Nonlinear Dynamics of the Perceived Pitch of Complex Sounds

Julyan H. E. Cartwright[1,*], Diego L. González[2,†], and Oreste Piro[3,‡]
[1] *Instituto Andaluz de Ciencias de la Tierra, IACT (CSIC-UGR), E-18071 Granada, Spain.*
[2] *Istituto Lamel, CNR, I-40129 Bologna, Italy*
[3] *Institut Mediterrani d'Estudis Avançats, IMEDEA (CSIC–UIB), E-07071 Palma de Mallorca, Spain*
(Physical Review Letters, **82**, 5389–5392, 1999)

We apply results from nonlinear dynamics to an old problem in acoustical physics: the mechanism of the perception of the pitch of sounds, especially the sounds known as complex tones that are important for music and speech intelligibility.

PACS numbers: 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. Like all sensations, pitch is a subjective quantity related to physical attributes of the stimulus, in this case mainly its component frequencies. For a pure tone with a single frequency component, the relation is monotonic and enables us to adopt an operational definition of pitch in terms of frequency: the pitch of an arbitrary sound is given by the frequency of a pure tone of the same pitch. The scientific investigation of pitch has a long history dating back to Pythagoras, but the origin of the pitch of complex tones with several frequency components is still not well understood. The first perceptual theories considered pitch to arise at a peripheral level in the auditory system [1–5]; more recently it has been thought that central nervous system processing is necessary [6–8]. However, the experimental evidence is that this processing is carried out before the primary auditory cortex [9]. The latest models integrate neural and peripheral processing [10,11]. Here we develop a nonlinear theory of pitch perception for complex tones that describes experimental results on the pitch perception of complex sounds at least as well as do current models, and that removes the need for extensive processing at higher levels of the auditory system.

A key phenomenon in pitch perception is known as the problem of the *missing fundamental*, *virtual pitch*, or *residue perception* [12], and consists of the perception of a pitch that cannot be mapped to any frequency component of the stimulus. Suppose that a periodic tone such as that shown in Fig. 1a is presented to the ear. Its pitch is perceived to be that of a pure tone at the frequency of the fundamental. The number of higher harmonics and their relative amplitudes give timbral characteristics to the sound, which allow one to distinguish a trumpet from a violin playing the same musical note. Now suppose that the fundamental and some of the first few higher harmonics are removed (Fig. 1b). Although the timbre changes, the pitch of the tone remains unchanged and equal to the missing fundamental; this is residue perception.

The first physical theory for the residue is due to von Helmholtz [3], 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, which are nontrivial solutions of the equation $p\omega_1 + q\omega_2 + \omega_C = 0$, where p and q are integers. For a harmonic complex tone (Fig. 1b), the difference combination tone $\omega_C = \omega_2 - \omega_1$ (i.e., $p = 1$, $q = -1$) between two successive partials has the frequency of the missing fundamental ω_0, that is $\omega_C = (k+1)\omega_0 - k\omega_0 = \omega_0$. However, a crucial experiment seriously challenged nonlinear theories of the residue. Schouten et al. [4] demonstrated that the behavior of the residue cannot be described by a difference combination tone: if we shift all the partials by the same amount $\Delta\omega$ (Fig. 1c), the difference combination tone remains unchanged, and the same should thus be true of the residue. Instead it is found that the perceived pitch also shifts, showing a linear dependence on $\Delta\omega$ (Fig. 1d). This phenomenon is known as the first pitch shift effect, and has been accurately measured in many experiments (psychoacoustic experiments on pitch can attain an accuracy of 0.2% [13]). A first attempt to model qualitatively the behavior of the pitch shift shows that the slopes of the lines in Fig. 1d depend roughly on the inverse of the harmonic number $k + 1$ of the central partial of a three-component complex tone. However, for small k at least, the change in slope is slightly but consistently larger than this, but smaller if we replace $k + 1$ by k. This behavior is known as the second pitch shift effect. Also, an enlargement of the spacing between partials while maintaining fixed the central frequency produces a decrease in the residue pitch. As this anomalous behavior seems to be correlated with the second pitch shift effect, it is usually included within it [4]. Pitch-shift experiments, and others that demonstrated that the residue is elicited dichotically (part of the stimulus exciting one ear and the rest of the stimulus the other, controlateral, ear) and not just monotically and diotically (all of the stimulus exciting one or both ears) led to the abandonment of peripheral (periodicity [2,4] and place [1,3]) theories and to the development of theories that considered the pitch of complex tones to be a result of central nervous system processing [6–8]; thence to integrated neural and peripheral models [10,11].

