CORTICAL REPRESENTATION OF THE ACOUSTIC SPECTRUM

Shihab A. Shamma
Electrical Engineering Department and Institute for Systems Research
University of Maryland, College Park, MD 20742 USA

ABSTRACT

Acoustic signals are characterized by their timbre, pitch, loudness, forms of modulation, and onset/offset instants. These descriptions of sound quality have a close relationship to the instantaneous spectral properties of the sound waves. The auditory system has developed elegant mechanisms to extract and represent this spectro-temporal information through noise-robust perceptual features. At the level of the auditory cortex, these processes are manifested by an elaborate multidimensional representation of the shape of the dynamic acoustic spectrum. Specifically, at each frequency, the local shape of the spectrum is decomposed in terms of its bandwidth and asymmetry. Such a representation turns out roughly to correspond to a *local* cepstral-like representation of the spectrum, or more accurately, a wavelet transform of the acoustic spectral profile. Mathematical descriptions of this representation have become feasible and functionally relevant, and can be fruitfully used to derive the principles underlying time-frequency analysis in the auditory system. In turn, these principles can be applied in various contexts involving detection, analysis, synthesis, and recognition of sound.

INTRODUCTION

The spectral profile and its evolution in time play a key role in the perception of timbre of broad band sounds such as speech and music [1]. It is therefore important to understand how and which features of a spectral profile are extracted and encoded by the central auditory system. In this paper, we review first the fundamental response properties of neurons in the primary auditory cortex (AI), the last processing stage along the primary auditory pathway. Next, we discuss the implications of these findings to the representation of stationary and dynamic speech spectra such as those of a sustained vowel and the transitions in a CV syllable. Specifically, we shall demonstrate that the shape of the acoustic spectrum is represented along at least three different axes: the usual frequency axis, a local bandwidth (or scale) axis, and a local asymmetry axis. For dynamic spectra, the latter two axes additionally represent the speed and direction of formant transitions.

AI is strictly tonotopically organized because of the topographic order of neural projections from the cochlea through several stages of processing (Fig.1). Thus, when tested with single tones, AI neurons are selective to a range of frequencies around a best frequency (BF) [2]. Within this range, responses change from excitatory to inhibitory in a pattern that varies from one cell to another in its width and asymmetry around the BF (Fig.2); This response pattern is usually called the *response area or field* (RF) of the neuron [3]. When a broad band spectrum is used as a stimulus, the cell's response can be thought of as the net effect of all excitatory and inhibitory influences induced by the spectral region within its RF. However, despite the diversity of RF shapes and the complexity of their responses, two simple organizational principles underlie the way in which AI responses encode the shape of the acoustic spectrum. These are *linearity* and *selectivity* of AI responses.

LINEARITY OF AI RESPONSES

To first order, AI responses to broad band spectra are linear in the sense that they satisfy the *superposition* principle [4]. This is illustrated in Fig.3 as follows: Given the response patterns R_A and R_B evoked along the tonotopic axis by each of the stimulus spectra S_A and S_B, then the response pattern due to the sum of the two spectral profiles, $S_A + S_B$, is, to within a gain factor, the sum of the responses, i.e., $R_A + R_B$. This rather surprising finding is demonstrated experimentally in Fig.4, where single unit responses to different

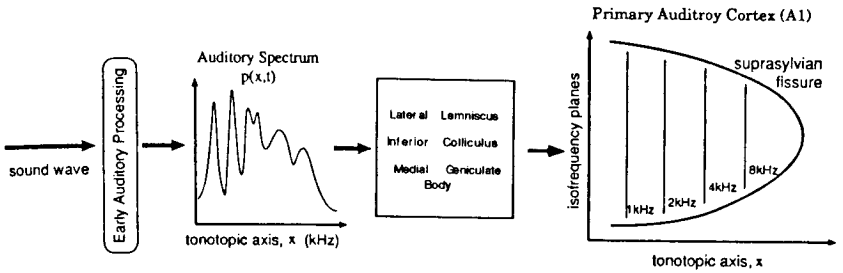


Figure 1. Schematic of the auditory pathway to the primary auditory cortex (AI) of the ferret. AI is tonotopically organizedwith units of similar BFs forming isofrequency planes as indicated.

