Entracking as a Brain Stem Code for Pitch: The Butte Hypothesis

Philip X Joris

Abstract The basic nature of pitch is much debated. A robust code for pitch exists in the auditory nerve in the form of an across-fiber pooled interspike interval (ISI) distribution, which resembles the stimulus autocorrelation. An unsolved question is how this representation can be "read out" by the brain. A new view is proposed in which a known brain-stem property plays a key role in the coding of periodicity, which I refer to as "entracking", a contraction of "entrained phase-locking". It is proposed that a scalar rather than vector code of periodicity exists by virtue of coincidence detectors that code the dominant ISI directly into spike rate through entracking. Perfect entracking means that a neuron fires one spike per stimulus-waveform repetition period, so that firing rate equals the repetition frequency. Key properties are invariance with SPL and generalization across stimuli. The main limitation in this code is the upper limit of firing (~ 500 Hz). It is proposed that entracking provides a periodicity tag which is superimposed on a tonotopic analysis: at low SPLs and fundamental frequencies > 500 Hz, a spectral or place mechanism codes for pitch. With increasing SPL the place code degrades but entracking improves and first occurs in neurons with low thresholds for the spectral components present. The prediction is that populations of entracking neurons, extended across characteristic frequency, form plateaus ("buttes") of firing rate tied to periodicity.

Keywords Autocorrelation · Brain stem · Phase-locking · Temporal

1   Introduction

> Pitch perception is a fascinating area because it appears so simple and yet is a process of considerable subtlety and complexity. (Green 1976)

P. X. Joris (✉)
Laboratory of Auditory Neurophysiology, University of Leuven, Herestraat 49 bus 1021,
3000 Leuven, Belgium
e-mail: Philip.Joris@med.kuleuven.be

> Pitch perception is considered to represent the heart of hearing theory, and is, without doubt, the topic most discussed over the years. (Plomp 2002)

> Pitch may be the most important perceptual feature of sound. (Yost 2009)

Despite more than a century of study, there is no consensus regarding the basic nature of pitch, causing a palpable level of frustration among hearing researchers. The wealth of behavioral research on pitch perception contrasts with the paucity of physiological research into its neural basis beyond the level of the auditory nerve (AN). Processing in the central nervous system (CNS) is expected to be fundamentally different for the various temporal vs. spectral schemes that have been proposed, so physiological insights have the potential to reveal which (combination) of the two classes of schemes underlies human pitch perception.

A robust but implicit code for pitch exists in the AN in the form of an across-fiber pooled interspike interval (ISI) distribution (Cariani and Delgutte 1996a, 1996b; Meddis and Hewitt 1991a, b), which resembles the stimulus autocorrelation. An unsolved question is how and whether this implicit temporal representation is transformed into a more explicit representation. Following an early model (Licklider 1951), various autocorrelation-type schemes have been proposed, in which periodicity-tuning is generated by some combination of coincidence detection and a source of delay. It is generally assumed that such a computation is implemented at a brainstem level, where responses are temporally precise over a broad range of frequencies. However, recordings have not revealed convincing evidence for level-invariant, periodicity-tuned neurons or sources of delay that cover a sufficiently broad temporal range (Neuert et al. 2005; Sayles et al. 2013; Sayles and Winter 2008a, 2008b; Verhey and Winter 2006; Wang and Delgutte 2012).

Here, simple properties of the early central auditory system are brought into focus and it is argued that the relevant representation is fundamentally different from autocorrelation-type schemes *à la* Licklider (1951). The key proposal is that entrained phase-locking (contracted to "entracking") generates a scalar rate code for pitch early in the brainstem.

