AFFERENT/EFFERENT PITCH PROCESSING

Matthew R. Flax <flatmax@ieee.org>

School of Electrical Engineering and Telecommunications,
University of New South Wales

ABSTRACT: An algorithm is developed to determine the pitch and pitch salience of complex auditory residues. The pitch extraction algorithm is based on both the afferent and efferent neural systems. It focuses on the cross processing nature and innervation of the afferent and efferent systems. This model assumes the simplification of direct feedback (from afferent to efferent signal paths). The cross processed scale (the 'pitch scale') is depicted qualitatively, but remains undefined.

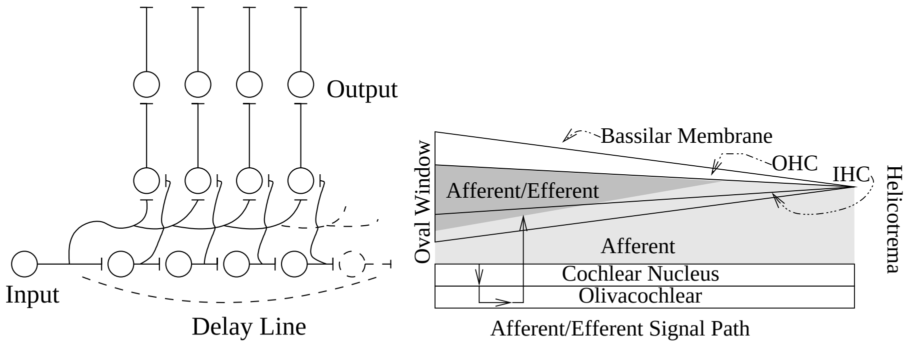
INTRODUCTION

Licklider (1951) suggests that the auditory nerve has surface layers (those close to the basilar membrane) which physically construct an autocorrelation function. Licklider explains that the neurons are inclined to form a delay line which is coupled with summing neurons to the original neural signal, Figure 1(a). Many of these micro delay line systems are shown to build a macro autocorrelation function.

Since Licklider's original review and interpretation of neural innervation in the organ of Corti/eighth nerve, technology for viewing such nerve tissue has improved. Cochlea micrography is the process of sectioning the cochlea into slivers which are visually analyzed using electron microscopes. Various methods of preparing and staining the tissue samples are used to render different morphological tissue features visible. The features of interest are the hair cells, nerve fibres and synapses. Micrography allows the analysis of functional flow within the cochlea neural system. Recent micrography by Rask-Andersen et al. (2000), Thiers et al. (2000) and Thiers et al. (2002) observes the following. i) Primates have a large amount of neural processing present in the spiral ganglion. They differ from other mammalian species in this attribute. ii) Outer Hair Cells (OHC) are part of the afferent and efferent processing system. One efferent nerve is connected to many OHC. iii) The afferent processing system is equally numerous throughout the spiral ganglion. This indicates that the cochlea tonotopic mapping is linearly projected onto the afferent neural system. iv) The efferent processing system exists in the first turn of the cochlea and is not present by the third turn. It is assumed for the purpose of this paper that efferent innervation gives an approximately linear frequency mapping (the basilar/afferent system warps the frequency mapping). In other words, a signal is warped (in frequency) by the basilar/afferent system and re-linearised by the efferent system. v) The efferent and afferent neural system is complexly interconnected. The efferent system is also connected by deeper neural processes. vi) Myelin and synaptic connections in the spiral ganglion are believed to restrict lateral signal processing in further stages. vii) The efferent system develops after birth. Consequently the complex neural processing in the spiral ganglion and around it is also known to develop after birth. Processing in the peripheral neural system is produced by a mixture of connected neural synapses. A coarse visual representation of the layout of the neural interaction of the pre-cortex portion of the auditory system is presented in Figure 1(b).

