Study on prediction of perceived spectral sequences based on spectral peak-tracking model

Noriyoshi SAKAGUCHI

School of Information Science,
Japan Advanced Institute of Science and Technology

February 13, 1998

Keywords: Spectral representation in the primary auditory cortex, Phonemic restoration, Stream segregation, Signal separation.

1 Introduction

Recently, Stream Segregation that a single stream is grouped from sequences and the prediction of frequency modulation have researched.

Aikawa *et al*[1] have reported that the mechanism of pitch tracking is described by second order system. However, it could predict only the peak of frequency. On the other hand, Masuda *et al*[2] have investigated perception of vowels at the end added noise of the vowel-to-vowel transition and suggest the existence of extrapolation for the temporal change of spectrum. They have reported that phonemic restoration is replicated using IFIS.

This papar presents, a method for extracting signal in interfering noise by predicting and tracking transitions of four parameters(frequency, amplitude, bandwidth, and symmetry) which represent spectral peaks using Auditory Cortex 1 model.

2 Spectral peak-tracking model

2.1 Spectral representation in the primary auditory cortex(A1)

In Figure 1(A), representation of A1 described by Shamma[3] is approximated by Gabor function $\psi(\omega)$ and is given from an input spectral envelope $p(\omega)$ by wavelet transforms along the frequency axis. The response $r(s, f, t)$ of A1 in time t is formulated by Equation(1).

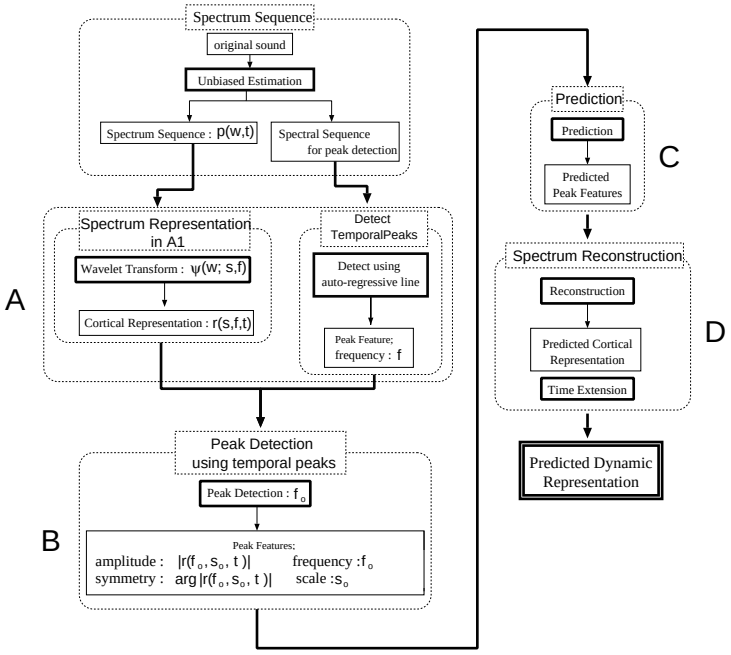

Figure 1: The spectral peak-tracking model

$$r(s,f,t) \quad = \quad \frac{1}{|s|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \psi^*(\frac{\omega - f}{s}) p(\omega, t) d\omega \tag{1}$$

where, s is the scaling parameter, f is the position of frequency, $|r|$ is the amplitude, $\arg r$ is the symmetry, and $*$ indicates the complex conjugate. Input for the spectral peak-tracking model is log spectral sequences estimated by the unbiased estimation method.

2.2 Algorithm for sampling spectral peaks

In Figure 1(B), peak frequencies are extracted from the input spectral envelope using auto-regressive line. This value is a candidate of the peak frequency given a maximum amplitude $|r|$. Scale $s_0(t)$, peak frequency $f_0(t)$, peak amplitude $|r|$, symmetry $\arg r$ are fixed by searching the maximum amplitude $|r(s_0, f_0, t)|$ within the limit of f.

2.3 Prediction of parameters

In Figure 1(C), the four parameters are predicted by a second differential Equation(2).

$$a(1-w)y''(t) + \{b(1-w) + cw\}y'(t) + y(t) = (1-w)x(t) \tag{2}$$

where, a, b, and c are constants, $x(t)$ and $y(t)$ are input and output corresponding to four parameters, and w is a weight function.

