A NEURAL MATRIX MODEL FOR ACTIVE TRACKING OF FREQUENCY-MODULATED TONES

Kiyoaki Aikawa [1], *Hideki Kawahara* and *Minoru Tsuzaki*

ATR Human Information Processing Research Laboratories
2-2 Hikaridai, Seikacho, Sorakugun, Kyoto 619-02 Japan

ABSTRACT

Previous paper has shown that the dynamic process of perceiving frequency-modulated (FM) tones can be well simulated by a sound tracking model represented by second-order systems. This paper proposes a new neural operational model for tracking FM tones using a tonotopic neural matrix. The neural cells are connected to build up an auto-regressive architecture. This neural matrix model is characterized by a novel spectral interpolation algorithm based on Lp-norm. The new tracking model enables simultaneous tracking of multiple FM tones. This paper demonstrates that the neural tracking model can successfully simulate perceptual effect such as pitch bounce caused by crossed linear sweep tones. The perceptual overshoot effect is also reasonably explained.

1. INTRODUCTION

The authors have discovered two important effects on pitch perception [1]. The first is that the perceived pitch image of a linear frequency glide is not in fact linear, but warped. The pitch change is slow at the tone initial and is accelerated afterward. The second is that an abrupt change in the sweep slope in a unidirectional FM tone induces perceived pitch ringing (oscillation).

The authors have shown that the dynamic characteristics of the FM perceiving process can be explained by assuming 2nd-order systems. The global warped shape is explained by a 2nd-order system showing a slow response. The natural frequency of a typical slow response is 2 Hz. The pitch ringing effect is explained by a 2nd-order system showing a fast response. The natural frequency of a typical fast response is 7 Hz. Other evidence is that the pitch ringing can be suppressed by shaping the frequency trajectory of the FM tone with the inverse filter of the 2nd-order FM tracking model. The ringing effect cannot be explained by a lower-order system.

This paper proposes a FM tracking model that can simultaneously track multiple FM tones. A novel neural computational algorithm is proposed for tracking FM tones on an AR neural matrix model. Simulations explain several important effects on pitch perception.

[1]Currently at:
NTT Human Interface Labs. Speech and Acoustics Lab.
Email: aik@nttspch.hil.ntt.jp

2. FM TRACKING MODEL

The FM tracking process is represented by the second order system in the s domain of the Laplace transform as

$$H(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \qquad (1)$$

$$\omega_n = 2\pi f_n \qquad (2)$$

where f_n denotes the natural frequency and ζ the damping factor [1].

The 2nd-order system is represented by a 2nd-order AR (auto-regression) model. Let the z transform of input and output variables $x(i)$ and $y(i)$ be $X(z)$ and $Y(z)$, then the AR model is given by

$$Y(z) = \frac{Gz^{-1}}{1 + \alpha_1 z^{-1} + \alpha_2 z^{-2}} X(z) \qquad (3)$$

$$\lambda = 2\pi \frac{f_n}{f_s} \qquad (4)$$

$$\alpha_1 = -2e^{-\lambda\zeta} \cos \lambda\sqrt{1-\zeta^2} \qquad (5)$$

$$\alpha_2 = e^{-2\lambda\zeta} \qquad (6)$$

where λ is a normalized frequency between $-\pi$ and π. α is a linear prediction coefficient. G is the gain constant of the system. f_s is the sampling frequency of peak frequency trajectories. Figure 1 shows the pitch tracking trajectories using 2nd-order systems with three different natural frequencies under critical damping conditions. Figure 2 shows the simulated pitch ringing. The natural frequency is 7 Hz and the damping factor is 0.2.

