

Theme 4: Pitch and Temporal Coding

Pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch depends mainly on the frequency content of the sound stimulus, but it also depends on the sound pressure and the waveform of the stimulus. (ANSI, 1994)

Pitch is often referred to as a primary parameter in music, a basic concept upon which other musical categories, such as pitch intervals and harmony, can be built (Snyder, 2000). While ANSI defines pitch primarily on a uni-dimensional space, Shepard (1982) postulated a multi-dimensional spatial model, such that the Euclidean distances between the musical tones (in the Western Tonal Tradition) reflect their perceived relations¹. This complex percept is important for human not only to appreciate melody and harmony in music (Dowling, 1999), but it is also used to convey prosody in speech (Winter *et al*, 2001); to segregate multi speakers talking simultaneously (Darwin, 1997); and in tonal languages, such as Cantonese, to carry lexical meanings (Fok, 1974). However, our understanding of the coding mechanism of pitch is only at a basic level. There are two competing theories of how frequency, a physical quantity that pitch percept is dependent on, is coded. The first one is the “place” theory, based on the spectral analysis performed in the inner ear and the fact that different frequencies excite different places along the basilar membrane (Moore, 2003). The other is the temporal coding theory, in which the frequency is related to the time pattern of the neural impulses evoked by the stimulus. There are also different mechanisms of complex pitch perception that have been postulated in the last few decades, ranging from Schouten’s temporal theory, and pattern recognition theory proposed by investigators such as Goldstein and Terhardt, to variations on Licklider’s autocorrelation model, which incorporates both the place and timing information in their pitch extraction algorithms. In this theme report, a sample of the vast and diverse research in pitch perception is presented. Literature reporting elements that influence pitch judgment, such as spectral weighting and auditory grouping mechanisms, are reviewed. Neurophysiological studies are surveyed to ascertain whether there are anatomical correlates that support the aforementioned pitch extraction models. Finally, a brief sample of studies in pitch perception at the cortical level is summarized.

In previous literature, there has been much consensus in the key role of resolved harmonics, and furthermore, many have postulated a dominance spectral region, suggested to be the first four or five harmonics, that are more weighted for pitch estimation². Dai (2000) assessed this relative influence of individual harmonics in terms of pitch judgment by *introducing small changes to the frequencies of the harmonics randomly and simultaneously on each stimulus presentation*. The assumption used is that the change in pitch from a nominal value P_0 can be expressed in a Taylor

series, $\Delta P = P - P_0 = \sum_{i=1}^n w_i \Delta \eta_i + \varepsilon + \xi$, where $\Delta \eta_i$ is the relative frequency change, ε represents

the higher-order terms in the Taylor series and ξ reflects contribution from internal noise. The weights of each harmonic, w_i , can be estimated using a correlation technique. The result shows that the harmonic weight is the strongest if that component is close to 600 Hz and therefore disagrees with some suggestions that the dominance region is roughly constant in harmonic rank. The author acknowledges that the subjects may be influenced by analytic listening, since there were no control mechanisms to monitor whether they were responding to spectral pitches involuntarily, even though they were instructed to respond to the virtual pitch. Therefore, the

¹ Shepherd (1982) proposed a model of seven structural components for the Euclidean synthesis of generalized helical representations for musical pitch. However, the more commonly accepted model is a 5-dimensional helix-based toroidal structure that incorporates tone height, chroma as well as the circle of fifths relationship.

² References in Dai (2000).

weighting functions obtained reflect the dominance regions for the most prominent pitch mode, which can be either virtual or spectral.

The author also discussed the possible mechanisms that might play a role in shaping the dominance region by examining some existing pitch theories. Based on computer simulations of 1000 trials using Wightman's pitch extraction model in the same test paradigm, it was concluded that the Wightman's pattern-transformation theory, which is based on the internal representation of the autocorrelation function of the excitation pattern, derives its weight mainly from the outer- and middle-ear transfer function. Even with the outer- and middle-ear transfer functions taken out of the model, the predicted weight functions have a broad peak near the 6th and the 7th harmonic, and thus it does not provide a good description of the measured weighting functions. However, based on the optimal-processor theory proposed by Goldstein, its weighting function, derived by assuming that the error function can be approximated by the frequency-difference limens (DLs) of harmonics measured within complex tones (Moore *et al*, 1984 in Dai, 2000), provides an excellent fit to the observed data. The dominance region observed in this study is also compared with Terhardt's virtual-pitch theory, in which the components around 600 to 700 Hz are assumed to be the most influential. However, the author pointed out that Terhardt's function is much too broad to produce the estimated perceptual weighting functions. It was concluded that the incorporation of an error function based on frequency DLs in pitch models could potentially improve their predictions.