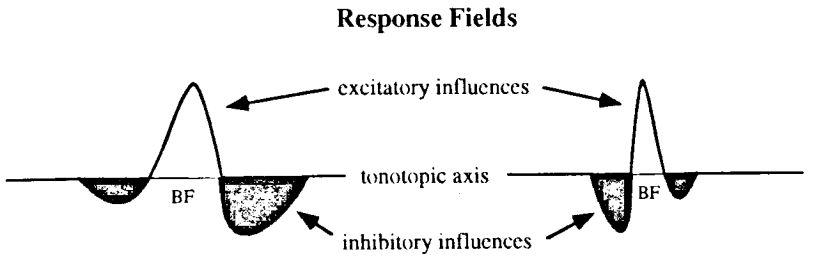


Figure 2. Schematics of two different tesponse fields (RF) measured in AI units. RFs may have different asymmetries with inhibition more prominant above the BF (left), below the BF (right), or simply symmetric. They also range in bandwidths from broad (left) to narrow (right).

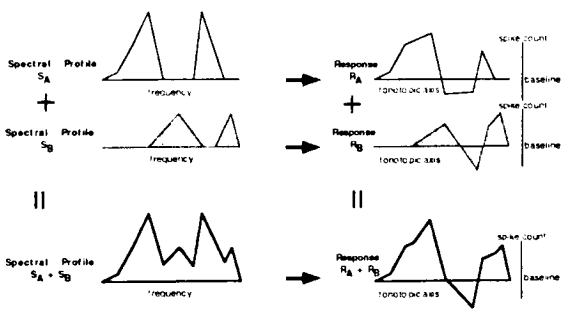


Figure 3. Linearity of AI responses imply the superposition principle. Acoustic spectral profiles S_A, S_B, and $S_A + S_B$ evoke schematic response patterns R_A, R_B, and R_A+R_B along the tonotopic axis. The response patterns are measured as spike counts relative to a "baseline" which is the response to a flat spectrum.

sinusoidal spectral profiles (also called *ripples*) are combined and compared with responses to superimposed ripple spectra. For instance, Fig.4A displays the responses to four spectra with different ripple densities (0.4, 0.8, 1.2, and 1.6 cycles/octave). In each case, the *dashed* response curve is constructed from the spike counts of the cell as the spectral profile is shifted to the left relative to the BF. As is evident, the responses track the sinusoidal shape of the input spectrum; They are largest when the ripple rate is 0.8 cycle/octave, and weakest at 1.6 cycles/octave. The *solid* curves are the best mean-square fits to the data.

When a complex spectrum is formed by the superposition of two ripples (Fig.4B), e.g., 0.4+0.8 cycles/octave (top) or 0.8+1.6 cycles/octave (bottom), linearity predicts that the response curves should resemble the superposition of the responses to the individual ripple spectra. This is confirmed by the similarity of the measured (*dashed*) and predicted (*solid*)) response curves in both cases. These results have been confirmed in a large number of tests involving spectral profiles composed of up to 10 superimposed spectra [5,6].

Linearity is a powerful simplifying principle that allows one to predict the responses to any arbitrary spectral profile. Specifically, if the responses to the basic set of rippled spectra are known, then it is possible to superimpose them uniquely to generate the responses to any arbitrary profile (this is the so-called *Fourier* decomposition) [4]. Therefore in the remainder of this paper, we shall examine in more detail the response properties of AI cells to various spectral ripple parameters.

SELECTIVITY OF AI RESPONSES

AI units are generally selective in that they respond only within a limited range of values of a given stimulus parameter. For instance, units are usually tuned along the tonotopic axis, i.e., they are driven by a relatively narrow range of frequencies around a BF as described earlier (Fig.2). AI responses are also selective to the parameters of a ripple spectrum, specifically the ripple frequency (or density, Ω) and ripple phase (Φ). For instance, in Fig.4A, the unit responds best around the ripple frequency 0.8 cycles/octave. Furthermore, the responses vary with the phase of the ripple, being excited in one-half cycle while suppressed in the other. This ripple selectivity can be efficiently displayed by a *transfer function* $T(\Omega)$ (Fig.5), where the amplitude and phase of the responses to different ripples are plotted as a function of ripple frequency. A complementary view of this information is contained in the unit RF which is (conceptually) formed by summing up the responses to the different ripples, or more accurately by *inverse Fourier* transforming $T(\Omega)$ [4].

