Proceedings of the 8th International Conference on Music Perception & Cognition, Evanston, IL, 2004
ICMPC8
S.D. Lipscomb, R. Ashley, R.O. Gjerdingen, & P. Webster (Eds.)
Adelaide, Australia: Causal Productions

A TEMPORAL MODEL FOR PITCH MULTIPLICITY AND TONAL CONSONANCE

Peter Cariani

Eaton Peabody Laboratory of Auditory Physiology, Mass. Eye & Ear Infirmary,
243 Charles St., Boston MA 02114 USA. Email: cariani@epl.meei.harvard.edu; www.cariani.com

ABSTRACT

One aspect of tonal consonance relevant to the pitch stability of chords and harmonic tension concerns the degree to which a collection of notes produces a unified, strong low pitch. We used a neurophysiologically-grounded temporal model for pitch to test Stumpf's general hypothesis that consonance is based on degree of tonal fusion. Patterns of consonance estimates generated from the model were compared with corresponding human judgments of consonance ("clearness") obtained by Kameoka & Kurogawa (1969).

The model used an auditory nerve simulation (72 fibers, CFs, 80-16000 Hz) to generate neural responses to pairs of tones separated by different musical intervals. The simulation used stimuli identical to those used in psychophysical experiments (pairs of pure tones and harmonic complexes, $F0_{root}$ = 440 Hz, n=1-6). Population-wide interspike interval distributions were then compiled from the responses of individual simulated auditory nerve fibers. The relative strengths of pitch-related interval patterns in these distributions were then measured using subharmonic sieves to yield predictions of which pitche(s) should be heard as well as their relative strengths (saliences). The estimated salience of the strongest pitch (maximum salience) is used as an estimate of pitch fusion and perceived consonance.

Consonance values estimated from the model were then compared with those obtained by K&K. The high degree of correspondences observed (r = 0.88 for "consonances" of pure tones and r = -0.97 for "absolute dissonances" of complex tones) support the viability of pitch-based models for consonance that depend on the strength of the fundamental bass.

1. BACKGROUND

The perception of concurrent and successive combinations of tones is central to psychological theories of musical harmony and melody. Consonance encompasses qualities that are created when tones are sounded together: from sensations of roughness to perceptions of unified, fused wholes (Tramo et al., 2001). Far from simple dependence on acoustics, perceived consonance is influenced by cochlear and neuronally-mediated processes that unfold on at least three different time scales. "Sensory" aspects of consonance are often used to describe perceived tone qualities of concurrent sounds presented in isolation. These aspects of consonance are likely to be mediated by cochlear filtering and the neuronal mechanisms responsible for perception of pitch and roughness. Contextual effects can arise when tones are preceded by other tones, and plausibly involve interactions between neuronal activity patterns associated with incoming sounds and activity patterns produced by sounds in the immediate past (~200 ms to seconds). These short-term contextual effects likely involve the same central neuronal mechanisms associated with echoic and working memory that mediate tonality induction (Leman, 2000) and melodic coherence (Deutsch, 1999; Snyder, 2000). Longer range contextual effects, such as particular patterns of tonal expectancies, almost inevitably involve experientially-acquired, (culturally) conditioned predispositions (hedonic and otherwise) that are mediated by long term memory (Bharucha, 1994).

Perception of consonance in music thus is thought involve "sensory" and "perceptual" processes that are relatively independent of context, as well as "cognitive" processes that (can) depend on musical context, both immediate and distant. We focus here on those "sensory" and "perceptual" aspects of consonance that apply to multiple concurrent tones presented in isolation. Models of consonance that are based on the perception of roughness examine the degree to which nearby harmonics beat (Helmholtz, 1885; Kameoka & Kuriyagawa, 1969a, 1969b; Plomp & Levelt, 1965; Sethares, 1999; Terhardt, 1974a), while pitch-based theories examine the strengths of virtual pitches created (Terhardt et al., 1982) and the degrees to which they fuse together (DeWitt & Crowder, 1987; Green & Butler, 2002; Lipps, 1905; Schneider, 1997).