There are other factors that affect the weighting of a harmonic component in pitch estimation. In some cases, primitive grouping mechanisms³, such as common onset time and lateral position cues, influence the simultaneous perceptual organization. Darwin *et al* (1995) investigated whether pitch perception is also subject to auditory grouping mechanisms that are concerned with sequential organization. The hypothesis is that if a preceding captor tonal sequence is able to perceptually remove a mistuned harmonic from a complex stimulus due to sequential integration causing auditory scene segregation (Bregman, 1990), the pitch shift due to the mistuned harmonic will be decreased. In their first experiment, pitch shift was measured in four testing configurations (see Figure 1). There were no significant differences in terms of pitch shift in the isolated ipsilateral and isolated contralateral presentations. However, there was a statistically significant interaction between the isolated/sequence factor and the level of mistuning (see Figure 2).

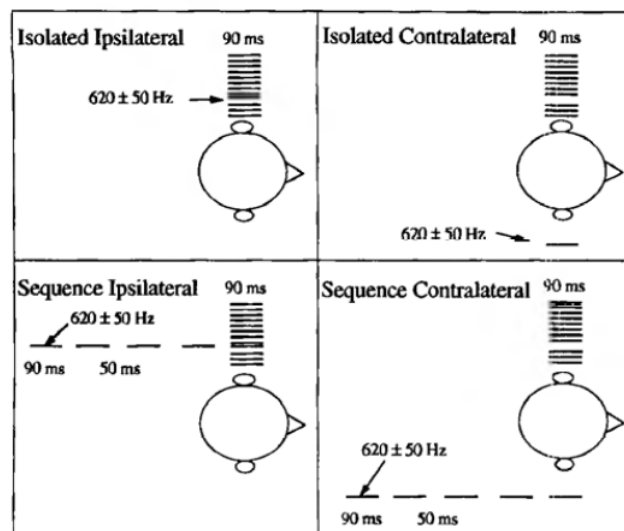


Figure 1 Stimulus configuration for the target complex tone and the mistuned component presented either ipsilaterally or contralaterally in isolated or sequence conditions. (Darwin *et al*, 1995, Figure 1).

³ Discussed in detail by Bregman (1990).

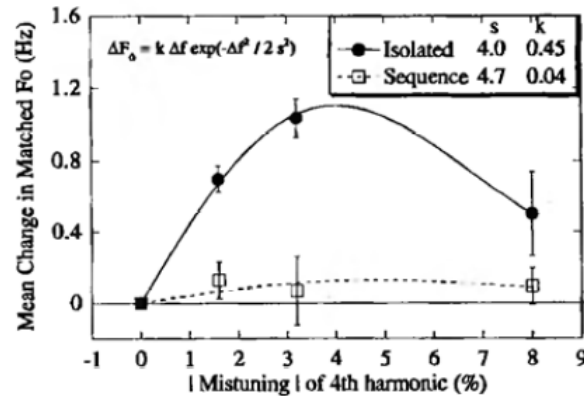


Figure 2 Mean pitch shifts produced by different mistunings of the 4th harmonic of the complex tone. Comparison between isolated and sequential presentations. (Darwin *et al*, 1995, Figure 4)

In their previous studies, the mistuned component was perceptually segregated using a different onset time, and thus it was not conclusive whether the reduction in pitch shift was due to auditory grouping mechanisms or peripheral adaptation. In their second experiment, the subject was presented with an isolated harmonic complex in the ipsilateral ear, and a mistuned harmonic in either the ipsilateral or the contralateral ear, varied in level. Results showed that as the level of the mistuned harmonic is reduced, the mean shift decreases for the ipsilateral condition, but not for the contralateral. Since the mistuned harmonic in the contralateral ear could not have been masked, the absence of any effect on pitch argues against the effects being attributed to adaptation, to the extent that adaptation effects can be equated to a decrease in physical level.