We demonstrate below that the crucial pitch-shift experiment of Schouten et al. can be accurately described in terms of generic attractors of nonlinear dynamical systems, such that a theory of pitch perception for complex tones can be constructed without the need to resort to extensive central processing. We model the auditory system as a generic nonlinear forced oscillator, and identify experimental data with structurally stable behavior of this class of dynamical system. We emphasize that since we are interested in universal behavior, the results we obtain are not dependent on the construction of a particular model, but rather represent the behavior of any dynamical system of this type.

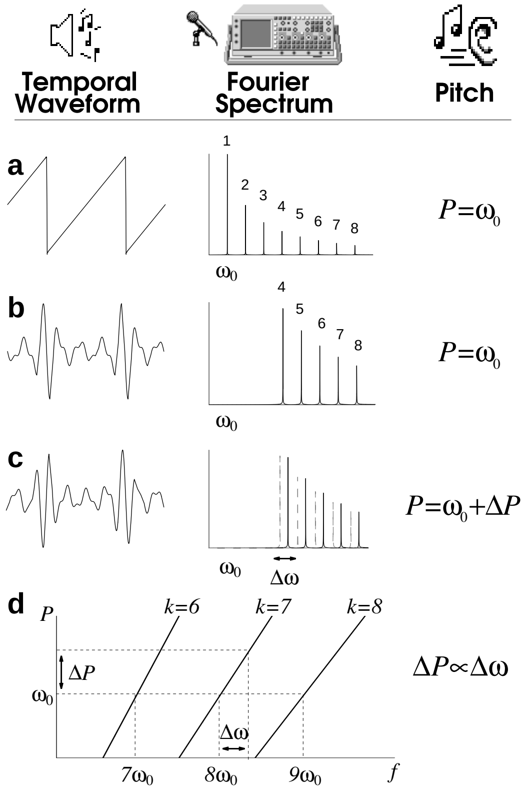

FIG. 1. Fourier spectra and pitch of complex tones. Whereas pure tones have a sinusoidal waveform corresponding to a single frequency, almost all musical sounds are complex tones that consist of a lowest frequency component, or fundamental, together with higher frequency overtones. The fundamental plus overtones are together collectively called partials. **a** A harmonic complex tone. The overtones are successive integer multiples $k = 2, 3, 4 \ldots$ of the fundamental ω_0 that determines the pitch. The partials of a harmonic complex tone are termed harmonics. **b** Another harmonic complex tone. The fundamental and the first few higher harmonics have been removed. The pitch remains the same and equal to the missing fundamental. This pitch is known as virtual or residue pitch. **c** An anharmonic complex tone. The partials, which are no longer harmonics, are obtained by a uniform shift $\Delta\omega$ of the previous harmonic case (shown dashed). Although the difference combination tones between successive partials remain unchanged and equal to the missing fundamental, the pitch shifts by a quantity ΔP that depends linearly on $\Delta\omega$. **d** Pitch shift. Pitch as a function of the central frequency $f = (k + 1)\omega_0 + \Delta\omega$ of a three-component complex tone $\{k\omega_0+\Delta\omega, (k+1)\omega_0+\Delta\omega, (k+2)\omega_0+\Delta\omega\}$. The pitch-shift effect is shown here for $k = 6$, 7, and 8. Three-component complex tones are often used in pitch experiments because they elicit a clear residue sensation and can easily be obtained by amplitude modulation of a pure tone of frequency f with another pure tone of frequency ω_0. When ω_0 and f are rationally related, $\Delta\omega = 0$, and the three frequencies are successive multiples of some missing fundamental. At this point $\Delta P = 0$, and the pitch is ω_0, coincident with the frequency of the missing fundamental.

