The dynamics of a sensory apparatus: The case of the auditory system

Julyan H. E. Cartwright*, Diego L. González[†] and Oreste Piro**

**Laboratorio de Estudios Cristalográficos, CSIC, Granada, Spain*

[†]Laboratorio di Acustica Musicale e Architettonica, Fondazione Cini-CNR, Venezia, Italy

***Institut Mediterrani d'Estudis Avançats, UIB-CSIC, Palma de Mallorca, Spain*

Abstract. The brain has to process and react to an enormous amount of information from the senses in real time. How is all this information represented and processed within the nervous system? A proposal of nonlinear and complex systems research is that dynamical attractors may form the basis of neural information processing. Here we show that this idea can be successfully applied to the human auditory system, and can explain our perception of pitch.

Keywords: auditory system, pitch perception, residue pitch, three-frequency resonances

PACS: 05.45.-a, 43.66.+y, 87.19.La

The pitch of a sound is where we perceive it to lie on a musical scale. For a pure tone with a single frequency component, there is a monotonic relationship between pitch and frequency. However, more complex signals also elicit a pitch sensation; see the stimuli in Fig. 1. All the stimuli, which may be termed complex tones, have a certain spectral periodicity. Many natural sounds exhibit this property, including vowel sounds in human speech and vocalizations of many other animals, and also sounds produced by the nonlinear interaction of two or more periodic sources, for example by amplitude or frequency modulation, and all of them produce a definite pitch sensation.

The scientific study of the pitch of sounds dates back to Pythagoras, but the mechanisms of pitch perception are still not fully understood. Models of pitch perception may be grouped into two main categories: place or spectral models consider that pitch is mainly related to spectral or Fourier properties of the stimulus [1], whereas periodicity or temporal models hold that its characteristics in the time domain are more important [2]. However, these models do not take into account the rôle played by active nonlinearities in pitch perception. Here we demonstrate that the pitch of complex tones can be described by three-frequency resonances: universal responses of nonlinear systems to quasiperiodic forcing.

Evidence for the importance of spectral periodicity in sound processing by humans is that noisy stimuli exhibiting this property also elicit a pitch sensation. An example is repetition pitch: the pitch of iterated ripple noise [3], which arises naturally when the sound from a noisy source interacts with a delayed version of the same, produced, for example, by a single or multiple echo. Thus it becomes clear that an efficient mechanism for the analysis and recognition of complex tones represents an evolutionary advantage for an organism. In this light, the pitch percept may be seen as an effective one-parameter categorization of sounds possessing some spectral periodicity [4].

For a harmonic stimulus like Fig. 1b, there is a natural physical solution to the

FIGURE 1. Stimuli: waveforms, Fourier spectra, and pitches. (a) 1 kHz pure tone; the pitch coincides with the frequency ω_0 . (b) Complex tone formed by 200 Hz fundamental plus overtones; the pitch is at the frequency of the fundamental ω_0 . (c) After high-pass filtering of the previous tone to remove the fundamental and the first few overtones, the pitch ω_0 remains at the frequency of the missing fundamental (dotted). (d) The result of frequency modulation of a 1 kHz pure tone carrier by a 200 Hz pure tone modulant; the pitch coincides with the difference combination tone ω_0 . (e) Complex tone produced by amplitude modulation of a 1 kHz pure tone carrier by a 200 Hz pure tone modulant; the pitch coincides with the difference combination tone ω_0 . (f) Result of shifting the partials of the previous tone in frequency by $\Delta\omega = 90$ Hz; the pitch shifts by $\Delta\omega_0 \approx 20$ Hz, although the difference combination tone does not. (g) Schematic diagram of the frequency line details (above the line) the pitch shift behaviour of (f) and (below the line) the three-frequency resonance we propose to explain it.

problem of encoding it with a single parameter: take the fundamental component of the stimulus as the pitch and all other components are naturally encoded as the higher harmonics of the fundamental. This is what nature does. However, a harmonic stimulus like Fig. 1c, which is high-pass filtered such that the fundamental and some of the first higher harmonics are eliminated, nevertheless maintains its pitch at the frequency of the absent fundamental. The stimulus (Fig. 1e) obtained by amplitude modulation of a sinusoidal carrier of 1 kHz by a sinusoidal modulant of 200 Hz is also of this type. As the carrier and modulant are rationally related, the stimulus is harmonic, the partials being integer multiples of the absent fundamental $\omega_0 = 200$ Hz. The perception of pitch for this kind of stimulus is known as the problem of the missing fundamental, virtual pitch, or residue perception [5]. A first physical theory for the phenomenon is due to von Helmholtz [6], who attributed it to the generation of difference combination tones in the nonlinearities of the ear. A passive nonlinearity fed by two sources with frequencies ω_1 and ω_2 generates combination tones of frequency ω_C (see the Appendix for clarification of the concepts from nonlinear dynamics used throughout the paper). For a harmonic complex tone such as Fig. 1e the difference combination tone $\omega_C = \omega_2 - \omega_1$ between two successive partials has the frequency of the missing fundamental ω_0 . However, in a crucial experiment, Schouten et al. [7] demonstrated that the behaviour of the residue cannot be described by a difference combination tone: if we shift all the partials in frequency by the same amount $\Delta\omega$ (Fig. 1f), the difference combination tone remains unchanged. However, the perceived pitch shifts, with a linear dependence on $\Delta\omega$.