2   AN to Brain Stem: A Change in Orientation

In the AN, ISI histograms of responses to low-frequency pure tones are always multimodal (Kiang et al. 1965); i.e., after firing a spike, fibers often skip one or more cycles before firing another spike. Cycle skipping allows temporal and average rate behavior to be uncoupled. For example, neither the sigmoidal shape of rate-level functions (firing rate as a function of stimulus SPL), their dynamic range, nor maximum firing rate, are dependent on stimulus frequency *per se* but on stimulus frequency *relative* to characteristic frequency (CF). This behavior, combined with cochlear band-pass filtering, underlies rate-place coding: spectral components are translated into a firing rate profile of the population of AN fibers (Cedolin and Delgutte 2005; Larsen et al. 2008). This is a tonotopic or "vertical" view of the audi-

tory system in which strength of response along the tonotopic axis is proportional to spectral energy. Experimental studies in which large populations of AN fibers are studied support this view (Delgutte and Kiang 1984; Kim and Molnar 1979; Sachs and Young 1979). Frequency *in an absolute sense* is inconsequential in these displays. For example, similar rate-place profiles would be expected for an animal with low-frequency hearing and one with high-frequency hearing, if the stimuli and filter widths and shapes could be appropriately scaled between the two species.

In contrast, in the CNS, skipping of cycles is often less prominent. At low stimulus frequencies, higher modes in the ISI distribution, at multiples of the stimulus period, can be strongly attenuated relative to the mode corresponding to the tone period (references: see Sect. 3). One corollary of this behavior is that average firing rate can approach the stimulus frequency. Perfect entracking means that a neuron fires one spike per cycle so that the firing rate equals the stimulus repetition frequency. A rate-place profile for neurons with perfect entracking would look quite different from that in the AN, and stimulus frequency *in an absolute sense* would now affect the display. For a low-frequency pure tone at some supra-threshold level, the rate-place profile would show a horizontal rather than a vertical pattern: all entracking neurons would be firing at the same firing rate, which would equal the stimulus frequency. Of course, only neurons with CF sufficiently close to the stimulus frequency would receive enough drive from their inputs to show entracking. The expected output is therefore a mixture of the vertical and horizontal pattern: a *butte* of activity in which all neurons would have the same average firing rate, and whose edges are formed by neurons whose CF is too far removed from the stimulus frequency for full entracking. A mixture of two low-frequency tones, if sufficiently separated in frequency, would be expected to generate two buttes with different height, corresponding to the two frequencies.

3   Entracking to Pure Tones

Pure tones are rare in nature but relevant in the context of pitch, not only to define and measure pitch, but also because of the dominant role of resolved harmonics (Plack and Oxenham 2005). Entracking to pure tones is visible in some of the early studies of the brainstem (Moushegian et al. 1967; Rose et al. 1974), and is extensively documented in studies from the Madison group (Joris et al. 1994b; Recio-Spinoso 2012; Rhode and Smith 1986). These data show that near-perfect entracking can be observed in low-CF neurons, or in the "low-frequency tail" of neurons with higher CFs. As expected from refractory behavior, there is an upper limit, which varies across neurons and across species. In rare instances neurons will fully entrack at 800 Hz or even higher, but more often the upper limit is near 500 Hz or lower. Over the frequency range of entracking, average firing rate increases linearly with frequency, and depends little on SPL except at low suprathreshold SPLs. One consequence is that neurons may fire at much higher rates to low-frequency stimuli in their "tail" than to tones near their CF (see Fig. 13 in Joris et al. 1994a).

Entracking is not a rare phenomenon. We and others have observed it in various cell types and nuclei, and in a number of species. In the ventral CN of the cat, most types of projection neuron display this behavior to some degree (spherical and globular bushy cells, octopus cells, commissural multipolar cells, stellate cells). We have observed it in the medial nucleus of the trapezoid body (Mc Laughlin et al. 2008) and in other neurons of the superior olivary complex (e.g. Joris and Smith 2008). We have observed entracking in different species (cat, chinchilla, gerbil; see also studies cited above), and have limited evidence in the CN of macaque monkey.

An important qualifier is that the entracking observed is not always the extreme form (exactly 1 spike/cycle), particularly at frequencies above a few 100 Hz. The defining property is that there is an effect of absolute frequency on average firing rate so that rate increases monotonically with frequency up to a certain maximum. Thus, in the brainstem, firing rate does not only depend on SPL and stimulus frequency relative to CF, as it does in the AN, but also depends on absolute stimulus frequency. For the extreme cases of this behavior a stronger statement can be made: SPL and stimulus frequency relative to CF have remarkably little effect, and absolute frequency is the overriding stimulus parameter determining response rate.