Since complex tones can be decomposed as a series of partials — a sum of purely sinusoidal components — and residue perception is elicited with at least two of these, we search for suitable attractors in the class of two-frequency quasiperiodically forced oscillators. Quasiperiodically forced oscillators show a great variety of qualitative behavior that falls into the three categories of periodic attractors, quasiperiodic attractors, and strange attractors (both chaotic and nonchaotic). We propose that the residue behavior we seek to explain is a resonant response to forcing the auditory system quasiperiodically. From stability arguments [14–17] we single out a particular type of two-frequency quasiperiodic attractor, which we term a three-frequency resonance, as the natural candidate for modelling the residue. Three-frequency resonances are given by the nontrivial solutions of the equation $p\omega_1+q\omega_2+r\omega_R = 0$, where p, q, and r are integers, ω_1 and ω_2 are the forcing frequencies, and ω_R is the resonant response, and can be written compactly in the form (p, q, r). Notice that 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. The investigation of three-frequency systems is a young area of research, and there is not yet any consistency in the nomenclature of these resonances in the scientific literature: as well as three-frequency resonances, they are also called weak resonances or partial mode lockings; see Baesens et al. [18] and references therein. It is known that three-frequency resonances form an extensive web in the parameter space of a dynamical system. In particular, between any two parent resonances (p_1, q_1, r_1) and (p_2, q_2, r_2) lies the daughter resonance $(p_1 + p_2, q_1 + q_2, r_1 + r_2)$ on the straight line in parameter space connecting the parents.

For pitch-shift experiments, the vicinity of the external frequencies 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 around the missing fundamental between the frequencies ω_1/k and $\omega_2/(k + 1)$, which correspond to the resonances $(0, -1, k)$ and $(-1, 0, k + 1)$. We suppose that the residue should be associated with the largest resonance in this interval. In numerical simulations and experiments with electronic oscillators [14–17], we have confirmed that for small nonlinearity the daughter of these resonances, $(-1, -1, 2k + 1)$, is the resonance of greatest width between its parents. Hence we propose that 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. Since in pitch-shift experiments the external frequencies are $\omega_1 = k\omega_0 + \Delta\omega$ and $\omega_2 = (k+1)\omega_0+\Delta\omega$, the shift of the response frequency with respect to the missing fundamental is $\Delta P = 2\Delta\omega/(2k + 1)$. This equation gives a linear dependence of the shift on the detuning $\Delta\omega$, in agreement with the first pitch shift effect.