Selectivity of an AI unit around a characteristic ripple frequency (Ωo) is intuitively inversely related to the width of its RF, or roughly to the bandwidth of the unit's frequency tuning curve [5]. Thus, the higher Ωo is, the more narrowly tuned the RF is. This suggests that a unit responds best (or is selective) to spectral patterns with a local bandwidth (or *scale*) that is comparable to that of its RF.

Similarly, selectivity to a particular ripple phase (characteristic phase, Φ_O) is directly reflected in the asymmetry of the RF. For a unit with Φ_O near zero, the RF exhibits a central excitatory region around the BF, flanked by symmetric inhibitory areas. If the Φ_O is positive (negative), the inhibition becomes asymmetrically strong below (above) the BF [3,5,7]. In this manner, an AI unit is selective to the local slope or asymmetry of the input spectral profile around the BF.

AI REPRESENTATION OF A VOWEL SPECTRUM

The combined selectivity of an AI unit to the asymmetry and scale around a local spectral region (BF) of the input profile means that it can encode explicitly the local shape of the spectrum. For example, the asymmetry of the RF in Fig.5 is directly responsible for the unit's selective responses (Fig.6) to the 2nd formant of the vowel spectrum /aa/, and not to the 1st formant. By having RFs with a range of BFs, bandwidths, and asymmetries, the AI can represent the shape of the entire input spectrum along three different axes. Such a representation is demonstrated in Fig.7 for the spectral profile of the vowel /aa/

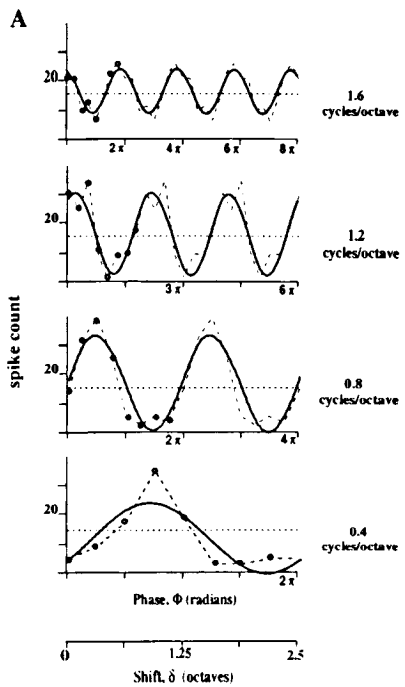


Figure 4. Superposition of responses to rippled spectra. (A) Responses of a unit to four input profiles at different ripple frequencies. In each case, the spectrum was shifted downwards relative to the BF of the unit to obtain the responses over the full cycle. (B) Comparison of recorded and predicted responses to spectra composed of ripple combinations 0.4+0.8 (top) and 0.8+1.6 (bottom) cycles/octave.

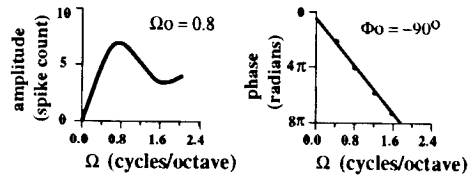
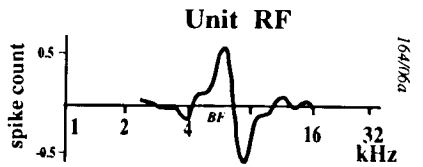


Figure 5. Ripple transfer function ($T(\Omega)$) of an AI unit. . The magnitude (left) and phase (right) of $T(\Omega)$ can be inverse Fourier transformed to obtain the RF (bottom).