The time domain input/output relation is given by

$$y(i) = Gx(i-1) - \alpha_1 y(i-1) - \alpha_2 y(i-2) \qquad (7)$$

Equation (7) must output the same value as the input if the input is stationary. Then the gain constant is given by

$$G = 1 + \alpha_1 + \alpha_2 \qquad (8)$$

$$0 < G < 1 \qquad (9)$$

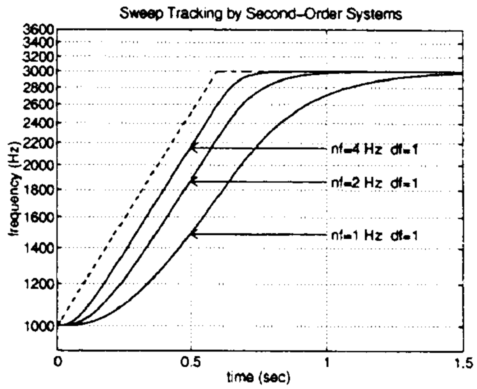

Figure 1. Tracking of a sweep tone with three kinds of 2nd order systems.

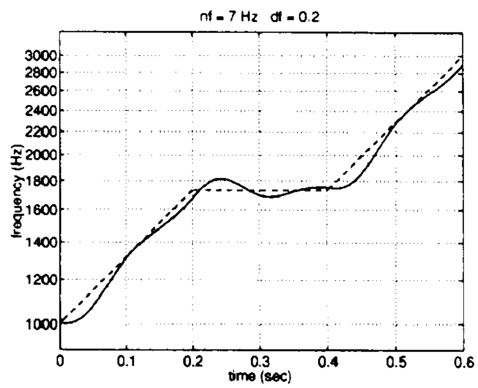

Figure 2. Simulated pitch ringing using a 2nd-order system.

3. TRACKING ALGORITHM

The primary auditory cortex shows tonotopic organization [2], and its response to FM tones has been analyzed [3]. However, the dynamic process of perceiving a continuous FM tone has not been clarified. This paper derives a neural model for tracking multiple FM tones based on neurophysiological findings.

3.1. TWO STEP TRACKING OPERATION

To track FM tones with a neural matrix, Eq. (7) is divided into a two-step operation such that

$$u(i) = \frac{1}{-\alpha_1 - \alpha_2}\{-\alpha_1 y(i-1) - \alpha_2 y(i-2)\} \quad (10)$$

$$y(i) = Gx(i-1) + (-\alpha_1 - \alpha_2)u(i) \quad (11)$$

where $(-\alpha_1 - \alpha_2)$ is a constant to adjust the gain of $u(i)$ for a constant input.

Equation (5) shows that α_1 is negative under the condition

$$2\pi\frac{f_n}{f_s}\sqrt{1-\zeta^2} < \frac{\pi}{2} \quad (12)$$

Equation (6) shows that α_2 is always positive and less than one. Therefore, $u(i)$ is obtained by extrapolating the direction from $y(i-2)$ to $y(i-1)$. This term acts as a momentum factor.

Since G is positive and less than 1,

$$-\alpha_1 - \alpha_2 = 1 - G \quad (13)$$

is positive. Then, $y(i)$ in Eq. (11) indicates the interpolation between $x(i-1)$ and $u(i)$.

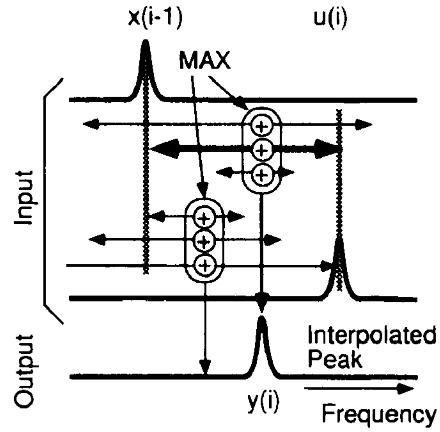

Figure 3. Finding an interpolated peak position between two input spectra. Bold arrows show the case of a given interpolation ratio.

3.2. SPECTRAL INTERPOLATION

The tracking frequency is to be indirectly represented by the peak positions of a firing histogram on a tonotopic neural array. A firing histogram can be regarded as a spectrum. This paper assumes that peaks on the firing histogram are steep enough. Equations (10) and (11) must be calculated indirectly between spectra.