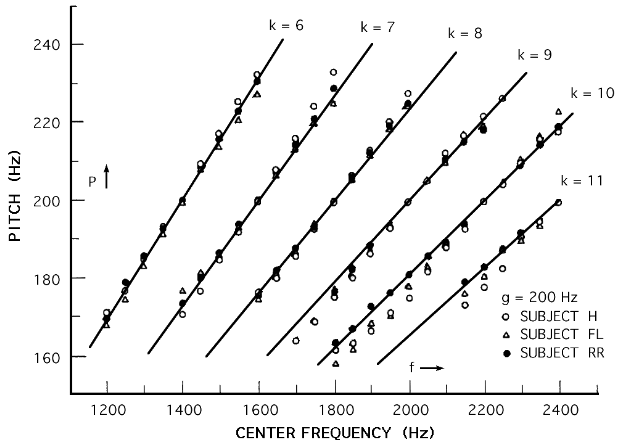

FIG. 2. Experimental data with our predictions superimposed show pitch as a function of the central frequency $f = (k+1)\omega_0+\Delta\omega$ for a three-component complex tone $\{f - g, f, f + g\}$, for $k = 6$, 7, 8, 9, 10, and 11. The component spacing is $g = \omega_0 = 200$ Hz. Circles, triangles and dots represent experimental data for three different subjects (from Schouten et al. [4]). The perceived pitch shifts linearly with the detuning $\Delta\omega$. The pitch shift effect we predict is shown superimposed on the data as solid lines satisfying the equation $P = g + 2(f - (k + 1)g)/(2k + 1)$. The lines describe the behavior of the response frequency in a three-frequency resonance formed with the two lowest-frequency components of the stimulus. The lines agree with the psychophysical data without any fitting parameters.

In Fig. 2 we have superimposed the behavior of the corresponding three-frequency resonances on data for the pitch shift for three different subjects. There is good agreement, which explains both the first pitch shift effect and the first aspect of the second pitch shift effect, because the predicted slope is $1/(k+1/2)$, between $1/k$ and $1/(k+1)$. The second aspect of the second pitch shift effect can be interpreted as follows: the term $2\Delta\omega$ in the equation for ΔP arises from the two equal contributions $\Delta\omega$ obtained by a uniform shift in the two forcing frequencies. If, while maintaining ω_2 fixed, we increase the interval to ω_1 — we enlarge the spacing between successive partials — the first contribution remains constant and equal to $\Delta\omega$ while the second diminishes, to give a decrease in the response frequency of the resonance and thus in the residue.

For small harmonic numbers k our model data fall within experimental error bars for most of the pitch-shift experiments reported [4,19,20]. For larger k there are systematic deviations of the residue slopes which become shallower than theoretical predictions of both our theory, and the central theories. An accepted explanation for this is that difference combination tones generated as a consequence of passive nonlinear mixing in the auditory periphery can play the rôle of the lowest frequency components of the stimulus. The same explanation is also valid for our approach. Our goal here, however, is simply to demonstrate that physical frequencies other than combination tones can accurately describe residue behavior in nonlinear terms without fitting parameters, and can overcome the objections against nonlinear theories raised by pitch-shift experiments.

Dichotic perception has been another argument against nonlinear theories, because it has been thought to imply central nervous system processing [21]. However, subcortical nervous paths, for example through the superior olivary complex, allow the exchange of information between both — left and right — peripheral auditory systems. Moreover, frequency information up to 5 kHz is preserved until at least this point in the auditory system [22,23]. These two factors allow frequency components of the stimulus that arrive at different ears to interact at a subcortical level. This implies that three-frequency resonances can be generated, not only monotically but also dichotically, by a mechanism such as that which we are proposing. Experimental evidence for the existence of multifrequency resonant responses is given by electrophysiological records in single units of the auditory midbrain nucleus of the guinea fowl [24]. Our mechanism is also consistent with neuromagnetic measurements performed on human subjects showing that pitch processing of complex tones is carried out before the primary auditory cortex [9].

Thanks to the residue we can appreciate music in a small radio with negligible response at low frequencies; but it is not just an acoustical curiosity. The residue seems to play an important rôle in music perception and speech intelligibility. The importance of the residue for the perception of musical sounds has long since been recognized. It has been proposed [8] that residue perception is at the heart of the fundamental bass of Rameau [25]. As the fundamental bass and its more modern counterparts form a key element for the melodic and harmonic structuration of musical sounds, a physical basis for the residue may contribute to the construction of an objectively grounded theory of music [14]. As for speech intelligibility, hearing aids that furnish fundamental frequency information produce better scores in profoundly hearing impaired subjects than simple amplification [26]. It is clear that a better knowledge of the basic mechanisms involved in pitch perception will allow a similar improvement in its applications to technology and medicine.

