successfully modelled the degree of saliency of the fundamental pitch itself (partial pitch-shift) when the other components in the

complex (the frame) are randomly mistuned. As in this paper, they used an across-channel weighting of the summarized autocorrelogram (Eq. 9). The presented model provides a rationale for the introduction of these weights: they relate to the conditional probabilities of a channel belonging to a cluster, $P(k|cluster;t)$.

5 Concluding remarks

An interesting aspect of this study is the use of the neuronal delay lines to compute simultaneously the pitch and a characteristic measure of the simulated auditory nerve firing probability time series. The predictions of the model include the perceptually segregated components in manipulated complex stimuli and speech. It also suggests a mathematical interpretation of the kind of computations that take place after peripheral processing of sound.

Acknowledgements This work was supported by EmCAP (Emergent Cognition through Active Perception, 2005–2008); a research project in the field of Music Cognition funded by the European Commission (FP6-IST, contract 013123). We thank Milica for proofreading the manuscript. E.B wishes to thanks Thomas Wennekers for his generous support.

Appendix

Other parameters used in the algorithm take the values indicated below. An updated Matlab© implementation of this method is freely available.

Number of lags (l): 200, from 0.3 to 25 ms, distributed in an equivalent rectangular bandwidth scale (Moore and Glasberg 1983; Denham 2005).

Number of ε: 100 (from stdev$(p(t,k))/200$ to stdev $(p(t,k))/2$ in stdev$(p(t,k))/2$ steps, as indicated in Kantz and Schreiber (1999).

Minimum number of neighbourhood vectors contained in a sphere of radius ε: 10,000 (otherwise the Eq. 4 computations are not further considered). There were no rejections in the stimuli tested in this paper.

The quantity $\tilde{C}(t,\varepsilon,k)$ can be easily interpreted in terms of more biologically meaningful magnitudes. For convenience, we wrote the distances in phase space as

$$\begin{aligned} M(t,l,k) &= \|\vec{x}(t)-\vec{x}(t-l)\|^2 \\ &= -F(t,l)+E(t)+E(t-l)-F(t-L,l) \\ &\quad +E(t-L)+E(t-l-L), \end{aligned} \tag{A1}$$

where $F(t,l)=2p(t)p(t-l)$; $E(t)=p(t)^2$; and the channel number k is omitted in the following equations. Next, we used half-wave rectification nonlinearity, $[\varepsilon - M(t,k,l)]_+$ instead of a threshold function (i.e., $[\varepsilon - M(t,k,l)]_+ = M(t,k,l)$ if $\varepsilon > M(t,k,l)$ and zero elsewhere). Thus, the sum across the lags in Eq. 7 is

$$\begin{aligned} \sum_{l(\varepsilon,t)} [\varepsilon - M(t,k,l)]_+ &= \text{number}\{l(\varepsilon,t)\}\cdot(E(t)+E(t-L)) \\ &\quad +\sum_{l(\varepsilon,t)} E(t-l)+E(t-L-l) \\ &\quad -\sum_{l(\varepsilon,t)} (F(t,l)+F(t-L,l)), \end{aligned} \tag{A2}$$

where $l(\varepsilon,t)$ is the set of lags corresponding to distances $M(t,l,k)$ smaller than ε at time t. Then, the Eq. 6 sum is approximately

$$\begin{aligned} \tilde{C}(t,\varepsilon,k) &\approx \sum_t \sum_{l(\varepsilon,t)+\{0\}} \beta(t)\cdot(E(t-l)+E(t-L-l)) \\ &\quad -\sum_{l(\varepsilon,t)+\{0\}} \sum_t (F(t,l)+F(t-L,l)), \end{aligned} \tag{A3}$$

where $\beta(t)$ =number of $l(\varepsilon,t)$ lags if $l=0$; $\beta(t)=1$ for $l \in l(\varepsilon,t)$ and otherwise zero. The positive summand, $\sum_t \sum_l \beta(t)(E(t-l)+E(t-L-l)$, is a double temporal integration (using a weighting factor β) of energy terms. The term $\sum_t (F(t,l)+F(t-L,l)$ are autocorrelations of the AN simulated signal at time t for the peripheral channel number k.