2. AIMS

Our general aims here parallel those of others (Helmholtz, 1885; Parncutt, 1989; Terhardt, 1974b) who have sought to determine which aspects of musical perception are direct consequences of the physiological mechanisms that subserve auditory perception. Like others before us (Boomsliter & Creel, 1962; Licklider, 1951; Patterson, 1986; Rose, 1980), we also seek to ground pitch-based theories of tonal consonance in interspike interval representations. Recently global, population-wide interval statistics have been used to explain a wide range of musical pitch phenomena (Cariani, 2002; Cariani & Delgutte, 1996a, 1996b; Leman, 2000; Leman & Carreras, 1997; Meddis & Hewitt, 1991; Meddis & O'Mard, 1997; Slaney & Lyon, 1991). Here we present consonance predictions based on computer simulations of auditory nerve responses to dyads of pure and complex tones and compare them to consonance judgments obtained in psychophysical experiments (Kameoka & Kuriyagawa, 1969a, 1969b).

Our ultimate goal is to apply such models for pitch multiplicity to predict the pitch stability of chords so as to directly link low-level physiologically-grounded neuronal representations of pitch with high-level music-theoretic notions of harmonic tension and relaxation. In order to achieve this ultimate goal, models need to incorporate effects of musical context, as embodied by neural processing operations that access neural representations held in both short and long-term memory stores (Bharucha, 1994; Leman, 2000). Such processes would lead to neuronal representations of music that dynamically unfold in time, combining successive presents with successive pasts.

3. METHOD

Auditory nerve model

Auditory nerve responses to dyads consisting of pure and complex tones were simulated in a computer model (Cariani, 2002). Ensembles of 72 auditory nerve fibers (ANFs) were simulated (3 fibers/CF), the 24 characteristic frequencies (CFs) being logarithmically-spaced between 80-16000 Hz. Single fiber responses were modeled using a series of relatively simple operations that parallel basic cochlear and neural processes: middle-ear filtering (attenuation: -20 dB < 100 Hz; 0 dB: 1-2 kHz; -20 dB < 10 kHz), cochlear band-pass frequency tuning (Bekesy-like gammatone filters; (Hartmann, 1998), pp. 254-256, filter order = 3, decay const. b = CF/5), hair cell transduction (half-wave rectification), synaptic filtering (low pass filter, rolloff > 2 kHz), neural rate-level functions, and uncorrelated, endogenous, “spontaneous” activity (Poisson process governed by driven - spontaneous rate (SR)). Three fibers representing three spontaneous rate/threshold classes (low, medium, and high SR) were simulated for each CF. An adaptive gain control similar to that of (Geisler & Greenberg, 1986) scales instantaneous firing rates according to the average input rms over the previous 4 ms so as to replicate rate compression and the quasi-linear character of fast amplitude modulations. Model parameters have been chosen that best replicate population-wide interspike interval distributions (PIDs) compiled from cat auditory nerve data (Cariani & Delgutte, 1996a). The model captures the response characteristics most relevant to temporal coding and auditory form perception: 1) the broadly-tuned nature of auditory nerve fiber responses at moderate-to-high levels, 2) faithful replication of the level-invariant forms of period histograms (Rose et al., 1971), impulse (click) responses, and threshold tuning curves (Kiang et al., 1965), and 3) rate saturation and the general spread of excitation at higher stimulus levels (Kim & Molnar, 1979). In addition, the model shows two-tone rate and synchrony suppression.

Stimuli

Waveforms were synthesized to match stimuli used in specific psychophysical experiments of Kameoka and Kuriyagawa, i.e. (Kameoka & Kuriyagawa, 1969a), fig. 5 and (Kameoka & Kuriyagawa, 1969b), fig. 7. Stimuli of 500 ms duration (44.1 kHz samples/s) were simulated at the specified levels: pure tone dyads at 57 dB SPL per component and complex tone dyads (harmonics 1-6) at 41-57 dB SPL per component.

Interspike interval distributions

The ANF model produces an array of instantaneous-firing-rate time series (post-stimulus time histograms, PSTHs). All-order interspike interval distributions from each fiber were estimated by computing the autocorrelation of the PSTHs (binwith = 0.1 ms). Interval distributions were weighted by SR fiber class and a human CF-distribution to form a population-interval distribution (PID) that serves as an estimate of the global all-order interval statistics of the human auditory nerve (Cariani & Delgutte, 1996a).