A majority of the pitch extraction algorithms are classed as either temporal or place models. Temporal models of pitch use temporal neural response patterns to determine pitch. Nerve fibres are simulated and their temporal firings are analysed for the most common firing period. This period corresponds to the period of the pitch. An example of a temporal model is the Meddis and Hewitt (1991) model. The Meddis and Hewitt model is based on psychologically modeling a signal from the outer ear to the neural level. Although the model does not incorporate simultaneous masking, it incorporates temporal masking in the form of neural relaxation. It is termed 'fiber refractory period'. A similar system is also presented by Slaney and Lyon (1990). Licklider's autocorrelation algorithm Licklider (1951) is used by temporal models and forms the neurological basis.

Place models of pitch analyse transformed power spectra. Autocorrelations on the power spectra reveal harmonic complexes (residues). Wightman's pattern-transformation model, Wightman (1973), is a good example of this technique. The model works like this. A coarse spectral analyser develops a rough power spectrum of the acoustic signal. This is a model of the tonotopic mapping to place within the cochlea. A Fourier transformer operates on the power spectrum producing a time related dimension. It



(a) Licklider's neural model.

(b) Afferent/Efferent signal flow.

Figure 1: (a) Neural autocorrelation proposed by Licklider. (b) Afferent/Efferent sectioning and pathways of the hair cells, spiral bundles and some subsequent stages (in the auditory cortex).

is argued that afferent neural wirings compute the Fourier transform which yields an autocorrelation of the power spectrum. This autocorrelation is then analysed for maxima (not at $lag \simeq 0$ seconds where it is obviously a maximum). The maxima then represent the period of the pitch. A concept of pitch strength is developed in Wightman's approach. Pitch strength is simply the magnitude of the maximum. It is found that the pitch strength indicates the perceptual presence of the waveform's pitch. Hence, the stronger the pitch strength, the stronger the perceived pitch. Other place models are based on peak picking techniques. Peak picking is the process of analysing a signal and isolating (indexing) peaks of interest. Peak picking is used to determine possible harmonic components. Terhardt (1979) and Duifhuis et al. (1982) for example pick peaks from power spectra using a masking threshold. Harmonic complexes are determined from the spectral peak set and the pitch is then known to be the residue pitch. It is possible to replace the searching mechanisms used in the fundamental determination mechanisms with an autocorrelation function. A clean set of harmonic components would be presented to the autocorrelation function. The result of autocorrelation would be to generate a maximum amplitude at the lag location matching the residue's fundamental period. This mechanism is not capable of determining virtual pitch, as in Terhardt et al. (1982), however it is capable of determining some formal salience and fundamental weighting. This accomplishes four of five steps in Terhardt's virtual pitch algorithm.

Both place and temporal models assume that autocorrelation mechanisms of some sort are present in the auditory nerve. This is the predominant method for analysing the temporal or place neural firings to determine harmonic or pitch related complexes. This autocorrelation method is linked directly to Licklider's article from 1951.

A listing of seven models Cohen et al. (1995); Bilsen and Goldstein (1974); Meddis and Hewitt (1991); Yost and Hill (1979); Terhardt (1979); Duifhuis et al. (1982); Wightman (1973) and their capabilities are given by Cohen et al. (1995). The authors indicate that the Meddis and Hewitt model for finding pitch is both the most tested and reliable. It is uniquely capable of accurate amplitude modulated noise pitch discrimination.

Cariani and Delgutte (1996a,b) discuss nerve fiber discharge patterns. Their study is primarily conducted on the neural pathways of cats. The concept of 'pitch salience' is derived from pooled autocorrelograms. Salience is defined to be the lag magnitude to mean magnitude ratio. A salient pitch is one with a high ratio. Again like pitch strength salience, it is a perceptual measure, meaning that a large salience is perceived more easily then a low salience.

Toward the future of pitch processing, Slaney (1995) states that processing systems are typically too focused on bottom up processing. Bottom up processing is defined as the processing of signal flow from sensory input to higher systems of analysis. Artificial intelligence assumes models of the world and hence these systems incorporate a top down signal flow and processing. People naturally process the world using a combination of both top down and bottom up processing. We have certain expectations of objects and interactions around us. Objects also have their own nature which makes them all different. For this reason we use both afferent and efferent signals to process and understand our environment.