2.3.1 Prediction in noise-free section

In noise-free section, Equation(2) is transformed into Equation(3) as $w = 0$.

$$y[n] = Gx[n-1] - \alpha_1 y[n-1] - \alpha_2 y[n-2] \tag{3}$$

$$\begin{aligned} G &= \frac{2}{2a+b} \\ \alpha_1 &= \frac{2(1-2a)}{2a+b} \\ \alpha_2 &= \frac{2a-b}{2a+b} \end{aligned}$$

where, α_1 and α_2 are linear prediction coefficients, and G is a gain constant of the system. Output $y[n]$ is given by the linear combination of input $x[n-1]$, output $y[n-1]$, and $y[n-2]$.

When signal is polluted by noise, it uses the output instead of the input as Equation(4).

$$G = 1 + \alpha_1 + \alpha_2 \tag{4}$$

2.3.2 Prediction in noisy section

Kurakata *et al*[4] conducted a psychophysical experiment about perception of sweep tone followed by noise. As a result, it is thought that auditory system predict a sound using information before noise. Then, frequency and symmetry parameters are predicted by Equation(5) because it is thought that a amount of change are constant in noisy section.

$$y[n] = y[n-1] + (y[n-1] - y[n-2]) \tag{5}$$

On the other hand, it is supposed that the power level of amplitude decrease and the band width of scale increase. Aikawa have reported that if noise is continued a measure of time, peak faded out. Therefore Equation(7) is transformed into Equation(3) as $w = 1$ in noisy section. The amplitude and scale parameter are predicted using Equation (6) and Equation(7) respectively.

$$y[n] = \frac{c}{c+1} y[n-1] \tag{6}$$

$$\text{amplitude} * \text{scale} = \text{const.} \tag{7}$$

2.4 Prediction of some peaks

If some peaks exist instantaneously, the question is whether next peak is chosen from some peaks and next peak is decided by choosing a nearest peak.

Table 1: Condition

Parameter	Value
sampling frequency	20 [kHz]
frame length	25.6 [msec]
shift length	6.4 [msec]
cepstrum order	60
natural frequency	18 [Hz]
dumping factor	1
c	0.0064

2.5 Reconstruction

In Figure 1(D), spectral envelopes are reconstructed by the inverse wavelet transform using predicted four parameters $\hat{r}(\hat{s}_0, \hat{f}_0, t)$.

$$p(\omega, t) = \frac{1}{2\pi C_\phi} \int_{-\infty}^{\infty} \frac{1}{|\hat{s_0}|^{\frac{1}{2}}} \hat{r}(\hat{s_0}, \hat{f}_0, t) \psi(\frac{\omega - \hat{f}_0}{\hat{s_0}}) \frac{d\hat{s_0} d\hat{f}_0}{\hat{s_0}^2} \tag{8}$$

3 Simulation

As simulated data vowels were synthesized by the Klatt formant synthesizer, whose pitch frequency is 140 Hz and the sampling frequency is 20 kHz.

- Synthesized Vowels
  - stationary vowels /a/ and /i/ whose duration is 200-ms.
  - each stationary vowel connected with 100-ms transition.
  - a transition from /a/ to /i/ is polluted by white noise from 50 to 100-ms.
- Simulation condition (Table 1)

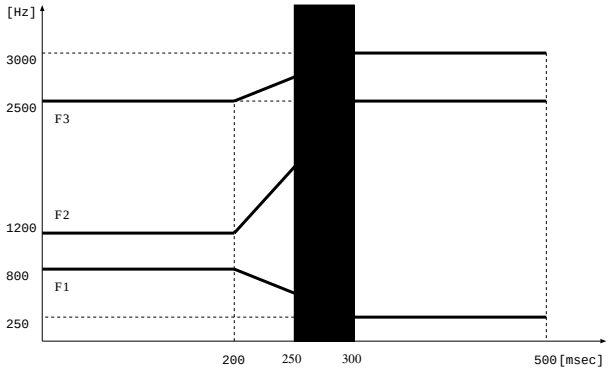

Figure 2: Outline of input spectrum

Transition of synthesized vowel from /a/ to /i/ is interfered by white noise, shown in Figure 2. Figure 4 shows the result of the simulation with noise. Frequency transition is perceived as the formants being extend into the noise by man. The synthesized vowel can be predicted even in noisy section.

