Title	周波数変化音に対する予測追跡の知覚モデルに関する研究
Author(s)	河本，晃一
Citation	
Issue Date	1997-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1017
Rights	
Description	Supervisor:赤木 正人，情報科学研究科，修士

A Perceptual Model of Prediction and Tracking for Frequency-Modulated Tones

Koichi Kawamoto

School of Graduate School of Information Science,
Japan Advanced Institute of Science and Technology, HOKURIKU.,
Japan Advanced Institute of Science and Technology

14, February, 1997

Keywords: pitch perception in sweep-tones, tracking for frequency modulation, spectral representation in the primary auditory cortex.

1 Introduction

If acoustical information is lacked by noise, it is difficult for computers to recognise speech.However human can restore lacked information and communicate each other using speech sounds correctly, such as, the phonemic restoration, the cocktail party effect, etc. This complementation leads to perceive sounds as a stream. The continuity of frequency modulation is one of factors for perceiving a stream. These facts suggest that there exists an auditory function for predicting and tracking frequency transition.

1.1 Perception for Sweep-tones and Phonetic Sounds

Aikawa et al. have reported two factts on pitch perception. The first is that the perceived pitch image of a logarithmic linear frequency glide is not in fact logarithmic linear, but warped. The second is that a sudden change in the sweep slope in a unidirectional Frequency-Modulated(FM) tone induces perceived oscillation.

Kurakata et al. conducted the psychophysical experiment about the perception for sweep tone followed by noise. The experiment showed that the sweep tone is perceived extended into the noise. They have suggested that there is an auditory function for predicting and tracking frequency modulation.

1.2 Spectral Representation in the Primary Auditory Cortex(A1)

Wang and Shamma have proposed a model of the spectral representation in A1. In the model, input spectral patterns are analyzed and represented along three independent dimensions : logarithmic frequency axis, spectral bandwidth axis, and symmetry axis. Since the processing in the model is considered to be multiresolution analysis, each local features in spectral envelope can be treated independently.

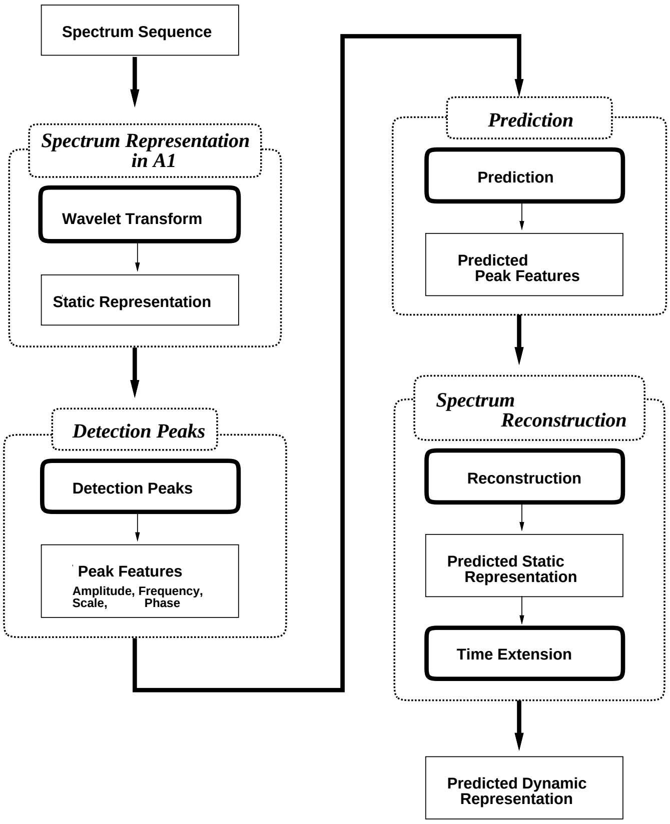


Figure 1: The flow chart of the prediction-tracking model

2 Prediction-Tracking(PT) Model for Frequency Modulation

Figure 1 illustrates the flow of the PT model.

In this model, features of spectral peaks are extracted independently from the static representation in A1. Each features is represented with four components : amplitude, frequency, bandwidth, and symmetry. Temporal transitions of the components are predicted and tracked using the second order system. Then the spectral envelope is reconstructed from predicted features. Moreover by expanding the spectral envelope sequence and summing them up along time axis.