The complex tone is now anharmonic. So how does nature encode an anharmonic complex tone into a single pitch? Intuitively, the shifted pseudofundamental depicted in Fig. 1g might seem to be a better choice than the unshifted fundamental, which corresponds to the difference combination tone. However, from a mathematical point of view, this is not obvious. The ratios between successive partials of the shifted stimulus are irrational and thus we cannot represent them as higher harmonics of a fundamental frequency because the true fundamental has frequency zero in this case. Some kind of approximation is needed. The approximation of two arbitrary frequencies ω_1 and ω_2 by the harmonics of a third one ω_R is equivalent to the mathematical problem of finding a strongly convergent sequence of pairs of rational numbers with the same denominator that simultaneously approximates the two frequency ratios ω_1/ω_R and ω_2/ω_R . If we consider the approximation to only one frequency ratio there exists a general solution given by the continued-fraction algorithm [8]. However, for two frequency ratios a general solution is not known. Some approximations have been proposed that work for particular values of the frequency ratios or which are weakly convergent [9]. An alternative approach we developed [10] has interesting dynamical applications. The idea is to equate the distances between appropriate harmonics of the pseudofundamental and the pair of frequencies we wish to approximate. In this way the two approximations are equally good or bad. The problem can then be solved by a generalization of the Farey sum [10]. This approach allows for the hierarchical classification of a class of dynamical attractors found in systems with three frequencies: three-frequency resonances $[p, q, r]$.

A classification of three-frequency resonances allows us to propose that nature might encode an anharmonic complex tone into a single pitch on the following basis: the pitch of a complex tone corresponds to a one-parameter categorization of sounds by means of a physical frequency whose harmonics are good approximations of the partials of

the complex. This physical frequency is naturally generated as a universal response of a nonlinear dynamical system under the action of an external force represented by the stimulus. Psychophysical experiments with multicomponent stimuli suggest that it is the lowest-frequency components that are dominant in determining residue behaviour [5]. Thus we represent the external force to a first approximation by the two lowest frequency components of the stimulus [11]. For pitch shift experiments such as those of Schouten et al. with small frequency detuning $\Delta\omega$, the vicinity of these two lowest components $\omega_1 = k\omega_0 + \Delta\omega$ and $\omega_2 = (k+1)\omega_0 + \Delta\omega$ to successive multiples of some missing fundamental ensures that $(k+1)/k$ is a good rational approximation to their frequency ratio. Hence we concentrate on a small interval between the frequencies ω_1/k and $\omega_2/(k+1)$ around the missing fundamental of the nonshifted situation. These frequencies corresponds to the three-frequency resonances $[0, -1, k]$ and $[-1, 0, k+1]$. We suppose that the residue should be associated with the largest three-frequency resonance in this interval: the daughter of these resonances, $[-1, -1, 2k+1]$. If our reasoning is correct, the three-frequency resonance formed between the two lower-frequency components of the complex tone and the response frequency $P = (\omega_1 + \omega_2)/(2k+1)$ gives rise to the perceived residue pitch P .

In Fig. 2 we have superimposed the behaviour of the corresponding three-frequency resonances on published experimental pitch-shift data [7, 12, 13]. There is good agreement with the three-frequency resonance produced by the two lowest-frequency components of the complex tone for intermediate harmonic numbers $3 \leq k \leq 8$ [11]. For high and low k values there are systematic deviations from predictions made using the two lowest components of the complex tone. Such deviations, noted in pitch-perception modelling, are explained by the dominance effect: peripheral prefiltering creates a frequency window of preferred stimulus components, so that not all components are equally important in determining residue perception [14]. For stimuli consisting only of high k components, the window of the dominance region is almost empty, and difference combination tones of lower k can become more important than the primary components in determining the pitch of the stimulus. In order to describe these slope deviations for high and low k values within our approach, it suffices, instead of taking the lowest component, to take some effective k that depends on the dominance effect, in which we take into account also the presence of difference combination tones, which provide some components with k s not present in the original stimulus. For higher values of k , the result of the modification is a saturation of the slopes that correctly describes the experimental data.