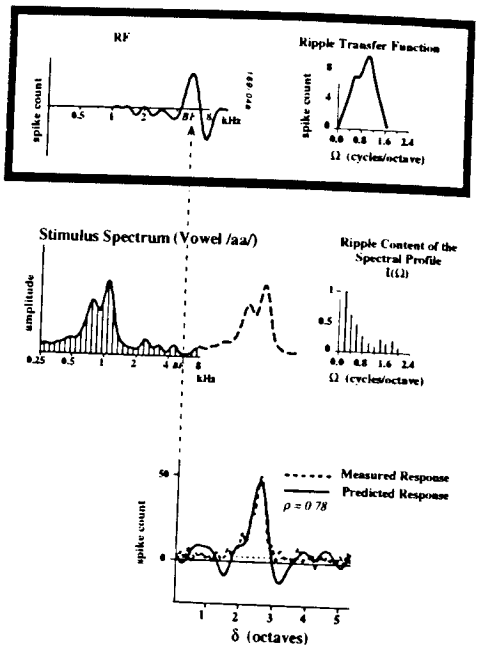


Figure 6. Measured and predicted single unit responses to the spectrum of a vowel /aa/. The unit's RF and T(Ω) are shown in the top box. The vowel spectrum (middle left) is synthesized by adding 10 ripples with amplitudes indicated by I(Ω) (middle right). Note, the unit responds well to the 2nd formant of the vowel (bottom); It does not respond to the 1st formant because of its asymmetric RF.

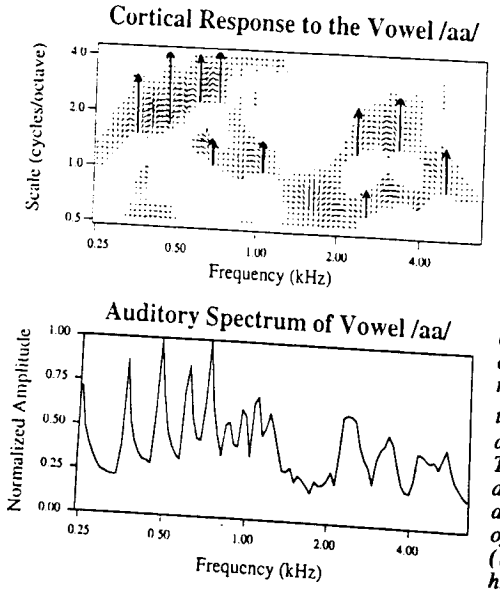


Figure 7. The spectrum of a naturally spoken vowel /aa/ (bottom), and the corresponding cortical representation (top). The scale axis is labeled by the characteristic ripple Ωo of the cortical cells. The characteristic phase is represented by the direction of the arrows (0 to 2π in clockwise direction, starting at up-arrow). The strength of the response is denoted by the length of the arrows. For clarity, trajectory of activated symmetric RFs (up-arrows) have been highlighted.

[7]. It is evident that the shape of the profile is dominated by the spectral peaks, i.e., the overall formant structure and the underlying harmonicity in the low frequency region (usually <1 kHz). These features are explicitly analyzed in the cortical representation in terms of their local asymmetry and bandwidth.

For example, the fine structure of the spectral harmonics is visible at the higher scales (usually > 1.5-2 cycles/octave). In contrast, the formants are relatively broad in bandwidth and thus are represented by the activity of units tuned to lower Ωo (< 2 cycles/octave). Sometimes, closely spaced formants are represented simultaneously at multiple scales as in the region of the 3rd formant (around 2.5 kHz), which is represented by activity near Ωo = 0.5 and 2 cycles/octave. The higher Ωo corresponds to the 3rd formant peak (approximately 0.25 octave in width). The lower Ωo captures the broad and skewed distribution of energy due to the combined 3rd and 4th formant peaks. A similar ``double-scale" representation occurs near 600-700 Hz, where the fine harmonic structure is represented at the higher scales, while the format structure (evident in the envelope of the harmonic peaks) is captured at lower Ωo.

The local asymmetry of the pattern in this representation is encoded by the direction of the arrow of the response. It provides a description of the local energy distribution in the spectrum. For example, the tonotopic locations at which the spectrum is locally symmetric (and hence represented by the up-arrows) closely reflect the positions of the peaks in the auditory spectrum; The left and right-arrows indicate whether the nearest spectral peak is at a higher or lower frequency. For instance, the spectral peak of vowel /aa/ at 3.25 kHz is not resolved at the broad scale, i.e., there is no up-arrow at Ωo =1 at this frequency. Instead, it is regarded as a trough (down-arrows) because it is flanked by two stronger peaks. However, the peak and its surrounding narrow valleys are resolved at a higher scale corresponding to twice the Ωo (around 2 cycles/octave).