Besides being present in various cell types, nuclei, and species, entracking has inherent properties that make it an attractive coding mechanism. It is remarkably invariant with SPL, i.e. once perfect entracking is reached, further increases in SPL do not affect average response rate: the rate-level function shows a limited (20 dB) dynamic range, and at higher SPLs the firing rate remains clamped at the stimulus frequency. At the population level, increases in SPL cause an increase in the number of entracking neurons. A second striking property is the low variability in firing rate. In some cases there is no variability: exactly the same number of spikes is generated in response to the same stimulus presented at the same or other SPLs.

4 Yes We SAM

Data regarding entracking to pitch stimuli other than pure tones are limited. An early striking example is to click trains: octopus cells in the CN can fire one spike per click up to ~700 clicks/s (Godfrey et al. 1975; Oertel et al. 2000). This behavior occurs at high CFs, to which octopus neurons are biased. As mentioned, octopus cells that can be driven by low-frequency tones also show entracking to tones. In AN fibers, firing rate also shows some dependence on click train frequency, but much weaker than in octopus cells.

We have also observed entracking to sinusoidally amplitude modulated (SAM) tones. Figure 3G in Joris and Smith (2008) shows data at different SPLs over a range of modulation frequencies for one monaural neuron, recorded just laterally to the MSO, in an area which may correspond to the mLNTB (Spirou and Berrebi 1996). Similar to entracking to pure tones mentioned in the previous section, firing rate increased linearly with modulation frequency; was similar for the different SPLs tested; and declined steeply once an upper limit is reached. The CF of the

neuron was 2.4 kHz, but the response to pure tones at CF was much lower than to low-frequency tones delivered in the tail of the tuning curve, to which the neuron entracked (Fig. 3C in Joris and Smith 2008).

5   Scalar *versus* Vector Code

Various schemes and neural mechanisms have been proposed for the encoding of pitch-related periodicity based on the temporal information carried by peripheral neurons. They have in common that they predict tuning in which a neuron is optimally responsive to a certain periodicity. Sounds that differ in pitch would activate different neurons; different sounds with the same pitch would activate the same neurons. Such schemes are referred to as vector codes (Churchland and Sejnowski 1992). Periodicity tuning is typically achieved by comparing spike trains with a delayed copy. The comparison involves multiplicative, subtractive, or additive neural interactions; and some source of delay from axons, cochlea, synapses, intrinsic membrane properties, etc. (reviewed by de Cheveigné 2005). From a physiological point of view, problems with this approach are that there is only limited evidence for such tuning in the brainstem and only over a limited range of periodicities; that no convincing delay mechanisms have been identified that cover a wide range of values (tens to tenths of ms); and that some of these schemes require elaborate (and biologically implausible) wiring.

The scheme proposed here dispenses with delays and suggests a scalar code, based on a property which is physiologically well documented. We surmise that entracking neurons at the brain stem level are not tuned to a particular periodicity but all code for a range of periodicities by their monotonic relationship between average firing rate and pitch-related period.

6   Buttes

To some extent (increasingly so with increasing SPL), the representation hypothesized here is "orthogonal" to the tonotopic representation. The rate-place profile in the AN to a low-frequency tone increasingly broadens with increasing SPL, and flattens due to rate saturation (Kim and Molnar 1979). If a population of central neurons shows entracking, this property causes a stratification in the rate-place profile. Instead of a vertical or "hilly" profile where resolved components cause local increases in firing rate and unresolved components lead to a broad mound of activity, entracking generates horizontally flattened profiles or buttes, for which firing rate is dictated by the dominant interval between spikes fed to neurons in these frequency channels. For multiple, truly resolved components, a staircase of buttes would result with increasing firing rates corresponding to the frequency of successive harmonics. For unresolved components, many factors come into play (filter

shape, limits to fine-structure and envelope, component phase), but buttes could be formed by neurons entracking to the dominant stimulus period.