The pattern recognition theories generally suffer two criticisms. Firstly, there is, thus far, lack of convincing biological correlates of the harmonic templates and a model from which they can be generated. Secondly, the models fail to take into account the residual pitch⁴ percepts and their properties. Shamma *et al* (2000) proposed a model for the formation of the harmonic templates, without relying on the exposure of harmonically rich stimuli such as speech and music. Their model is shown in Figure 3. The analysis stage of this model consists of the peripheral component of cochlear filters performing the frequency tuning that is a three-step process: a high-pass filter accounting for velocity coupling of the hair cell cilia; a sigmoid function that describes the nonlinear transducer channels; and a low-pass filter representing the leakage in hair cell currents that gradually attenuates phase-locked responses. The central component of lateral inhibitory network and temporal sharpening enhance the features of the harmonic templates. The final stage of the model is the coincidence matrix, integrating all results over time to produce its final output. Harmonic templates emerge as regions of high coincidence between responses of harmonically related cochlear filters, after the model is passed through broadband noise or click trains (see Figure 4). The network relies on the half-wave rectification to generate nonlinear, but harmonic, distortions and that these distorted harmonics from any filter can be partially coincident with the outputs of other filters that are at harmonically related characteristic frequencies (CFs). The authors argue that even though harmonics at the coincidence stage are generally not synchronized in phase, the rapid phase change near the CF ensures that there will be cases of coincidences.

⁴ "Residue pitch" is associated with unresolved harmonics, very slow click trains, or the envelope of amplitude modulated noise and sinusoids. (Shamma *et al*, 2000)

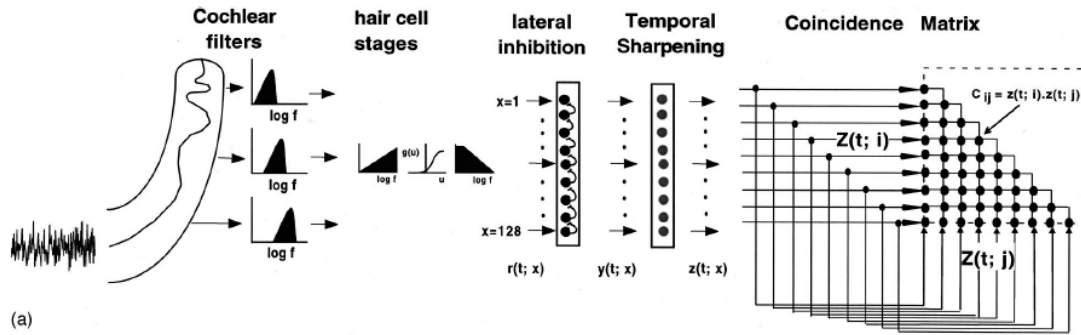


Figure 3 Schematic model of early auditory stages. (Shamma *et al*, 2000, Figure 1a)

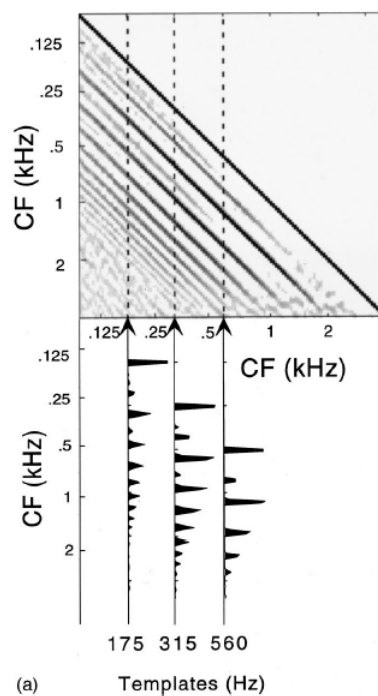


Figure 4 The harmonic templates integrated output of the coincidence matrix. Templates emerge as regions of high coincidence that run parallel to the main diagonal, and are exactly spaced at harmonically related CF distances. These templates are generated by a broadband stimulus, but similar patterns can be found for random click trains stimulus. (Shamma *et al*, 2000, Figure 3a)

The authors suggest that since phase-locking up to at least 2kHz is necessary at the input of the matrix, the coincidence matrix would exist at, or prior to, the inferior colliculus. They also draw the parallel of the synchronized responses at the input of the coincidence matrix to the variety of onset neurons in the cochlear nucleus. It is also conceivable that the octopus cells, the presumed onset-l cells, can serve as the coincidence matrix themselves. The model also tries to incorporate a systematic delay network at the cochlear filter outputs in order to account for the computation of residue pitch. This suggests that it requires another network to extract the pitch. Furthermore, the authors are unsure about the exact nature of the integration of the virtual and residual pitch percepts.