A new algorithm is proposed for interpolating two spectra. A spectrum is capable of representing multiple FM tones by its peaks. To calculate an interpolated spectrum,

1. set the output channel number k to one,
2. get the sum of firing levels of two neural cells located at a given distance ratio from channel k,
3. take the max of the level sum to be the output of the channel k.
4. repeat (2) to (4) incrementing k by one.

Figure 3 schematically shows the interpolation process. When k is at the given interpolation ratio from two peak positions, the level sum indicates the maximum value. The bold arrow shows this case. The spectral interpolation can be calculated by the neural connection shown in Fig. 4. This connection is called a counter tonotopic connection. The connection in Fig. 4 shows the case that the interpolation ratio is 1:2.

3.3. Lp-NORM

To avoid successive search for the maximum value, a new formulation for spectral interpolation is proposed using Lp-norm. Let the input spectra be $f(k)$ and $g(k)$. k is the

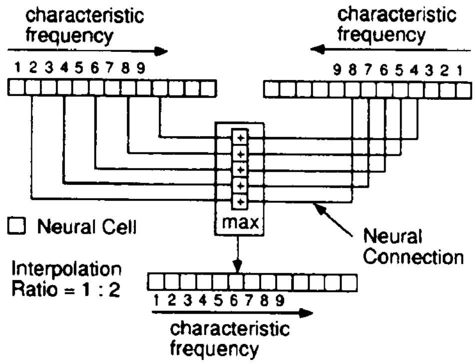

Figure 4. A counter tonotopic neural connection architecture for interpolating two input spectra. Interpolation ratio is 1:2.

channel number. The output firing spectrum $q(k)$ is obtained by

$$h(k,l) = f(k+rl) + g(k-(1-r)l) \tag{14}$$

$$p(k) = \left\{ \frac{\sum_{l=-N}^{N} h(k,l)^{\eta} w(l)^{\eta}}{\sum_{l=-N}^{N} w(l)^{\eta}} \right\}^{1/\eta} \tag{15}$$

$$q(k) = R[p(k) - \theta(k)] \tag{16}$$

where N is the number of channels and r denotes the interpolation ratio. The number of channels was fixed to 200 in the experiments.

Equation (14) calculates the sum of two firing levels in each spectrum at a given distance ratio r from an output channel. Equation (15) calculates the Lp-norm of the firing level sum. In the experiments, the value of η was fixed to 10.

$R[]$ is a rectifier. The spectral peak positions of $f(k)$ and $g(k)$ represent $y(i-2)$ and $y(i-1)$ in Eq. (10). The spectral peak positions of $f(k)$ and $g(k)$ represent $u(i)$ and $x(i-1)$ in Eq. (11). $w(k)$ denotes the neighbor weight and $\theta(k)$ the peak detection threshold.

3.4. NEIGHBOR WEIGHT

If multiple FM tones exist, it is possible to calculate every interpolation among the spectral peaks. To separately track each stream of a FM tone, a neighbor weight is applied to spectral interpolation. The shape of the neighbor function is Gaussian such that

$$w(k) = \exp(-\frac{k^2}{2\sigma^2}). \tag{17}$$

where σ controls the neighbor width. σ was 1/8 in the experiments. The number of tones to be tracked simultaneously depends on the value of σ.