2.1 Input Spectral Envelope

Input for the PT model is the spectrum sequence analyzed by the improved cepstral method. The frequency of spectra is ordered on the ERB rate scale, which has a good relation with physiological and psychological findings.

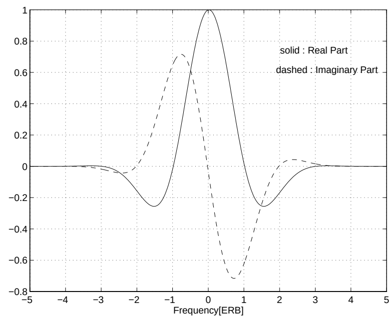


Figure 2: Approximate shape of the spectro response function

2.2 Spectral Representation Model in A1

Shamma reported that the A1 nerve cells severally consist neurons exhibiting similar response characteristic, which are localized around the characteristic frequency(CF), and which has function turning by the bandwidth and symmetry. Considering lateral inhibition, this spectro response characteristics in the PT model are represented by transformed Gabor function, indicated in Figure 2.

A1 cells fully respond, if spectro response characteristics match the input spectral envelope. Mathematically, a cell response is an inner product between the input and a spectro response function. In the PT model, an inner product is defined as wavelet transform, given by Equation(1).

$$\begin{aligned} r(a,b) &= < p(x), \psi(x;a,b) >_x \\ &= \frac{1}{|a|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \psi^*(\frac{x-b}{a}) p(x) dx. \end{aligned} \quad (1)$$

Where a and b denote a scaling parameter and a shifting parameter respectively, and * denotes the complex conjugate. a indicates the position along scales, and b indicates the position along frequencies.

The temporal static response is given by wavelet analysis for an input spectral envelope along the frequency axis using Eq.(1).

When presented with a stationary stimulus at CF, A1 cells usually respond strongly only near the onset of the stimulus. After that, an amplitude and a bandwidth of the response go on decreasing. Assuming this temporal response characteristic to be exponentially decaying, a temporal response function can be described by Equation(2).

$$\lambda(t) = exp(\frac{-t}{\tau}). \quad (2)$$

Where τ denotes time constant.

Moreover, the temporal-spectro response is given by the convolution of a spectro response function and a temporal response function. The response of A1 in any time is obtained by temporal summation of all cell responses.

2.3 Spectral Representation using Peak features

At prediction and tracking process in the PT model, it uses features of spectral peaks that content four parameters. It has be confirmed that a spectral envelope can approximately be reconstructed by using only the peak features. Especially, it can exactly be reconstructed around spectral peaks.

2.4 FM Prediction and Tracking Process

The FM prediction and tracking process is represented by the second order system as Equation(3)

$$
\begin{aligned}
H(z) &= \frac{Gz^{-1}}{1+\alpha_1 z^{-1}+\alpha_2 z^{-2}} \\
\alpha_1 &= -2\exp^{-\zeta\lambda}\cos(\sqrt{1-\zeta^2}\cdot\lambda) \\
\alpha_2 &= \exp^{-2\zeta\lambda} \\
\lambda &= 2\pi\frac{f_n}{f_s} \\
G &= 1+\alpha_1+\alpha_2.
\end{aligned}
\tag{3}
$$

Where α_1 and α_2 denote linear prediction coefficients. G is the gain constant of the system, and f_s denotes the sampling frequency. ζ and f_n denote the damping factor and the natural frequency of the system, respectively.

When a sweep tone is in noise, it uses the output before one step instead of the input before one step, because peak features can not be extracted in the noise.