All previous models for deriving pitch are based on afferent processing systems. This may be a valid method for pitch perception within the wider mammalian kingdom, however primates are known to have rather complex afferent/efferent interaction. This two way signal influence is known to appear within the first layers of the auditory spiral ganglion. These are the first layers of neurons within the auditory nerve.

It is apparent that a model for pitch processing which incorporates efferent and afferent signals is required. This article outlines a pitch determination algorithm which mimics the neural innervation of the cochlea periphery. This article compares the purely afferent approach to determine pitch with that of the new afferent/efferent cross processed approach. The next section (section 1) defines what is meant by an afferent/efferent cross processing system. Sections 2 and 3 define the algorithm and experiments with a place model. Section 4 concludes.

MIMICKING THE PRIMATE PERIPHERAL AUDITORY NEURAL PROCESSING SYSTEM

One may subdivide neural auditory processing into a system of afferent autocorrelation, and a system of efferent/afferent processing, Figure 2(a). For simplicity, assume that the efferent feed back system contains neural spike trains which are gathered by the myelinated afferent system. As part of this assumption, efferent feedback is an un-processed copy of the afferent signal, which occurs at the periphery of the auditory cortex (Figure 2(a), block a]). Assume further that there are many efferent feedback paths which occur deeper within the auditory cortex (block b]). This model only addresses the peripheral feedback path, which is a direct copy of the afferent signal (block a]). The rest of this section will define the nature of the afferent/efferent processing system (block c]). This processing occurs at the periphery of the auditory nerve, where it is reported that there is a large amount of afferent/efferent interaction or cross processing. The result of this cross processing system (signal d]) is the topic of this article. It is believed that this signal is channelled to the deeper auditory cortex as a separate cue (separate to the afferent pitch mechanism described earlier).

Basilar membrane tonotopic mapping is modeled using the number of ERBs scale (ERBS) ($f_E = 21.4log_{10}(4.37f + 1)$ where f is in kHz) Glasberg and Moore (1990). The ERBS scale (f_E) warps the frequency scale. It stretches the lower frequencies and compresses the higher frequencies. The ERBS scale arises due to the mechanics of the cochlea de Boer (1995). Energy is processed by the outer ear (which is not incorporated into this model for pitch perception) and is presented to the cochlea through the oval window using the stapes. Neural transduction is accomplished within the organ of Corti as a process of the physical forces and pressures. The pressures and forces are imposed on the outer and inner hair cells which stimulate the afferent neural system.

As the cochlea is a frequency transformation device, a signal $s(t)$ which is presented to the cochlea is frequency transformed ($S(f) = \Im\{s(t)\}$) where $\Im$ represents the Fourier Transform. The signal is also frequency warped according to the organ of Corti response

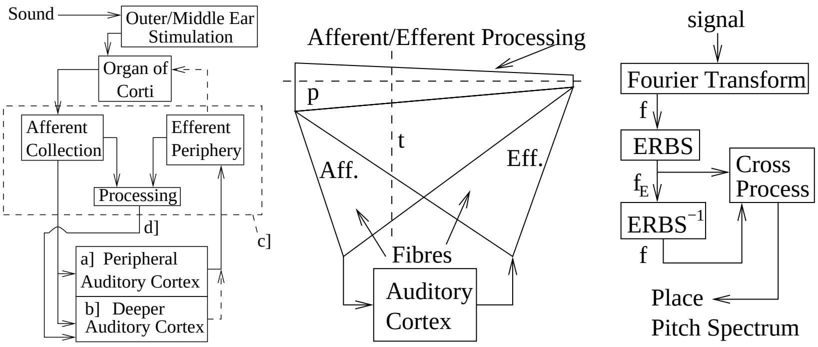
$$S(f_E) = S(ERBS(f)) \tag{1}$$

where $ERBS$ is a frequency warping operator. Consequently, the afferent neural system is stimulated with the signal $S(f_E)$. The neural signal phase is known to be shifted somewhat, however for the purposes of determining the pitch of the signal, only the signal magnitude ($|S(f_E)|$) will be processed.