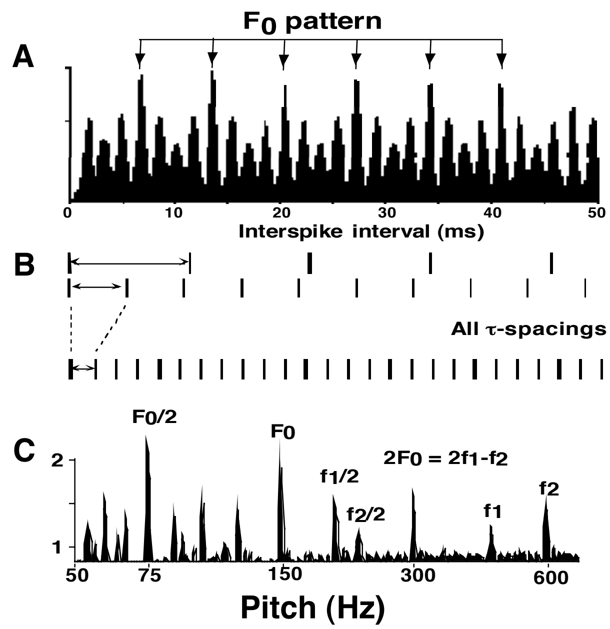

Figure 1. Pitch multiplicity and salience estimation. A. PID in response to two pure tones a fourth apart, 440 & 587 Hz. B. Representative sub-harmonic interval sieves for estimating pattern-saliences. C. Pitch map: observed distribution of pattern salience values. Exponential interval weighting window is not shown.

Pitch salience estimation

Our working hypothesis is that the neural correlate of a periodicity pitch percept involves a pattern of interspike intervals (e.g. intervals at the pitch period (1/F0) and its multiples (n/F0, the "undertone" series) rather than the relative number of intervals at only the pitch period itself. We want to estimate the relative pattern-strengths of different sets of interspike intervals that are associated with different pitches. The estimation procedure consists of two operations, the weighting of interspike intervals by their respective durations and application of periodic sieve that measures relative numbers of intervals associated with each pitch (Fig 1).

Interval weighting

Although we desire to include intervals at multiples of pitch periods in our analysis, a limit on the maximum interval duration that is processed in a central autocorrelation analyzer is needed to