Here we wish to concentrate here on the more complex case of low k stimuli. For these stimuli not only quantitative but also qualitative differences arise between the two-lowest-component theory [11] and experiment. As with high values of k , a saturation of slopes can be seen in the experimental data for decreasing values of k . This effect can be explained in terms of the dominance region. For a 200 Hz stimulus spacing, the region is situated at about 800 Hz; this implies that stimulus components with harmonic numbers n and $n+1$ other than the two lowest ones (i.e., $n > k$) become more important for determining the the three-frequency resonance that provides the residue pitch. The more interesting feature, however, which can be observed in Fig. 2, is a second series of pitch-shift lines centred around the pitch of 100 Hz. To understand these we must recall that the three-frequency resonance is determined using the property that for small frequency detuning $\Delta\omega$ the frequency ratio between adjacent stimulus components

can be approximated by the quotient of two integers differing by unity, i.e. $\omega_2/\omega_1 = (n+1)/n$. However, if we relax the small detuning constraint, so that $\Delta\omega$ becomes large, we can move to a situation where this approximation is no longer valid, where ω_2/ω_1 can be better approximated by $(n+2)/(n+1)$. But, by the usual Farey sum operation between rational numbers, we know that there exists between these two regions an interval where the frequency ratio can be better approximated by $(2n+3)/(2n+1)$. In this interval, then, the main three-frequency resonance is $[-1, -1, 4n+4]$, giving a response frequency $P = (\omega_1 + \omega_2)/(4n+4)$, which produces a pitch-shift line with slope $1/(2n+2)$ centred at $\omega_0/2 = 100$ Hz for the case analysed. Of course, if prefiltering produces a saturation of the slopes of the primary pitch-shift lines, the same should also occur for these secondary ones. In Fig. 2 we also show our predictions for the secondary lines taking in account the dominance effect. The agreement, both qualitative and also quantitative, is impressive. Moreover, a small group of data points indicates the existence of a tertiary level of pitch-shift centred at 50 Hz in the region between a primary and a secondary pitch-shift line. We can understand this tertiary level in the same way as above, and we plot our prediction for the tertiary pitch-shift line in Fig. 2. This is clear evidence for the hierarchical arrangement of the perception of pitch of complex tones entirely consistent with the universal structure that dynamical systems theory predicts for the three-frequency resonances in quasiperiodically forced dynamical systems. Further evidence comes from psychophysical experiments with pure tones. These, presented under particular experimental conditions, also elicit a residue sensation. The extremes of the three-frequency staircase correspond to subharmonics of only one external frequency and thus these are the expected responses when only one stimulus component is present. As the results of Houtgast [15] show, these subharmonics are indeed perceived.

A dynamical attractor can be studied by means of time or frequency analysis; both are common techniques in dynamical systems analysis, but one is not inherently more fundamental than the other, nor are these the only two tools available. For this reason, our results cannot be included either in the spectral [1] or the temporal [2] classes of models of pitch perception. What we have developed is not another model, but a metamodel: a mathematical basis for the perception of pitch that uses the universality of responses of dynamical systems to address the question of why the auditory system should behave as it does when confronted by stimuli consisting of complex tones. Not all pitch perception phenomena are explicable in terms of universality; nor should they be, since some will depend on the specific details of the neural circuitry, however this is a powerful way of approaching the problem that is capable of explaining many data considered difficult to understand. Future pitch models can surely incorporate these results in their frameworks; spectral theories [1], because they make consistent use of different kinds of harmonic templates and three-frequency resonances offer in a natural way optimized candidates for the base frequency of such templates without the need to include stochastic terms; temporal theories [2], because they need some kind of locking of neural spiking to the fine structure of the stimulus and, as we have shown, three-frequency resonances are the natural extension of phase locking to the more complicated case of quasiperiodic forcing which is typically related to the perception of complex tones.

We have shown that universal properties of dynamical responses in nonlinear systems are reflected in the pitch perception of complex tones. In previous work [11], we have argued that a dynamical-systems approach backs up experimental evidence for subcor-

tical pitch processing in humans [16]. The experimental evidence is not conclusive, as studies with monkeys have found that raw spectral information is present in the primary auditory cortex [17]. However, whether this processing occurs before, or in, the auditory cortex, the dynamical mechanism we envisage greatly facilitates processing of information into a single percept.