While Shamma *et al* (2000) explored the possible biological correlates for the template model using neural-network modeling, Winter *et al* (2001) turned to the ventral cochlear nucleus (VCN) of

anaesthetized guinea-pigs and examined the temporal discharge patterns of single units in response to iterated rippled noise⁵ (IRN) evoked by stimuli with delays ranging from 1/32 to 1/1000s, seeking for possible neural correlates for pitch extraction algorithms that are based on autocorrelation models. The authors analyzed the first-order and the all-order interspike interval (ISI) distribution of different units to cosine-phase harmonic complexes (CPH), random-phase harmonic complexes (CPH) and IRN stimuli. They showed the degree of interval enhancement⁶ in the all-order, but more prominently in the first-order, *interval histograms* are tuned to the respective IRN delay (or the psuedo-period of the IRN stimuli). Furthermore, the authors showed that there is a strong correlation for the onset-chopper (OC) and the sustained chopper (CS) neurons between these preferred delays for IRN and the preferred delay of the neurons in response to white noise. This suggests that these aforementioned neurons in the VCN could be well-suited to encode the delay of IRN in their first-order ISI that would facilitate the temporal coding of the pitch of complex sounds.

There are, however, some questions in this paper left unanswered. Figure 5 shows that the OC unit has a peak interval enhancement at 10 ms, corresponding to the delay of the IRN stimulus, but the preferred delay is shifted in response to a 100 Hz RPH and CPH stimuli. Why does the preferred delay correspond to a weaker pitch percept for IRN stimulus and not for the stronger pitch percept stimuli? What neural correlates are necessary to utilize the interval enhancement information? The range of the pitch tested on the guinea pig is lower than the range of their vocalization frequency. How would these proposed neural candidates response in a higher frequency range?

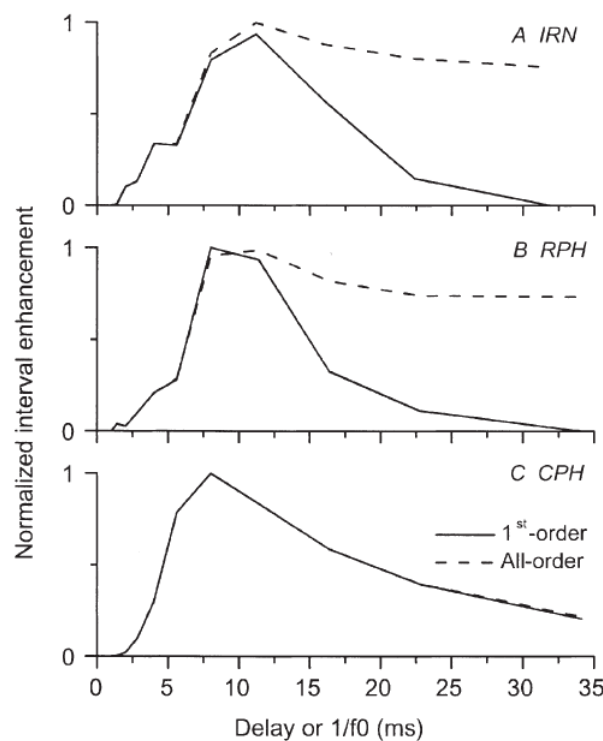


Figure 5 Normalized interval enhancement in an OC unit (Winter et al, 2001, Figure 6)

⁵ IRN is produced by delaying a copy of the noise by d ms and adding it back to the original and iterating the above process n -times. As n increases, the pitch becomes more salient. The authors here argued that IRN precludes the network from using interval histogram for pitch analysis, since it lacks periodicity.

⁶ Interval enhancement is defined by these authors to be the increase in intervals beyond the magnitude of the white noise response.

In recent times, magnetoencephalography (MEG) and other imaging techniques have been used to further our understanding of pitch perception in the auditory cortex. Many studies use the neuromagnetic response N100m to study pitch perception at the cortical level. However N100m is evoked by the onset of sound energy, independent of whether the sound produces pitch. Krumbholz *et al* (2003) used regular interval sound⁷ (RI) to isolate the onset of pitch information, by presenting the subjects with noise smoothly transitioned to RI sound. They showed that the change from noise to RI elicits a transient neuromagnetic response, referred to as the pitch-onset response (POR) but this response is absent if the stimulus changes from RI to noise. Furthermore, Figure 6 shows that the amplitude of the average dipole moments measured in the POR varies with pitch strength and the latency varies with the pitch of the RI sound. Furthermore, their data indicate that the latency of the response is 4 times the delay of the RI (or the inverse of the RI pitch) with an offset of 120ms. The authors also used psychoacoustical studies to ascertain the pitch discrimination threshold⁸ and they concluded that the auditory system has to integrate over a duration of at least 4 times the delay to derive a rough estimate of the pitch for these RI sounds, which is comparable to the results found for POR latency.