AI RESPONSES TO DYNAMIC SPECTRA

Remarkably, AI units exhibit the same response properties of linearity and selectivity to *dynamic* spectral profiles. Thus, responses evoked by any combination of dynamic inputs can be roughly predicted from a linear sum of responses to the individual inputs. This is demonstrated in Fig.8 using a rippled spectrum (Ω = 0.8 cycles/octave) that moves to the left along the tonotopic axis with different *angular* velocities ω (4-24 cycles/sec). As is evident in Figs.8A and B, these stimuli evoke in this unit well synchronized responses to all ripple velocities. Combining ripples with different ω and Ω (Fig.8C) produce responses that are predictable by superposition of responses to the individual moving ripples. Again, the significance of linearity and of the basic set of moving ripple stimuli is seen through the *Fourier* decomposition theorem which allows us to generate and predict the responses to arbitrarily complex dynamic spectra, such as those of speech CV-syllables.

As with stationary ripples, responses to moving ripples are selective in that a given unit responds over a restricted range of velocities around a characteristic rate, ω_o [8,9]. Furthermore, there does not seem to be a relationship between ω_o and Ωo in a given unit, i.e., in a large population of AI units in the ferret, all combinations approximately within Ωo < 2 cycles/octave and ω_o < 20 cycles/sec may occur. It is likely that these ranges vary significantly across species reflecting their acoustic environment.

One possible implication of the selectivity to ω is the ability to encode the *rate* of spectral transitions. In addition, AI units are readily selective to the direction of a spectral transition by virtue of their RF asymmetries [7]. Combining those two features, together with those of bandwidth and BF creates a multidimensional cortical representation which explicitly extracts and maps out a variety of stationary and dynamic measures of the shape of the acoustic spectrum [7].

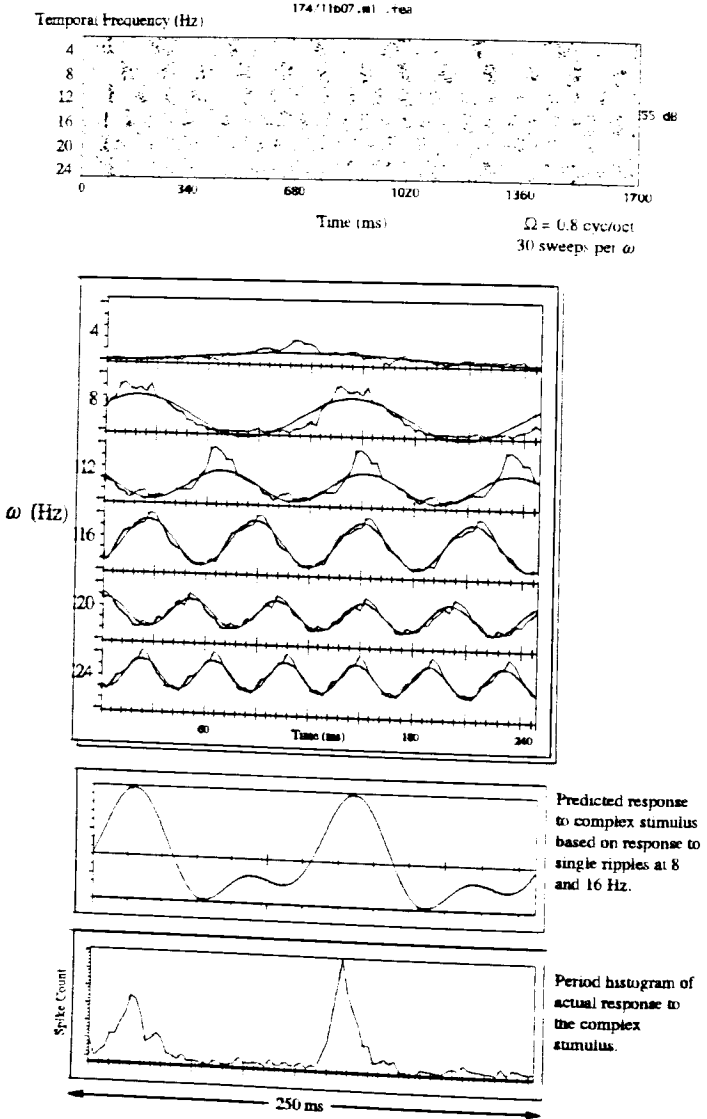


Figure 8. Responses to rippled spectra. (A) A 0.8 cycle/octave ripple moving at increasing angular velocities ω (4-24 cycles/sec) evokes synchronized responses. The ripple begins to move at time 0 ms, and is turned on at time 50 ms. Following an onset response, the unit synchronizes to the stimulus periodicity as seen in the period histograms below. (B) Period histograms of the responses and their mean-square error sinusoidal fits. (C) Measured (bottom) and predicted (top) responses for a two 0.8 cycle/oct ripples travelling at different velocities ω (8 and 16 cycles/sec). The top plot is constructed from the sum of the sinusoidal fits at 8 and 16 Hz shown in (B) above. The bottom plot is a 250 ms period histogram of the responses.