An obvious limitation in this scalar code is the upper limit of firing (usually near ~500 Hz). We surmise that at low SPLs and fundamental frequencies >500 Hz, a spectral or place mechanism codes for pitch (Cedolin and Delgutte 2005; Larsen et al. 2008). With increasing SPL the place code degrades but entracking improves and first occurs in the neurons with the lowest thresholds for the spectral components present.

For a given stimulus and time window, a neuron can only have one firing rate. One may wonder what distinguishes the low firing rate of an entracking neuron to a low frequency component vs. the low firing rate to a component above the entracking limit. The key would again be in the uniform firing rate, with low variability, across neurons in the case of a low frequency component. Above the entracking limit, firing rate is no longer clamped to the value dictated by the dominant ISI of a neuron's inputs and will vary across neurons.

7   Discussion

The main goal of this chapter is to introduce a new way of thinking about pitch coding, grounded in CNS physiology. If there is a robust representation of pitch in the dominant ISI distribution in the AN (Cariani and Delgutte 1996a, 1996b), and if some neurons convert ISI directly into a corresponding firing rate, then it seems possible that the dominant ISI interval is coded as predominant firing rate. Evidently, the scheme proposed here is incomplete. Questions arise how and where a butte profile would be read out; how such a representation would mesh with the spectral representation needed for F0 above ~500 Hz; how phase-invariant this representation would be; etc. We conclude with some interrelated issues.

An important issue is the effective bandwidth of central neurons. Several CN neuron types integrate over wide frequency regions (Godfrey et al. 1975; Winter and Palmer 1995): partials that are resolved at the level of the AN may be unresolved in these CN populations. The autocorrelogram-like display in the dominant ISI hypothesis sums across frequency channels (Cariani and Delgutte 1996a, 1996b): for such an operation a wide bandwidth would be beneficial. Another issue is whether there is a specific physiological subset of neurons or even a separate brainstem nucleus which codes periodicity via entracking. Entracking is observed in a diversity of structures and neuron types, suggesting a distributed mechanism, but this does not exclude the existence of a brainstem "pitch center" specialized in this form of encoding. For CN neurons showing entracking, convergence of multiple inputs from the AN is obviously required; the degree of entracking in responses beyond the CN suggests that there are multiple stages of such convergence. A strong form of the butte hypothesis is based on perfect entracking; a weaker form only requires a monotonic relationship between firing rate and pitch-related period without attaining equality. One of the most critical issues is phase invariance, which we see

as an experimental issue. Some neurons in the CN show good envelope coding to quasi-frequency-modulated (QFM) stimuli (Rhode 1995). Possibly there are always some neurons entracking at the pitch-related period, no matter what the phase spectrum is. Finally, entracking is invariably accompanied by exquisite phase-locking. One could debate whether it fundamentally is a temporal or rate code. The temporal aspects of the response are disregarded here in the sense that it is the constancy in rate, within and across neurons and stimuli, that codes for pitch, while the phase of spiking is surmised to be irrelevant.

References

Cariani P, Delgutte B (1996a) Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. J Neurophysiol 76:1717–1734

Cariani P, Delgutte B (1996b) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J Neurophysiol 76:1698–1716

Cedolin L, Delgutte B (2005) Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve. J Neurophysiol 94(1):347–362

Churchland P, Sejnowski TJ (1992) The computational brain. MIT Press, Cambridge

De Cheveigné A (2005). Pitch perception models. In: Plack CJ, Fay RR, Oxenham AJ, Popper AN (eds) Pitch. Springer New York, New York, pp 169–233

Delgutte B, Kiang NYS (1984) Speech coding in the auditory nerve: i. Vowel-like sounds. J Acoust Soc Am 75:866–878

Godfrey DA, Kiang NYS, Norris BE (1975) Single unit activity in the posteroventral cochlear nucleus of the cat. J Comp Neurol 162:247–268

Green D (1976) An introduction to hearing. Lawrence Erlbaum Associates, Hillsdale

Joris PX, Smith PH (2008) The volley theory and the spherical cell puzzle. Neuroscience 154(1):65–76

Joris PX, Smith PH, Yin TC (1994a) Enhancement of neural synchronization in the anteroventral cochlear nucleus. II. Responses in the tuning curve tail. J Neurophysiol 71(3):1037–1051