The existence of D_2 (Eq. 4) requires a strongly linear plateau in the $\ln(C(t,\varepsilon,k))$ versus $\ln\varepsilon$ plot (Kantz and Schreiber 1999). In our closely related calculations, we are looking for a similar linear plateau in the relationship between $\ln(\tilde{C}(t,\varepsilon,k))$ and $\ln\varepsilon$. We define $\tilde{D}$ using an algorithm, which uniquely characterizes this slope. $\tilde{D}$ is the slope of the linear part of $\ln(\tilde{C}(t,\varepsilon,k))$ versus $\ln\varepsilon$ that shows the maximum r^2 with the following restrictions: the required minimum length of the region is 30 consecutive epsilon values; and the minimum required r^2 is 0.98, $p=0.001$. This algorithm provides automatically unambiguous $\tilde{D}$ values in all of the stimuli tested, because both restrictions are satisfied uniquely in all of the cases. Therefore, using this restrictive definition, the statistical errors in $\tilde{D}$ within this set of points are negligible and are omitted in the figures.

Table 1 shows the approximate D_2 values using the Takens estimator (Theiler 1988) of the AN output in different frequency channels. We selected a 50 ms steady-state section of the stimuli used in this paper. These values are not used in the present study and cannot be directly compared with $\tilde{D}$, as explained in Sect. 3.1.

The estimation uses the longest linear region of the plot $\ln C(\varepsilon)$ versus $\ln\varepsilon$. The D_2 defined as the slope of this plot differs, in some cases from the Takens estimator (Kantz and Schreiber 1999), these values are indicated in parenthesis. The order of magnitude of the errors is 10^{-1}, therefore, the values are rounded to the first decimal point. The time separation plots method (Provenzale 1992) indicates that $i_{\min}$ should vary approximately from 20 to 250 samples (Eq. 3), depending on the time series. The ε ranges and the minimum

number of neighbourhood vectors contained in a sphere of radius ε takes the same values as the ones used in the model.

To conclude, the consistency of the method presented in this paper is demonstrated using a pure tone stimulus. The left and middle plots in Fig. 8 illustrate the $\tilde{D}$ values for the cochlear model responses to a 440 Hz sinusoid computed at 140,160 and 180 ms. The $\tilde{D}(t,k)$ values fluctuate up to 140 ms and, after that, stabilize at the same value for the majority of the channels (left plot).

As indicated in Sect. 2, the fluctuations of the $\tilde{D}$ values in the first three channels are negligible and a single cluster groups together the thirty frequency channels. Therefore, the pitch strength computation is strongly peaked only at 440 Hz (right plot), as expected.

References

Balaguer-Ballester E, Denham S (2006) Unified temporal method for perceptual fusion and pitch perception. In: Fifteenth annual computational neuroscience meeting CNS 31–32

Balaguer-Ballester E, Palomares A, Martin JD, Soria E (2006) Predicting service request in support centres based on nonlinear dynamics, ARMA models and neural networks. Expert Syst Appl 10:1016–1025

Bernstein JGW, Oxenham AJ (2005) An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination. J Acoust Soc Am 117:3816–3831

Brunstrom JM, Roberts B (1998) Profiling the perceptual suppression of partials in periodic complex tones:Further evidence for harmonic template. J Acoust Soc Am 104:3511–3519

Brunstrom JM, Roberts B (2000) Separate mechanisms govern the selection of the spectral components for perceptual fusion and for the computation of global pitch. J Acoust Soc Am 107:1566–1577

Culling JF, Darwin CJ (1993) Perceptual separation of simultaneous vowel: within and across-formant grouping by F0. J Acoust Soc Am 93:3454–3467

Dayan P, Abbot LF (2001) Theoretical neuroscience. Cambridge University press, Cambridge USA

Cheveigné A (2005) Pitch perception models. In: de Plack CJ, Oxenham AJ, Fay RR, Popper AN (eds) Pitch: neural coding and perception. Springer, New York

Cheveigné A, Kawahara H (2002) YIN, a fundamental frequency estimator for speech and music. J Acoust Soc Am 111:1917–1930

Cheveigné A, Pressnitzer D (2006) The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction. J Acoust Soc Am 119:3908–3918

Cheveigné A (1998) Cancellation model of pitch perception. J Acoust Soc Am 103:1261–1271

Denham SL (2005) Dynamic Iterated Ripple Noise: further evidence for the importance of temporal processing in auditory perception. Biosystems 79:199–206