We have left out of this analysis the question of what the output of a dynamical system representing the auditory system should be when fed with other stimuli apart from complex tones. In work yet to be published, we show that these ideas are also able to account for phenomena such as the pitch of iterated ripple noise that we mentioned in the introduction. Pitch processing may then prove to be a further example in which universality in nonlinear dynamics can explain complex experimental results in biology. The auditory system possesses an astonishing capability for real-time pitch-related information processing; here we have demonstrated why, at a fundamental level, this is so.

FIGURE 2. (Color online) Experimental data (red dots) from Gerson & Goldstein [12] (0–800 Hz range) and from Schouten et al. [7] (1200–2200 Hz range) show pitch as a function of the lower frequency $f = k\omega_0 + \Delta\omega$ of a complex tone $\{k\omega_0 + \Delta\omega, (k+1)\omega_0 + \Delta\omega, (k+2)\omega_0 + \Delta\omega, \dots\}$ with spacing $g = \omega_0 = 200$ Hz apart. The data of Schouten et al. are for three-component tones monotonically presented (all of the stimulus entering one ear); those of Gerson & Goldstein for four-component tones dichotically presented (part of the stimulus entering one ear and the rest of the stimulus entering the contralateral ear); the harmonic numbers of the partials present in the stimuli are shown beside the data. The pitch-shift effect we predict from the resonance model, taking into account the dominance region, is shown superimposed on the data as solid lines given by the Eqs. $P = g + (f - ng)/(n - 1)$ (primary lines), $P = g/2 + (f - (n + 1/2)g)/(2n + 2)$ (secondary lines), and $P = g/4 + (f - (n - 1/4)g)/(4n + 1)$ (tertiary lines); the harmonic numbers used to calculate the pitch-shift lines are shown enclosed in red squares; for primary lines these harmonic numbers correspond to n and $n + 1$, for secondary lines to $2n + 1$ and $2n + 3$, and for tertiary lines to $4n + 1$ and $4n + 5$. A red circle, instead of a square, signifies that the component is not present in the stimulus, but corresponds to a combination tone. The inset graph displays the slopes of the data averaged over the distinct experimental conditions as a function of harmonic number. The blue squares are the data of Gerson & Goldstein, the red squares are those of Schouten, and the blue diamonds are those of Patterson [13] for six and twelve-component tones. The black diamonds correspond to our theory.

APPENDIX

Universal behaviour of nonlinear systems

Nonlinear systems exhibit universal responses under external forcing:

Harmonics from periodically forced passive nonlinearities

A single frequency periodically forcing a passive nonlinearity generates higher harmonics (overtones) $2\omega_1, 3\omega_1, \dots$ of a fundamental ω_1 : given by $p\omega_1 + \omega_H = 0$ with p integer. This is seen in acoustics as harmonic distortion.

Combination tones from quasiperiodically forced passive nonlinearities

A passive nonlinearity forced quasiperiodically by two sources generates combination tones $\omega_1 - \omega_2, \omega_1 + \omega_2, \dots$, which are solutions of the Eq. $p\omega_1 + q\omega_2 + \omega_C = 0$ where p and q are integers. They are found as distortion products in acoustics.

Subharmonics, or two-frequency resonances from periodically forced dynamical systems

With a periodically forced active nonlinearity — a dynamical system — more complex subharmonic responses $\omega_1/r, 2\omega_1/r, \dots, (r-1)\omega_1/r$ known as mode lockings or two-frequency resonances are generated. These are given by $p\omega_1 + r\omega_{2R} = 0$ when p and

r are integers. As some parameter is varied, different resonances are found that remain stable over an interval. A classical representation of this, known as the devil's staircase, is shown below. The rotation number, which in this case coincides with the frequency ratio, i.e., $\rho = -p/r = \omega_{2R}/\omega_1$, is plotted against the period of the external force.

We see that the resonances are hierarchically arranged. The local ordering can be described by the Farey sum: If two rational numbers a/b and c/d satisfy $|ad - bc| = 1$ we say that they are unimodular or adjacents and we can find between them a unique rational with minimal denominator. This rational is called the mediant and can be expressed as a Farey sum operation $a/b \oplus c/d = (a+b)/(c+d)$. The resonance characterized by the mediant is the widest between those represented by the adjacents [18].