Joris PX, Carney LH, Smith PH, Yin TC (1994b) Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. J Neurophysiol 71(3):1022–1036

Kiang NYS, Watanabe T, Thomas EC, Clark LF (1965) Discharge patterns of single fibers in the cat's auditory nerve. MIT Press, Cambridge. Research Monograph No 35

Kim DO, Molnar CE (1979) 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 42:16–30

Larsen E, Cedolin L, Delgutte B (2008) Pitch representations in the auditory nerve: two concurrent complex tones. J Neurophysiol 100(3):1301–1319

Licklider JCR (1951) A duplex theory of pitch perception. Experientia 7(4):128–134

Mc Laughlin M, van der Heijden M, Joris PX (2008) How secure is in vivo synaptic transmission at the calyx of Held? J Neurosci 28(41):10206–10219

Meddis R, Hewitt MJ (1991a) Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: phase sensitivity. J Acoust Soc Am 89(6):2883–2894

Meddis R, Hewitt MJ (1991b) Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: pitch identification. J Acoust Soc Am 89(6):2866–2882

Moushegian G, Rupert AL, Langford TL (1967) Stimulus coding by medial superior olivary neurons. J Neurophysiol 30(5):1239–1261

Neuert V, Verhey JL, Winter IM (2005) Temporal representation of the delay of iterated rippled noise in the dorsal cochlear nucleus. J Neurophysiol 93(5):2766–2776

Oertel D, Bal R, Gardner SM, Smith PH, Joris PX (2000) Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus. Proc Natl Acad Sci USA 97(22):11773–11779

Plack CJ, Oxenham AJ (2005) The psychophysics of pitch. In: Plack CJ, Oxenham AJ, Fay RR, Popper AN (eds) Pitch: neural coding and perception, vol 24. Spinger, New York, pp 7–55

Plomp R (2002) The intelligent ear: on the nature of sound perception. Lawrence Erlbaum Associates, Mahwah

Recio-Spinoso A (2012) Enhancement and distortion in the temporal representation of sounds in the ventral cochlear nucleus of chinchillas and cats. PloS ONE 7(9):e44286

Rhode WS (1995) Interspike intervals as a correlate of periodicity pitch in cat cochlear nucleus. J Acoust Soc Am 97(4):2414–2429

Rhode WS, Smith PH (1986) Encoding timing and intensity in the ventral cochlear nucleus of the cat. J Neurophysiol 56:261–286

Rose JE, Kitzes LM, Gibson MM, Hind JE (1974) Observations on phase-sensitive neurons of anteroventral cochlear nucleus of the cat: nonlinearity of cochlear output. J Neurophysiol 37:218–253

Sachs MB, Young ED (1979) Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. J Acoust Soc Am 66(2):470–479

Sayles M, Winter IM (2008a) Ambiguous pitch and the temporal representation of inharmonic iterated rippled noise in the ventral cochlear nucleus. J Neurosci 28(46):11925–11938

Sayles M, Winter IM (2008b) Reverberation challenges the temporal representation of the pitch of complex sounds. Neuron 58(5):789–801

Sayles M, Füllgrabe C, Winter IM (2013) Neurometric amplitude-modulation detection threshold in the guinea-pig ventral cochlear nucleus. J Physiol 591(Pt 13):3401–3419

Spirou GA, Berrebi AS (1996) Organization of ventrolateral periolivary cells of the cat superior olive as revealed by PEP-19 immunocytochemistry and Nissl stain. J Comp Neurol 368(1):100–120

Verhey JL, Winter IM (2006) The temporal representation of the delay of iterated rippled noise with positive or negative gain by chopper units in the cochlear nucleus. Hear Res 216–217:43–51

Wang GI, Delgutte B (2012) Sensitivity of cochlear nucleus neurons to spatio-temporal changes in auditory nerve activity. J Neurophysiol 108(12):3172–3195

Winter IM, Palmer AR (1995) Level dependence of cochlear nucleus onset unit responses and facilitation by second tones or broadband noise. J Neurophysiol 73:141–159

Yost WA (2009) Pitch perception. Atten Percept Psychophys 71(8):1701–1715