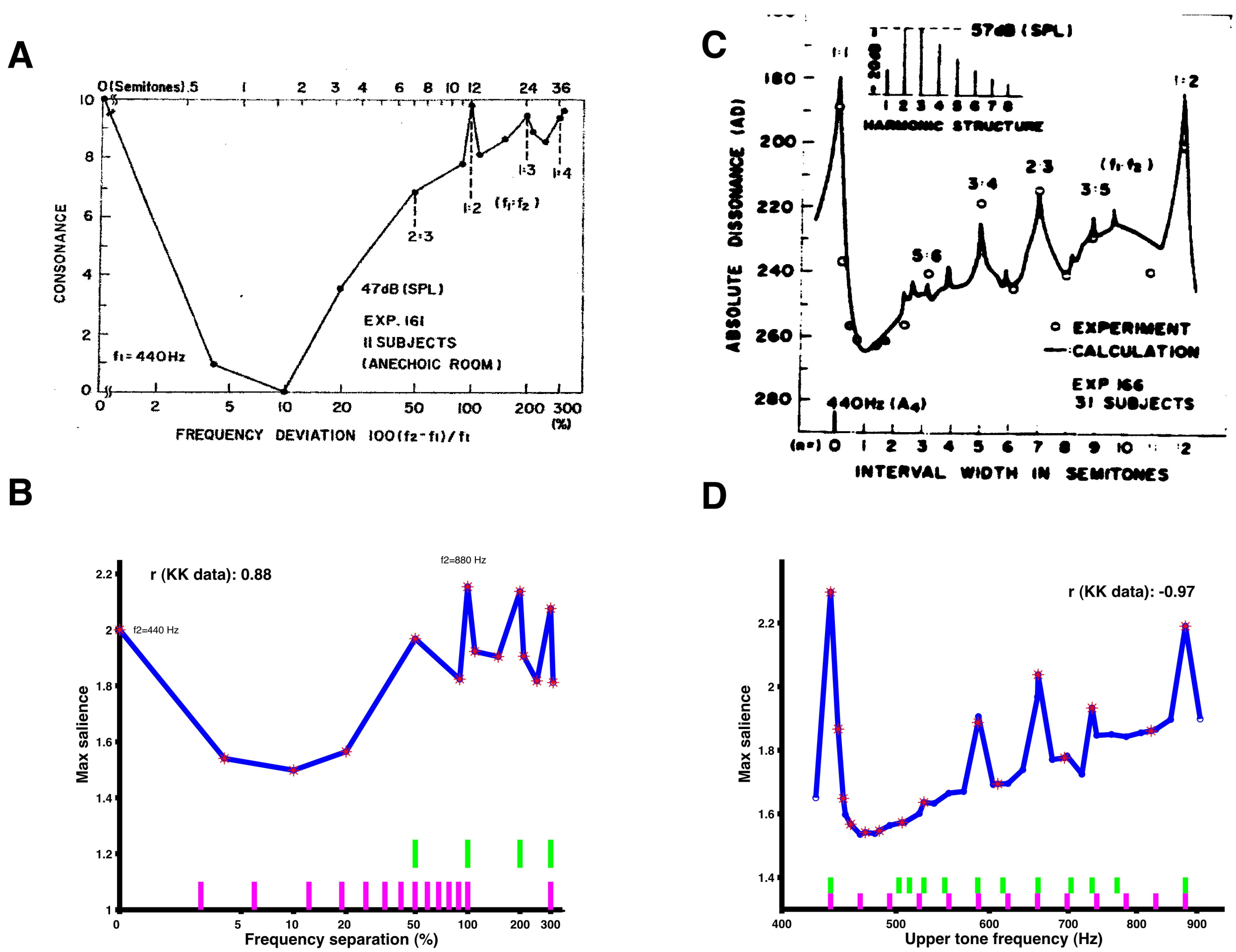

Figure 2. Comparisons of simulation-based consonance estimates with psychophysical observations. A. Subjective ratings of consonance ("clearness" vs. "tubidity") of pure tone dyads with different frequency separations in 11 human subjects, Fig. 5 in (Kameoka & Kuriyagawa, 1969a). B. Maximum pitch salience values for the same pure tone dyads computed from simulated auditory nerve population-interval distributions. Lower ticks: semitone intervals; upper ticks: just intervals (3:2, 2:1, 3:1, 4:1) C. Subjective ratings of absolute dissonance of complex tone dyads in 21 human subjects, Fig. 7 in (Kameoka & Kuriyagawa, 1969b). D. Maximum pitch salience values for complex tones (harmonics 1-6) computed from simulations as a function of the F0 of the upper tone (root F0 always 440 Hz).

account for the lower limit of musical pitch (~ 30 Hz, (Pressnitzer et al., 2001)), and the limited frequency resolution of the system. Discounting the influence of successively longer intervals achieves these effects. A neurocomputational interpretation would be that the interval analysis mechanism that subserves pitch has mostly only short time delays (< 33 ms) at its disposal. Simulations experiments with linear- and exponentially-tapering weights that decline with interval duration have suggested that an exponential decay with a time constant tau of ~9 ms best models the decline of pitch salience for periodicities below 100 Hz. The sharper the decay, the lower the saliences of periodicities < 100 Hz. The present model used CF-dependent taus: 100 Hz (.03s), 440 Hz (.016s), 880 Hz (.012s), 1320 Hz (.01s), and above (.009s).

Interval sieve

A dense set of periodic sieves (25-1000 Hz in 2 Hz steps) was applied to duration-weighted interval histograms in order to measure the relative pattern-strength of interval patterns associated with different perceived pitches. Sieves counted all bins that overlapped with the 100 us sieve holes. The salience of a particular pitch is estimated by dividing the mean density (intervals/bin) of pitch-related intervals in sieve-bins by the mean density of the whole distribution (background).

Maximum pitch salience

The pitch with the maximum salience value amongst all possible pitches is predicted to be the strongest pitch present and the one is most likely to be heard (cf. (Parncutt, 1989)).