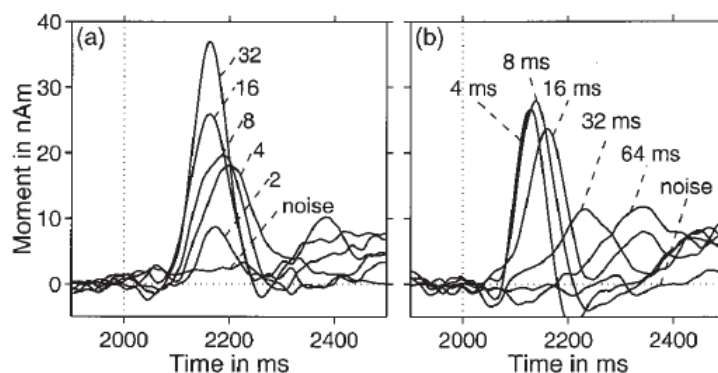


Figure 6 Average dipole moments as a function of time in response to the transition from a noise to a RI sound. (Krumbholz *et al*, 2003, Figure 4)

Besides imaging, another research method to establish the functions performed by different parts of the brain is through the study of subjects with known areas of brain lesion. Zatorre (1988) examined the effect of unilateral temporal-lobe lesion on missing fundamental pitch perception⁹. He concluded that 'only subjects with right temporal lobectomy in whom Heschl's gyri were excised committed significantly more errors than the normal control group on this task... and the results suggest that Heschl's gyri and surrounding cortex in the right cerebral hemisphere play a crucial role in extracting the pitch corresponding to the fundamental from a complex tone.' However, there are a number of experimental uncertainties, in my opinion, that could influence the interpretation of his results. In his control condition, the results reflect a ceiling effect across all tested groups. Using these results as a reference for statistical comparison could skew the data substantially. The exact location and the size of the lesion were not reported. Macro-anatomical landmarks cannot accurately infer borders of functional regions in the cortex. Moreover, the volume of the lesion was not reported, and it could influence the results enormously, especially the excisions were closed to the hippocampus and parahippocampal gyrus. Some subjects were tested 2 weeks following the surgery, others were seen in follow-up examination 1 or more years after surgery. Therefore,

⁷ Equivalent to the IRN stimuli used by Winter *et al* (2001) described previously.

⁸ The subject is required to indicate which of two RI sounds had the higher pitch and the pitch discrimination threshold was defined to be the minimum difference in delay required for statistically reliable discrimination.

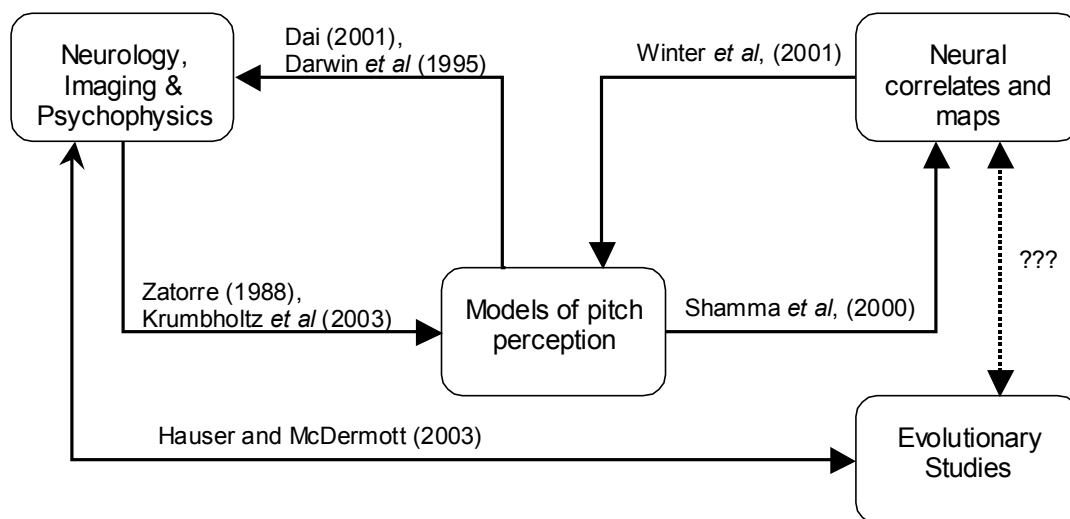
⁹ The observation that a pitch could be associated with the fundamental of a complex stimulus even when the fundamental was absent in the spectrum of the complex stimulus is called the case of the missing fundamental. (Yost, 2000)

plasticity of the brain is yet another possible factor that can affect the inferences drawn by the location of the excision linking to its function.