It is known that the efferent neural system connects to the OHC. For this reason, the efferent neural system must affect the physical response of the cochlea. This feedback is not allowed for in any of the physical models (although other models of feedback are incorporated). Consequently the interaction between the efferent neural system and the organ of Corti will not be incorporated into this model of neural pitch processing. It is envisaged that its incorporation would improve cochlea and pitch processing models.

The efferent system is driven by the afferent system through the auditory cortex. For the purposes of this algorithm it is assumed that the efferent system interconnects to the afferent system in an arbitrary (possibly random) way. This method of connection is modeled by a cross correlation process (similar results are obtained when using convolution).

The innervation of the efferent system is large for high frequencies, a medium amount for mid frequencies and a very small amount for low frequencies. In other words, the high frequency efferent neurons are connected to many high frequency afferent neurons. This means that local to the first turn of the



(a) Cochlea and Neural auditory system block diagram.

(b) Neurophysiologic depiction.

(c) Algorithm block diagram

Figure 2: (a) Auditory block diagram. (b) Neurophysiologic depiction of the processing system, nerves and auditory cortex. Location is also shown of place ('p' in diagram) and temporal('t' in diagram) models. (c) Block diagram of the place algorithm for pitch extraction.

cochlea, the efferent system over-samples the afferent system. The mid frequency efferent neurons are connected to a similar amount of mid frequency afferent neurons. This means that local to the mid turn of the cochlea, the efferent system samples the afferent system linearly. The low frequency efferent neurons are connected to very few afferent neurons. This means that local to the last turn in the cochlea, the efferent system under-samples the afferent system. This sampling corresponds to a de-warping of the frequency ordinate of the afferent signal. The de-warping is such that it stretches the high frequencies, leaves the mid frequencies unaltered and compresses the low frequencies. This is in opposition to the function of the ERBS. It is a re-linearisation of the ERBS, Equation 1. For the purposes of experiment, the re-linearisation will use an inverse ERBS ($ERBS^{-1}$) where

$$f = ERBS^{-1}(f_E) \tag{2}$$

A PLACE MODEL, AFFERENT/EFFERENT PITCH EXTRACTION ALGORITHM

The place and temporal models of pitch processing were discussed in the introduction. The place algorithm will be defined here. Figure 2(b) depicts a block neurological image of the auditory system. It depicts the afferent/efferent processing system, the fibres which connect it to the cortex and the cortex itself.

Place models of pitch processing operate on signals which are power spectra. This type of model operates at one instant in time. It is therefore a lateral snapshot of the processing which occurs where the afferent and efferent systems connect (cross). It is depicted by the lateral line 'p' in Figure 2(b). The algorithm may be dictated in the following way, as depicted in Figure 2(c) :
a] Fourier Transform a window of the signal.
b] Warp the frequency amplitudes using $ERBS$.
c] Re-linearise the frequency spectrum using $ERBS^{-1}$
d] Cross process the linear and warped amplitude spectra to produce a pitch spectrum.

EXPERIMENTAL RESULTS

A C4 major triad (chord) is played by a viola and used for analysis. The fundamental frequency of the 1st note in a C4 major chord is at 261.63 Hz. This is considered its most salient pitch and is the pitch which defines the perceptual tuning of the chord. The afferent/efferent processing system is analysed by cross correlating the afferent signal ($|S(f_E)|$) with the efferent signal ($|S(f)|$). Results are compared to the traditional purely afferent techniques.