The salience of this pitch produced by a given dyad or triad is one measure of the degree to which one pitch predominates over all other unrelated pitches. Harmonically-related pitches share intervals at their common subharmonics (nonexclusive allocation of intervals), so that their respective interval patterns interfere the least vis-à-vis the salience measure, while unrelated pitches reduce the salience of each other (because interval peaks in one raise the mean (background) density for the other). While the maximum pitch salience measure has its limitations, it captures many of the qualitative properties of "pitch fusion"(DeWitt & Crowder, 1987), "pitch unity", "tonal unity" (Lipps, 1905), fundamental bass, and virtual pitch strength (Parncutt, 1989).

4. RESULTS

Maximum saliences observed for pure and complex tone dyads are shown in Figure 2. For both pure and complex tones, maximal salience is highest for unison and the octave separations and lowest for separations near 1 semitone. In both cases, there was strong correspondence between psychophysical consonance ("clearness") and "absolute dissonance" ratings experimentally obtained by Kameoka and Kuriyagawa: r = 0.88 for consonances of pure tones and r = -0.97 for absolute dissonances of complex tones ((Kameoka & Kuriyagawa, 1969b), Fig. 7). In most cases, our population-interval model predictions followed those of K & K's roughness-based model for complex tones (C, solid lines, vs. D).

5. CONCLUSIONS

Representations based on all-order interspike intervals contain both overtone and undertone series, and in most respects their analysis parallels spectral operations based on coinciding harmonics and subharmonics (Terhardt et al., 1982). The present simulation demonstrates that interval-based models of low, "virtual" pitch can plausibly account for the consonance of pairs of pure and complex tones that are presented in isolation (without preceding tonal context). Since there exists a close relationship between frequency separations that give rise to maximal dissonance and critical band phenomena (Plomp & Levelt, 1965), these pitch fusion models may also predict critical bandwidths. It thus remains to be seen whether a neural basis for critical band phenomena might lie in the properties of population-wide interspike interval distributions of the auditory nerve (whose forms are partially determined by the properties of cochlear filters). Further studies will explore properties of population-interval distributions to model pitch multiplicities, stabilities, and similarities of triads along theoretical lines parallel to those of Parncutt (Parncutt, 1989) and Krumhansl (Krumhansl, 1990).

[This work was supported in part by NSF-EIA-BITS-013807]

6. REFERENCES

1. Bharucha, J. J. Tonality and expectation. In R. Aiello & J. Sloboda (Eds.), Musical Perceptions (pp. 213-239). New York: Oxford University Press, 1994.
2. Boomsliter, P., & Creel, W. "The long pattern hypothesis in harmony and hearing." J. Music Theory, 5, 2-31, 1962.
3. Cariani, P. "Temporal codes, timing nets, and music perception." J. New Music Res., 30(2), 107-136, 2002.
4. Cariani, P. A., & Delgutte, B. "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience." J. Neurophysiol., 76(3), 1698-1716, 1996a.
5. Cariani, P. A., & Delgutte, B. "Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch." J. Neurophysiol., 76(3), 1717-1734, 1996b.
6. Deutsch, D. The processing of pitch combinations. In D. Deutsch (Ed.), The Psychology of Music (pp. 349-411). San Diego: Academic Press, 1999.
7. DeWitt, L. A., & Crowder, R. G. "Tonal fusion of consonant musical intervals: The oomph in Stumpf." Perception and Psychophysics, 41(1), 73-84, 1987.
8. Geisler, C. D., & Greenberg, S. "A two-stage nonlinear cochlear model possesses automatic gain control." J Acoust Soc Am, 80(5), 1359-1363, 1986.
9. Green, B., & Butler, D. From acoustics to Tonpsychologie. In T. Christensen (Ed.), The Cambridge History of Western Music Theory (pp. 246-271). Cambridge: Cambridge University Press, 2002.
10. Hartmann, W. M. Signals, Sound, and Sensation. New York: Springer, 1998.
11. Helmholtz, H. v. On the Sensations of Tone as a Physiological Basis for the Theory of Music (1954 Reprint). New York: Dover, 1885.
12. Kameoka, A., & Kuriyagawa, M. "Consonance theory I. Consonance of dyads." J. Acoust. Soc. Am., 45, 1451-1459, 1969a.
13. Kameoka, A., & Kuriyagawa, M. "Consonance theory II. Consonance of complex tones and its calculation method." J. Acoust. Soc. Am., 45, 1460-1469, 1969b.
14. Kiang, N. Y. S., Watanabe, T., Thomas, E. C., & Clark, L. F. Discharge Patterns of Single Fibers in the Cat's Auditory Nerve. Cambridge: MIT Press, 1965.
15. Kim, D. O., & Molnar, C. E. "A population study of cochlear nerve fibers: comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones." J. Neurophysiol., 42(1), 16-30, 1979.
16. Krumhansl, C. L. Cognitive Foundations of Musical Pitch. New York: Oxford University Press, 1990.