Goldstein JL (1973) An optimum processor theory for the central information of the pitch of complex tones. J Acoust Soc Am 54:1496–1516

Grose JH, Hall JW, Buss E (2002) Virtual pitch integration for asynchronous harmonics. J Acoust Soc Am 112:2956–2961

Hall JW, Peters RW (1981) Pitch for nonsimultaneous succesive harmonics in quiet and noise. J Acoust Soc Am 69:509–513

Hancock KE, Davis KA, Voltg HF (1997) Modelling inhibition of type II units in the dorsal cochlear nucleus. Biol Cybern 76:417–428

Hartman WM (1996) Pitch periodicity and auditory organization. J Acoust Soc Am 100:3491–3502

Kantz H, Schreiber T (1999) Nonlinear time series analysis. Cambridge University press, Cambridge USA

Krumbholz K, Patterson RD, Seither-Preisler A, Lammertmann C, Lutenhoner L (2003) Neuromagnetic evidence for a pitch processing centre in Heschl's Gyrus. Cerebral Cortex 13:765–772

Li JY, Hartman WM (1998) The pitch of a mistuned harmonic: evidence for a template model. J Acoust Soc Am 103:2608–2617

Licklider J (1951) A duplex theory of pitch perception. Experientia 7:128–134

Licklider J (1959) Three auditory theories. In: Koch S (eds) Psychology a study of a science. McGraw-Hill, New York, pp 41–144

Lopez-Poveda EA, Meddis R (2001) A human nonlinear cochlear filter bank. J Acoust Soc Am 110:3170–3118

Lyon RH (1984) Range and frequency dependence of transfer function phase. J Acoust Soc Am 76:1433–1439

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

Meddis R, Hewitt MJ (1991) Virtual pitch and phase sensitivity of a computer model of the auditory periphery: II. Phase sensitivity. J Acoust Soc Am 89:2883–2894

Meddis R, Hewitt MJ (1992) Modelling the identification of concurrent vowels with different fundamental frequencies. J Acoust Soc Am 91:233–245

Meddis R, O'Mard L (1997) A unitary model of pitch perception. J Acoust Soc Am 102:1811–1820

Meddis R, O'Mard L (2006) Virtual pitch in a computational physiological model. J Acoust Soc Am 120:3861–3868

Moore BCJ, Glasberg BR (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J Acoust Soc Am 74:750–753

Plack CJ, White LJ (2000) Perceived continuity and pitch perception. J Acoust Soc Am 108:1162–1169

Provenzale A, Smith LA, Vio R, Murante G (1992) Distinguishing between low-dimensional dynamics and randomness in measured time series. Physica D 5:28–31

Roberts B (2005) Spectral pattern Grouping and the pitches of complex tones and their components. Acta Acustica United Acustica 91:945–957

Roberts B, Bailey PJ (1996) Spectral regularity as a factor distinct from harmonic relations in auditory grouping. J Exp Physchol Hum Percept Perform 22:604–614

Roberts B, Brunstrom JM (2001) Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch. J Acoust Soc Am 110:2479–2490

Roberts B, Holmes SD (2006) Grouping and the pitch of a mistuned fundamental component: Effects of applying simultaneous multiple mistunings to the other harmonics. Hear Res 222:79–88

Sauer T, Yorke JA, Casdagli M (1991) Embedology. J Stat Phys 65:579–590

Sumner CJ, Lopez-Poveda EA, O'Mard LP, Meddis R (2003) Adaptation in a revised inner-hair cell model. J Acoust Soc Am 113:893–901

Takens F (1981) Detecting strange attractors in turbulence. Springer Lecture Notes in Mathematics vol. 898. Springer, New York

Theiler J (1988) Lacunarity in a best estimator of fractal dimension. Phys Lett A 135:195–210

Wiegrebe L (2001) Searching for the time constant in of neural pitch extraction. J Acoust Soc Am 107:1082–1091

Winter IM (2005) The neurophysiology of pitch. In: Plack CJ, Oxenham AJ, Fay RR, Popper AN (eds) Pitch: neural coding and perception.. Springer, New York

Yost WA (1996) Pitch and pitch strength of iterated rippled noise: Is it the envelope or fine structure?. J Acoust Soc Am 100:2720–2730