The principle of economy, widespread in nature, suggests that by means of our proposed mechanism, complex tones may be preprocessed at a subcortical level, to feed optimized pitch candidates to more central zones of the nervous system, freeing them of an enormous quantity of calculation. In this way, the nonlinear behavior of the auditory periphery may greatly contribute to explain the astonishing real-time analytical capabilities of the auditory system [27].

JHEC and OP acknowledge the financial support of the Spanish Dirección General de Investigación Científica y Técnica, contracts PB94-1167 and PB94-1172. We are pleased to thank Egbert de Boer, Manfred Schroeder, and Gene Stanley for helpful conversations.

* julyan@galiota.uib.es, WWW http://formentor.uib.es/∼julyan.
† diego@indra.lamel.bo.cnr.it.
‡ piro@imedea.uib.es, WWW http://www.imedea.uib.es/∼piro.
[1] G. S. Ohm, Ann. Phys. Chem. **59**, 513 (1843).
[2] A. Seebeck, Ann. Phys. Chem. **60**, 449 (1843).
[3] H. L. F. von Helmholtz, *Die Lehre von dem Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (Braunschweig, 1863).
[4] J. F. Schouten, R. J. Ritsma, and B. L. Cardozo, J. Acoust. Soc. Am. **34**, 1418 (1962).
[5] R. Plomp, J. Acoust. Soc. Am. **41**, 1526 (1967).
[6] J. L. Goldstein, J. Acoust. Soc. Am. **54**, 1496 (1973).
[7] F. L. Wightman, J. Acoust. Soc. Am. **54**, 407 (1973).
[8] E. Terhardt, J. Acoust. Soc. Am. **55**, 1061 (1974).
[9] C. Pantev, M. Hoke, B. Lütkenhöner, and K. Lehnertz, Science **246**, 486 (1989).
[10] R. Meddis and M. J. Hewitt, J. Acoust. Soc. Am. **89**, 2866 (1991).
[11] M. A. Cohen, S. Grossberg, and L. L. Wyse, J. Acoust. Soc. Am. **98**, 862 (1995).
[12] E. de Boer, in *Handbook of Sensory Physiology. Auditory System*, edited by W. D. Keidel and W. D. Neff (Springer, 1976), Vol. V, pp. 479–584.
[13] W. M. Hartmann, J. Acoust. Soc. Am. **100**, 3491 (1996).
[14] D. L. González, Ph.D. thesis, Universidad Nacional de la Plata, Argentina, 1987.
[15] O. Calvo *et al.*, IEEE Trans. Circuits and Systems, to appear (1999).
[16] J. H. E. Cartwright, D. L. González, and O. Piro, Phys. Rev. E **59**, 2902 (1999).
[17] O. Calvo *et al.*, Int. J. Bifurcation & Chaos, to appear (1999).
[18] C. Baesens, J. Guckenheimer, S. Kim, and R. S. MacKay, Physica D **49**, 387 (1991).
[19] R. D. Patterson, J. Acoust. Soc. Am. **53**, 1565 (1973).
[20] A. Gerson and J. L. Goldstein, J. Acoust. Soc. Am. **63**, 498 (1978).
[21] A. J. M. Houtsma and J. L. Goldstein, J. Acoust. Soc. Am. **51**, 520 (1972).
[22] N. Y. S. Kiang, *Discharge patterns of single fibers in the cat's auditory nerve* (MIT, 1965).
[23] D. Johnson, J. Acoust. Soc. Am. **68**, 1115 (1980).
[24] G. Langner, Exp. Brain Res. **52**, 333 (1983).
[25] J. P. Rameau, *Traité de l'harmonie* (Paris, 1722).
[26] A. Faulkner *et al.*, J. Acoust. Soc. Am. **91**, 2136 (1991).
[27] Sample sound files to demonstrate the phenomena we describe in this Letter may be found at the URL http://www.imedea.uib.es/∼piro/PitchPage/index.html.