3.5. PEAK DETECTION THRESHOLD

The interpolated spectrum $p(k)$ in Eq. (15) indicates the maximum value of the input spectra if k is not at the interpolated position. This bias can be removed by a rectifier with the threshold

$$d(k,l) = f(k+l)^{\eta} + g(k+l)^{\eta} \tag{18}$$

$$\theta(k) = \gamma \left\{ \frac{\sum_{l=-N}^{N} d(k,l) w(l)^{\eta}}{\sum_{l=-N}^{N} w(l)^{\eta}} \right\}^{1/\eta} \tag{19}$$

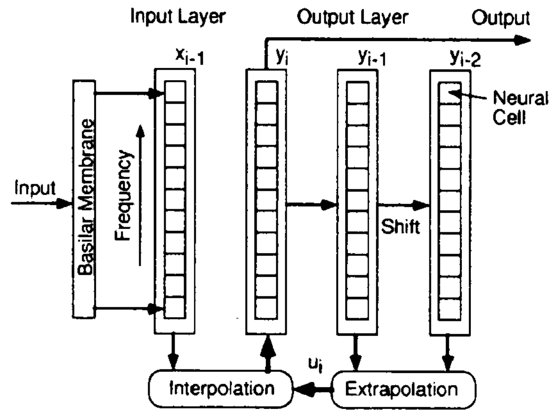

Figure 5. An FM tracking model represented by an AR neural matrix.

where γ is the multiplier to remove low level peaks and side-lobes generated by interpolation process. γ was set to 1.2 in the experiments.

To detect new spectral peaks, Eq. (18) is replaced by

$$d(k,l) = f(k+l)^{\eta}. \tag{20}$$

In this case, the threshold is determined only by the history of the input spectra and is independent of the new input.

3.6. PEAK NORMALIZATION

One way to normalize the amplitude of spectral peaks is to normalize the maximum peak to 1. However this is not necessarily appropriate for a spectrum having multiple peaks.

To avoid the gain decay by iterative tracking, a sigmoid function is employed. To avoid widening of spectral peaks, lateral inhibition is also applied in combination with the sigmoid.

A lateral inhibition function can be given by the difference of two Gaussian distributions of different variance.

3.7. AR NEURAL MATRIX MODEL

The frequency tracking of Eq. (7) can be performed by the AR neural matrix model shown in Fig. 5. The input layer of the neural matrix corresponds to $x(i-1)$ in Eq. (7). $u(i)$ is calculated from $y(i-1)$ and $y(i-2)$. Then $y(i)$ is calculated from $u(i)$ and $x(i-1)$. The iteration successively proceeds by shifting $y(i-1) \to y(i-2)$ and $y(i) \to y(i-1)$.

4. TRACKING EXPERIMENT

4.1. MULTIPLE FM TONES

Three FM tones were simultaneously tracked by the proposed AR neural matrix model. Figure 6 shows the tracking trajectories. The dashed lines show the stimulus FM tones. Each peak has a narrow Gaussian distribution. The natural frequency of the 2nd-order system is 4 Hz and the damping factor is 1. The dotted patterns show the neural firing patterns. Each FM tone is tracked with a 2nd-order response.

4.2. CROSSED FM TONES

A crossed ascending and descending FM tone complex is perceived as a bounced pitch image. An explanation for this has been proposed [4]; however, it is not clear enough. A simulation experiment was therefore carried out to track

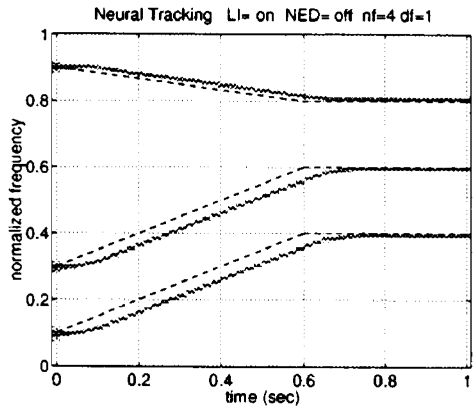

Figure 6. Simultaneous tracking of three FM tones. LI: lateral inhibition; NED: new event detection; nf: natural frequency (Hz); df: damping factor.