17. Leman, M. "An auditory model of the role of short-term memory in probe-tone ratings." Music Perception, 17(4), 481-510, 2000.

18. Leman, M., & Carreras, F. Schema and Gestalt: Testing the hypothesis of psychoneural isomorphism by computer simulation. In M. Leman (Ed.), Music, Gestalt, and Computing (pp. 144-165). Berlin: Springer, 1997.

19. Licklider, J. C. R. "A duplex theory of pitch perception." Experientia, VII(4), 128-134, 1951.

20. Lipps, T. Consonance and Dissonance in Music (1995 translation) (W. Thompson, Trans.). San Marino, CA: Everett Books, 1905.

21. Meddis, R., & Hewitt, M. J. "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification." J. Acoust. Soc. Am., 89(6), 2866-2882, 1991.

22. Meddis, R., & O'Mard, L. "A unitary model of pitch perception." J. Acoust. Soc. Am., 102(3), 1811-1820, 1997.

23. Parncutt, R. Harmony: A Psychoacoustical Approach. Berlin: Springer-Verlag, 1989.

24. Patterson, R. D. "Spiral detection of periodicity and the spiral form of musical scales." Psychology of Music, 14, 44-61, 1986.

25. Plomp, R., & Levelt, W. J. M. "Tonal consonance and critical bandwidth." J. Acoust. Soc. Am., 38, 548-560, 1965.

26. Pressnitzer, D., Patterson, R. D., & Krumboltz, K. "The lower limit of melodic pitch." J. Acoust. Soc. Am., 109(5), 2074-2084, 2001.

27. Rose, J. E. Neural correlates of some psychoacoustical experiences. In D. McFadden (Ed.), Neural Mechanisms of Behavior (pp. 1-33). New York: Springer Verlag, 1980.

28. Rose, J. E., Hind, J. E., Brugge, J. R., & Anderson, D. J. "Some effects of stimulus intensity on response of single auditory nerve fibers of the squirrel monkey." Journal of Neurophysiology, 34(4), 685-699, 1971.

29. Schneider, A. "Verschmelzung", tonal fusion, and consoance: Carl Stumpf revisited. In M. Leman (Ed.), Music, Gestalt, and Computing. Berlin: Springer, 1997.

30. Sethares, W. Tuning, Timbre, Spectrum, Scale. London: Springer, 1999.

31. Slaney, M., & Lyon, R. F. (1991). Apple Hearing Demo Reel [videotape and report]. Cupertino, CA: Apple Computer, Inc.

32. Snyder, B. Music and Memory. Cambridge: MIT Press, 2000.

33. Terhardt, E. "On the perception of periodic sound fluctuations (roughness)." Acustica, 30, 201-213, 1974a.

34. Terhardt, E. "Pitch, consonance, and harmony." J. Acoust. Soc. Am., 55(May), 1061-1069, 1974b.

35. Terhardt, E., Stoll, G., & Seewann, M. "Pitch of complex signals according to virtual-pitch theory: test, examples, and predictions." J. Acoust. Soc. Am., 71(3), 671-678, 1982.

36. Tramo, M. J., Cariani, P. A., Delgutte, B., & Braida, L. D. Neurobiology of harmony perception. In I. Peretz & R. J. Zatorre (Eds.), The Cognitive Neuroscience of Music (pp. 127-151 (Reprint of Tramo et al (2001), Annals NY Acad Sci, 2990:2092-2116)). New York: Oxford University Press, 2001.