Three-frequency resonances from quasiperiodically forced dynamical systems

Quasiperiodically forced dynamical systems show a great variety of qualitative behaviour that falls into three main categories: there are periodic attractors, quasiperiodic attractors, and chaotic and nonchaotic strange attractors. Here we concentrate on the three-frequency resonances produced by two-frequency quasiperiodic attractors as the natural candidates for modelling the residue [19]. Three-frequency resonances are given by the nontrivial solutions of the Eq. $p\omega_1 + q\omega_2 + r\omega_{3R} = 0$, where p , q , and r are integers, ω_1 and ω_2 are the forcing frequencies, and ω_{3R} is the resonant response, and

can be written compactly in the form $[p, q, r]$. Combination tones are three-frequency resonances of the restricted class $[p, q, 1]$. This is the only type of response possible from a passive nonlinearity, whereas a dynamical system such as a forced oscillator is an active nonlinearity with at least one intrinsic frequency, and can exhibit the full panoply of three-frequency resonances, which include subharmonics of combination tones. Three-frequency resonances obey hierarchical ordering properties very similar to those governing two-frequency resonances in periodically forced systems. In the interval $(\omega_2/p, \omega_1/q)$, we may define a generalized Farey sum between any pair of adjacents as $a_1/c \oplus a_2/d = (a_1 + a_2)/(c + d)$. The daughter three-frequency resonance characterized by the generalized median is the widest between its parents characterized by the adjacents [10]. Thus three-frequency resonances are ordered very similarly to their counterparts in two-frequency systems, and form their own devil's staircase:

Contrarily to the case of periodically driven systems, where plateaux represent periodic solutions, here they represent quasiperiodic solutions (only the third frequency is represented in the vertical axis). We have investigated these properties in three different systems: the quasiperiodic circle map, a system of coupled electronic oscillators and a set of ordinary nonlinear differential equations, with the same qualitative results, which confirm the theoretical predictions [20].

ACKNOWLEDGMENTS

J.H.E.C. acknowledges the financial support of the Spanish Ministerio de Ciencia y Tecnología grant CTQ2004-04648; O.P. acknowledges the financial support of grants CONOCE2 (FIS2004-00953) and HIELOCRIS (200530F0052). This review is based in part on our article [19].

REFERENCES

1. M. A. Cohen, S. Grossberg, and L. L. Wyse, *J. Acoust. Soc. Am.* **98**, 862–878 (1995).
2. R. Meddis, and M. J. Hewitt, *J. Acoust. Soc. Am.* **89**, 2866–2882 (1991).
3. W. A. Yost, *J. Acoust. Soc. Am.* **100**, 511–518 (1996).
4. B. Roberts, and P. J. Bayley, *J. Exp. Psychol.* **22**, 604–614 (1996).
5. E. de Boer, “On the “residue” and auditory pitch perception,” in *Handbook of Sensory Physiology. Auditory System*, edited by W. D. Keidel, and W. D. Neff, Springer, 1976, vol. V, pp. 479–584.
6. H. L. F. von Helmholtz, *Die Lehre von dem Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, Braunschweig, 1863.
7. J. F. Schouten, R. J. Ritsma, and B. L. Cardozo, *J. Acoust. Soc. Am.* **34**, 1418–1424 (1962).
8. A. Y. Kinchin, *Continued Fractions*, University of Chicago Press, 1964.
9. S. Kim, and S. Ostlund, *Phys. Rev. Lett.* **55**, 1165–1168 (1985).
10. J. H. E. Cartwright, D. L. González, and O. Piro, *Phys. Rev. E* **59**, 2902–2906 (1999).
11. J. H. E. Cartwright, D. L. González, and O. Piro, *Phys. Rev. Lett.* **82**, 5389–5392 (1999).
12. A. Gerson, and J. L. Goldstein, *J. Acoust. Soc. Am.* **63**, 498–510 (1978).
13. R. D. Patterson, *J. Acoust. Soc. Am.* **53**, 1565–1572 (1973).
14. R. D. Patterson, and F. L. Wightman, *J. Acoust. Soc. Am.* **59**, 1450–1459 (1976).
15. T. Houtgast, *J. Acoust. Soc. Am.* **60**, 405–409 (1976).
16. C. Pantev, M. Hoke, B. Lütkenhöner, and K. Lehnertz, *Science* **246**, 486–488 (1989).
17. Y. I. Fishman, D. H. Reser, J. C. Arezzo, and M. Steinschneider, *Brain Res.* **786**, 18–30 (1998).
18. D. L. González, and O. Piro, *Phys. Rev. Lett.* **50**, 870–872 (1983).
19. J. H. E. Cartwright, D. L. González, and O. Piro, *Proc. Natl. Acad. Sci. USA* **98**, 4855–4859 (2001).
20. O. Calvo, J. H. E. Cartwright, D. L. González, O. Piro, and O. Rosso, *Int. J. Bifur. & Chaos* **9**, 2181–2187 (1999).