The results of place processing for both a sine tone (at the fundamental frequency C4=261.63 Hz) and the C4 major chord are plotted in Figure 3. A raw power spectrum is also depicted for comparison (Figure

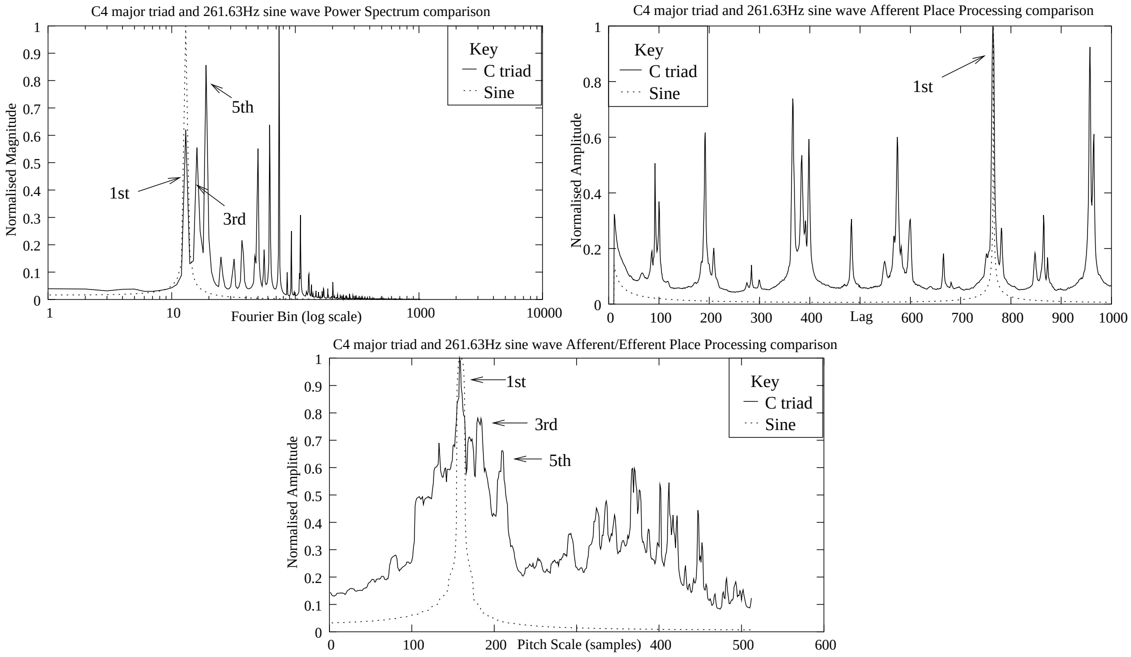


Figure 3: Comparison of the place processing technique. (a) The Fourier power spectrum must be shown on a log scale as all detail is squashed in the lower Fourier bins. It is not clear that the 1st peak corresponds to the fundamental frequency. (b) Purely afferent place pitch processing. The fundamental frequency is correctly highlighted. Detail is spread evenly across the pitch scale. (c) Afferent/efferent place pitch processing. The pitch scale clearly depicts the fundamental frequency as the most salient and the other major triad components are clearly salient too. Detail is spread evenly across the scale.

3(a)). It is clear that the raw power spectrum has an incorrect salience for the fundamental frequency (it is less salient then the 5th component). The most salient pitch is much higher then any of the first, third or fifth constituent fundamentals. All information is also compressed into the lower Fourier frequency bins. The design of an algorithm which would pick the fundamental frequency from the raw Fourier power spectrum is an extremely difficult task for complex residuals (such as that shown). This is why afferent based pitch research has been developed for the past fifty years. The purely afferent processed signal (Figure 3(b)) attributes accurately a strong pitch salience for the fundamental frequency. The salience is small. The afferent/efferent processed signal (Figure 3(c)) attributes accurately a strong pitch salience for the fundamental frequency. Numerically, the salience is 3. This matches the best salience found for simple pitched signals by Cariani and Delgutte (1996a). For the first (C4) note on its own (not pictured here), a pitch salience of 4.5 is obtained. A single note is still a relatively complex signal, as the viola is rich in overtones. The 3rd and 5th fundamentals in the major triad are clearly visible with strong salience as well. It is a simple task to write an algorithm which peak picks the pitch from this processed signal.