Main point is that one cannot conclude that the deficit is truly species to F0 perception.

Take home message / Summary / Discussion for Theme 4:

- Pitch is a multi-dimensional percept. We still do not have a clear agreement in the scientific world on the exact mechanism for pitch estimation in a monophonic presentation. Other grouping mechanisms would certainly add on *[or help solve?]* to the challenges in solving this long-standing problem.
- Neurological studies (accompanied by behavioral data) can be useful for determining the functional roles of different part of the cortex in auditory perception. However, the exact location and the volume of the lesion can have a huge effect on the interpretation of the data. Are the macro-anatomical landmarks like gyral and sulcal borders accurate enough to define the functional area in the cortex? How variable are they between subjects, especially in face of a surface that is almost certainly not sufficiently characterized by traditional Euclidean geometry that is based upon the Cartesian coordinate system?
- One branch of scientific research was not represented in this theme report discussion. Could evolutionary studies, especially that on the evolution of the music faculty, shed new lights on how we search for neural correlates for processing pitch? Hauser and McDermott (2003) cited interesting behavioral studies that monkeys, but not songbirds, have the capacity for melodic transpositions, except for atonal melodies. Previous studies have shown that rhesus monkeys showed octave generalization. Could we draw inferences from studies of how animals prefer or are indifferent to consonant and dissonant music on the pitch processing network being more at the peripheral or cognitive level?
- A similar diagram has been presented in Theme Report 1. This flowchart diagram was conceived by Brad Buran, and fine-tuned by Anne Dreyer, Daryush Mehta and myself. It shows how we conceive the different branches of scientific research contribute in solving the pitch perception conundrum.



Papers discussed:

Dai H (2000). On the relative importance of individual harmonics in pitch judgments. *J. Acoust. Soc. Am.* **107**:953-959.

Darwin CJ, Hukin RW, and Al Khatib BY (1995). Grouping in pitch perception: Evidence for sequential constraints. *J. Acoust. Soc. Am.* **98**:880-885.

Krumbholz, Patterson RD, Seither-Preisler A, Lammertmann C, and Lütkenhöner B (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex* **13**:765-772.

Shamma S, and Klein D (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *J Acoust Soc Am* **107**:2631-2644.

Winter IM, Wiegrobe L, and Patterson RD (2001). The temporal representation of the delay of iterated ripple noise in the ventral cochlear nucleus of the guinea pig. *J Physiol* **537**:553-566.

Zatorre RJ (1988). Pitch perception of complex tones and human temporal lobe lesions. *J Acoust Soc Am* **84**:556-572.

Reference:

ANSI, (1994). American National Standards. *Acoustical Terminology S1*. Acoustical Society of America, New York.

Bregman, A.S., (1990), *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT, Cambridge.

Dowling, WJ, (1999). *The Development of Music Perception and Cognition*, in Deutsch, D., (ed.) *the Psychology of Music*, 2nd Edition. Academic Press, New York, pp 603-625.

Fok, YY, (1974). *A perceptual Study of Tones in Cantonese* (Centre of Asian Studies, University of Hong Kong, Hong Kong), Vol. 18.

Hauser, M, McDermott, J (2003). The evolution of the music faculty: a comparative perspective, *Nature Neuroscience*, **6**:7: 663-668.

Moore, BCJ (2003). *An Introduction to the Psychology of Hearing* (5th ED). Academic Press, UK, pp 195-196.

Shepard, RN., (1982). "Structural representations of pitch", in Deutsch, D., (ed.), *The Psychology of Music*, 1st Edition. Academic Press, New York, pp 343-90.

Synder, B (2000). *Music and Memory*, Massachusetts Institute of Technology, Cambridge, pp 195-196.

Yost, W (2000). *Fundamentals of Hearing*, Academic Press, London, p197.