3 Simulations

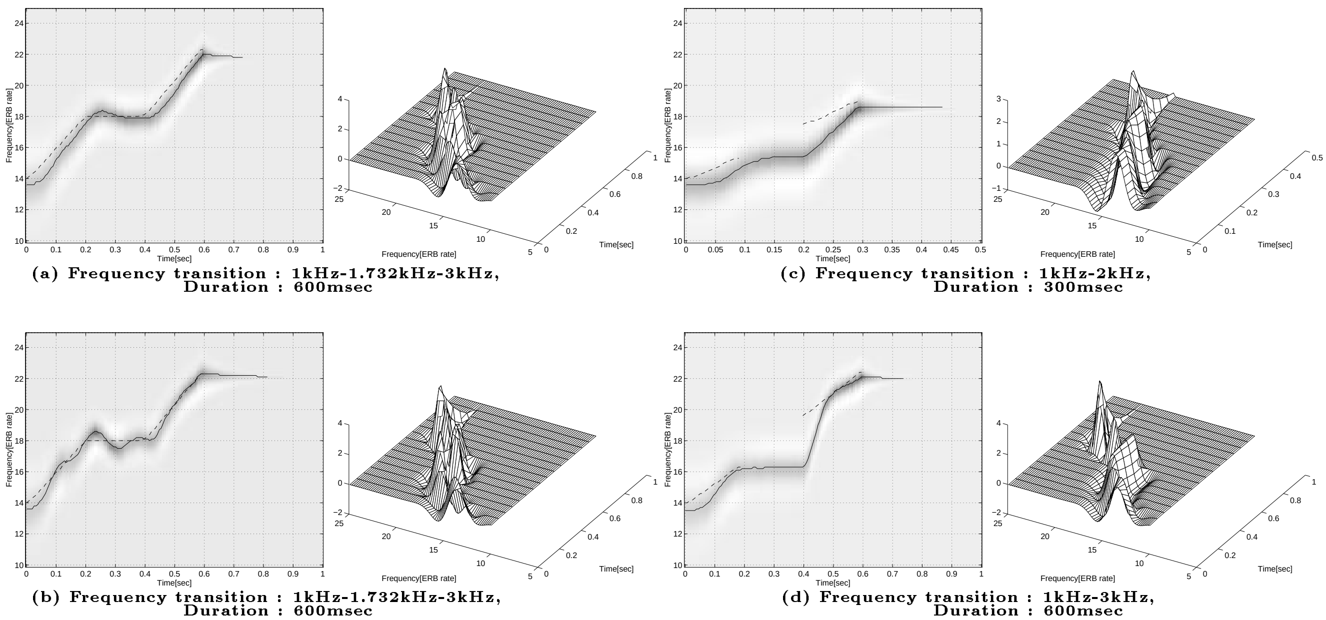
Figure 3 shows predicted trajectories of some sweep tones by using PT model. (a) and (b) in Figure3 are results of the simulation for step sweep tones that include a sudden change. (c) and (d) are that for the interrupted sweep tones that include an interruption by noise of which duration are 100msec in case of (c), and 200msec in case of (d). Setting parameters are described in each figure.

At (a) and (b) in Figure 3, pitch oscillation occur around a part of changing slope suddenly. This disposition is similar to the result of the psychological physical experiment by Aikawa et al.. Moreover the degree and the number of repetition of pitch oscillation can be controlled by changing the natural frequency and the damping factor.

At (c) and (d), if a duration of the interruption by noise is short such as (c), the frequency transition is perceived as the sweep being extended into the noise, but if a interrupted duration is long such as (d), the frequency transition can not be perceived in the noise. Kurakata et al. have reported that the limit of duration for the extended perception is about 100msec. Because the PT model uses features only before interruption by noise, above disposition seems to be similar to the result of the psychological physical experiment by Kurakata et al..

4 Conclusion

In this study, the model is proposed to predict and to track the frequency modulation. This model uses the spectral representation in A1. Accordingly, a spectral peak is represented by



(a) Frequency transition : 1kHz-1.732kHz-3kHz, Duration : 600msec

(c) Frequency transition : 1kHz-2kHz, Duration : 300msec

(b) Frequency transition : 1kHz-1.732kHz-3kHz, Duration : 600msec

(d) Frequency transition : 1kHz-3kHz, Duration : 600msec

Step sweep tones

Interrupted sweep tones

Figure 3: Results of Prediction Tracking Simulation

features that contents four parameters : amplitude, frequency, bandwidth, symmetry. A spectral envelope can approximately be reconstructed by using only peak features. From this view point, features of the spectrum is compress owing to using the spectral representation in A1.

In the PT model, prediction and tracking process is represented by the second order system(Eq.(3)), and it treats features of spectral peaks that extracted from the spectral representation.

The PT model was evaluated by simulation experiments using some kind of sweep tones. The results of experiments were similar to the perceptions that is found by the psychological physical experiment for those sweep tones. Therefore, those perceptions can be represented on signal processing level because of the PT model uses.

References

[1] 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.