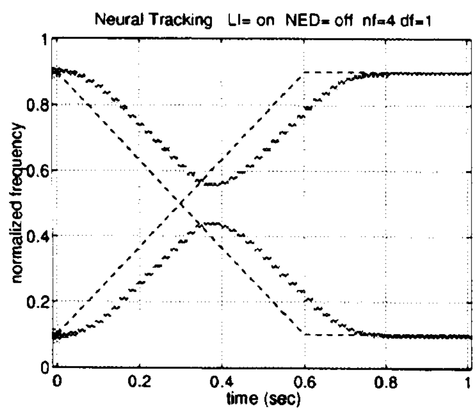

Figure 7. Tracking of two crossed FM tones under a critical damping condition. LI: lateral inhibition; NED: new event detection; nf: natural frequency (Hz); df: damping factor.

the crossed FM tones by the proposed AR neural matrix model. Figure 7 shows the tracking trajectory with the 2nd-order system. Its natural frequency is 4 Hz and the damping factor is 1. This figure shows that the neural matrix model well simulate the pitch bounce effect.

4.3. OVERSHOOT

When a sweep tone is followed by white noise, the final frequency is perceived as the sweep being extended into the noise [5]. Moreover, when a vowel transition is followed by noise, a similar overshoot effect is observed [6].

Figure 8 shows the simulated result of perceptual overshoot using the neural matrix model. The tracking proceeds even after the stimulus tone is turned off, if the tracking is not terminated. This figure indicates that the perceptual overshoot effect can be explained by the effect of the momentum term of the 2nd-order system.

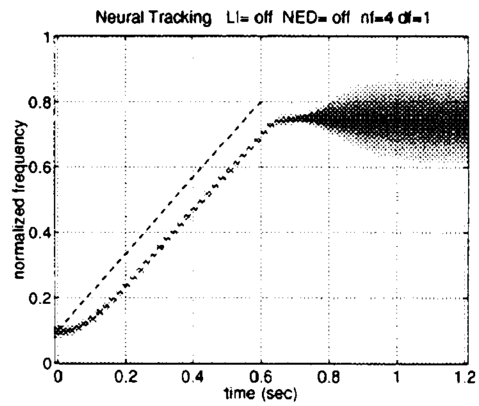

Figure 8. FM tone tracking in case that the FM tone is turned off at 600 ms. The tracking continues toward the sweep direction even after the end of the FM tone.

5. CONCLUSIONS

A neural matrix model for tracking FM tones has been proposed. The model can simultaneously track multiple FM tones with a second-order response. A novel countertonotopic neural network architecture has been proposed in combination with a spectral interpolation algorithm using Lp-norm. The whole tracking algorithm is represented by a neural matrix having a tonotopic auto-regression architecture. Simulation experiments have demonstrated that the proposed neural tracking model can explain perceptual effects such as pitch bouncing and pitch overshoot.

REFERENCES

[1] K. Aikawa, M. Tsuzaki, H. Kawahara and Y. Tohkura, "Pitch ringing induced by frequency-modulated tones.", *J. Acoust. Soc. Am.*, vol. 98, no. 5, Pt.2, pp. 2926 (1995-11).

[2] Y. Hosokawa, J. Horikawa and M. N. I. Taniguchi, "Spatiotemporal binaural patterns of guinea pig auditory cortex studied by the method of optical recording", *Technical report of ASJ*, vol. H-94, no. 26, pp. 1–8 (1994).

[3] J. R. Mendelson, C. E. Schreiner, M. L. Sutter and K. L. Grasse, "Functional topography of cat primary auditory cortex: responses to frequency-modulated sweeps", *Exp Brain Res*, vol. 94, pp. 65–87 (1994).

[4] A. S. Bregman: "Auditory Scene Analysis: The perceptual organization of Sound", MIT Press (1990).

[5] K. Kurakata, M. Matsui and A. Nishimura, "Perceptual trajectory of the continuity effect of frequency glided tone", *Technical report of ASJ*, vol. H-94, no. 71, pp. 1–9 (1994-12).

[6] I. Masuda, K. Aikawa and M. Tsuzaki, "Effect of following noise on the perception of transient in continuous vowels", *Technical report of ASJ*, vol. H-95, no. 61, pp. 1–8 (1995).