4 Conclusion

A sound signal is extracted by predicting and tracking the trajectries of four parameters(frequency, amplitude, bandwidth, and symmetry) using Auditory Cortex 1 model.

References

[1] Aikawa ,Tsuzaki ,Kawahara ,"Dynamic Response of the Sweep Tone Tracking System,The Journal of the Acoustical Society of Japan, H-95-31, 1-8, 1995.

[2] Masuda ,Aikawa ,Tsuzaki ,"A Method for predicting of perceived spectral sequences based on a FM-tracking model,Technical Report of IEICE, vol.96, SP96-2, pp.9-16, 1996.

[3] K. Wang, S. A. Shamma, "Spectral Shape Analysis in the Central Auditory System, IEEE Trans. Speech Audio Processing, vol.3, no.5, pp.382-395, 1995.

[4] Kurakata, Matsui, Nishimura, "Perceptual Trajectory of the Continuity Effect of Frequency Glided Tone. The Journal of the Acoustical Society of Japan, H-94-71, 1994.

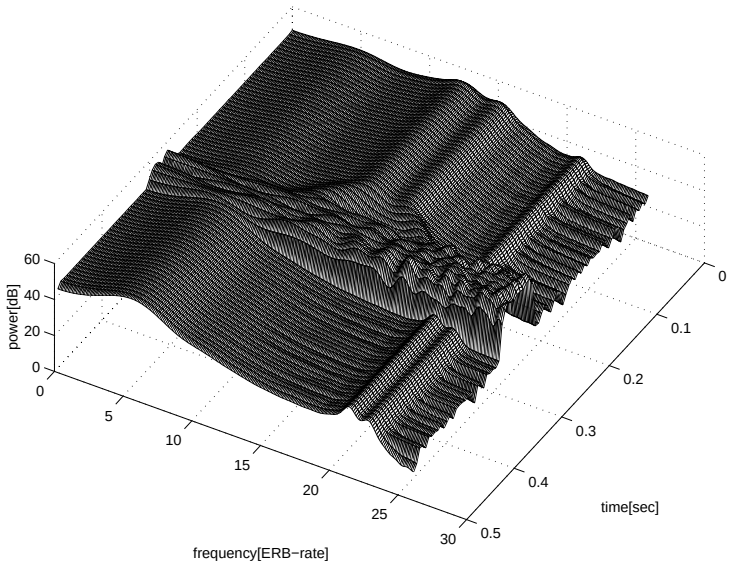

Figure 3: Input spectral sequences

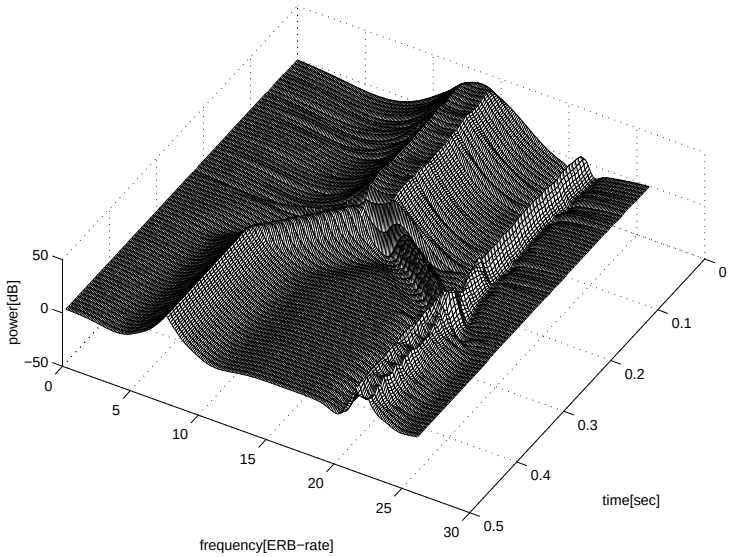

Figure 4: Output spectral sequences