The place model which incorporates afferent/efferent cross processing seems to retain perceptual salience. Namely in the C4 triad, all of the constituent notes are visibly present with strong salience. This information is lost in other place techniques. The afferent/efferent cross processing technique is unique in its capability of retaining and communicating this information. This afferent/efferent model of pitch determination successfully extracts pitches from other complex stimuli not reported here. These include pitch shifting, modulated noise and others.

CONCLUSION

This model of primate pitch detection incorporates efferent processing. Afferent/efferent cross processing produces another pitch cue for the auditory cortex. The pitch salience of the afferent/efferent cross processed cue is perceptually relevant to primates. A salient cue is produced by the place processing model which incorporates afferent/efferent processing. Important complex pitch features are retained in the afferent/efferent place model.

Efferent and afferent neural innervation is used to determine the type of cross signal processing which occurs at the auditory nerve periphery. Efferent nerve fibre innervation reverses the tonotopic mapping of the afferent system, which is described by the ERB scale. The tonotopic reversal and cross signal processing is only applicable to primates. This is because primates are the only known mammals to have a complex afferent/efferent processing system which extends to the spiral ganglion. It is possible that the deeper auditory cortex determines pitch from several independent cues. Those of the purely afferent system and the cross afferent/efferent processing system.

ACKNOWLEDGEMENTS Thanks to A/Prof. W. Harvey Holmes for useful suggestions on the format of this article. Thanks to A/Prof. Jesse S. Jin for overall PhD guidence. Thanks to the School of Electrical Engineering & Telecommunications UNSW, for support in my last period of PhD research.

REFERENCES

Bilsen, F. and Goldstein, J. (1974). Pitch of dichotically delayed noise and its possible spectral basis. *Journal of the Acoustical Society of America*, 55:292–297.

Cariani, P. and Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, 76:1698–1716.

Cariani, P. and Delgutte, B. (1996b). 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. *Journal of Neurophysiology*, 76:1717–1734.

Cohen, M., Grossberg, S., and Wyse, L. (1995). A spectral network model of pitch perception. *Journal of the Acoustical Society of America*, 98:862–879.

de Boer, E. (1995). On equivalence of locally active models of the cochlea. *Journal of the Acoustical Society of America*, 98:1400–1409.

Duifhuis, H., Willems, L., and Sluyter, R. (1982). Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception. *Journal of the Acoustical Society of America*, 71:1568–1580.

Glasberg, B. and Moore, B. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47:103–138.

Licklider, J. (1951). A duplex theory of pitch perception. *Experientia*, 7:128–134.

Meddis, R. and Hewitt, M. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America*, 89:2866–2882.

Rask-Andersen, H., Tylstedt, S., Kinnefors, A., and Illing, R. B. (2000). Synapses on human spiral ganglion cells: a transmission electron microscopy and immunohistochemical study. *Hearing Research*, 141:1–11.

Slaney, M. (1995). A critique of pure audition. *International Joint Conference on Artificial Intelligence*, pages 13–18.

Slaney, M. and Lyon, R. (1990). A perceptual pitch detector. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1:357–360.

Terhardt, E. (1979). Calculating virtual pitch. *Hearing Research*, 1:155–182.

Terhardt, E., Stoll, S., and Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America*, 71(3):679–688.

Thiers, F., Burgess, B., and Nadol, J. (2000).]hrefhttp://www.elsevier.com/locate/issn/03785955Prevalence and ultrastructural morphology of axosomatic synapses on spiral ganglion cells in humans of different ages. *Hearing Research*, 150:119–131.

Thiers, F., Burgess, B., and Nadol, J. (2002). Axodendritic and dendrodendritic synapses within outer spiral bundles in a human. *Hearing Research*, 164:97–104.

Wightman, F. (1973). The pattern-transformation model of pitch. *Journal of the Acoustical Society of America*, 54:407–416.

Yost, W. and Hill, R. (1979). Models of the pitch and pitch strength of ripple noise. *Journal of the Acoustical Society of America*, 66:400–410.