REPRESENTATION OF PITCH

So far, we have focused exclusively on the representation of spectral profiles. Another important percept that can be integrated in this representation in various ways is that of pitch. It is well established that the resolved harmonics of voiced sounds (such as those of the vowel /aa/ in Fig.7) contribute significantly to the perception of pitch [1]. On the logarithmic frequency tonotopic axis of the auditory system, resolved harmonics fit within a typical pattern which, except for a shift along this axis, remains unchanged regardless of pitch value. The harmonics in turn create a similarly stable cortical activation pattern at the higher scales from which pitch values and strength can be determined [7]. Other temporal mechanisms for the encoding of pitch can also be theoretically integrated in the cortical representation if the appropriate ranges of ω temporal selectivities are used, e.g., in the manner already suggested by [10] at lower auditory centers.

DISCUSSION

The auditory cortical representation of the acoustic spectrum is evidently quite redundant in that it expands the profile along several additional axes (e.g., ripple scale and phase, and temporal rate). This redundancy potentially serves many important functions. One is making explicit the spectral features responsible for the recognition of different phonemes, the evaluation of pitch, the perception of voice quality, and other auditory perceptual tasks. Another function is endowing the spectral representation with added stability and noise-robustness [7].

Another interesting area of speculation concerns the question of how the cortical representation can be gracefully mapped unto vocal tract parameters or models. This is an important issue both from a biological and an applications points of view since vocal tract models are heavily utilized in systems for data compression, vocoders, synthesizers, and speech recognizes.

ACKNOWLEDGMENT

Many colleagues participated in the experiments and analyses presented in this paper. They include Huib Versnel, Didier Depireux, Nina Kowalski, Kuansan Wang, Po-wen Ru, Tony Owens, and Preetham Gopalaswamy. This work is supported by the Air Force Office of Scientific Research, The Office of Naval Research, the National Science Foundation through NSF Grant (# NSFD CD 8803012), and the National Institutes of Health.

REFERENCES

[1] Plomp, R. (1976), *Aspects of Tone Sensation*, Academic Press

[2] Merzenich, M., Knight, P., Roth, G. (1975), "Representation of cochlea within primary auditory cortex", *J. Neurophysiology*, vol 28, pp.231-249.

[3] Shamma, S. , Fleshman, J., Wiser, P., Versnel, H. (1993) "Organization of response areas in ferret primary auditory cortex", *J. Neurophysiology*, vol.69(2), pp.367-383.

[4] Oppenheim, A., Willsky, A., and Young, I. (1983), *Signals and Systtems*, Prentice Hall.New Jersey.

[5] Shamma, S., Versnel, H., and Kowalski, N. (1995) "Ripple analysis in ferret primary auditory cortex. I. Response characteristics of single units to sinusoidally rippled spectra", *Auditory Neuroscience*, vol.1(2) *(in press)*.

[6] Shamma, S., and Versnel, H. (1995) "Ripple analysis in ferret primary auditory cortex. II. Prediction of single unit responses to arbitrary spectral profiles", *Auditory Neuroscience* , vol.1(2) *(in press)*.

[7] Wang, K., and Shamma, S. (1995) "Spectral shape analysis in the primary auditory cortex", IEEE *Trans. Speech and Audio (in press)*.

[8] Schreiner, C. and Urbas, J. (1988) "Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields", *Hearing Res.*, vol.32, pp.49-64.

[9] Eggermont J. and Smith G. (1995) "Synchrony between single unit activity and local fiel potentials in relation to periodicity coding in primary auditory cortex", *J. Neurophysiology* , vol.73(1), pp.227-245.

[10] Schreiner, C., and Langner, G., (1988) "Periodicity coding in the inferior colliculuc of the cat", *J. Neurophysiology*, vol.60, pp.1823-1840.