



**Acoustics'08
Paris**
June 29-July 4, 2008
www.acoustics08-paris.org

Local-pitch identification accuracy depending on the trajectory of frequency-modulated tones

Yuki Hiruma and Kiyooki Aikawa

School of Media Science, Tokyo Univ. of Technology, 1404-1 Katakuracho, Hachioji, 192-0982
Tokyo, Japan
aik@media.teu.ac.jp

Local-pitch identification accuracies were analyzed for frequency-modulated (FM) tones. The problem was whether every portion of a continuous sound was perceived at the same accuracy or not. Psychophysical experiments revealed that the local-pitch identification accuracies were significantly different among the nodes of continuous FM tones. Also, the accuracies were dependent on the frequency trajectory shape. The stimuli were two types of piecewise-linear FM tones of up-down-up and down-up-down glide sequences. Each tone included four nodes; the initial, two intermediate points, and the final. The duration of each linear glide was 100 ms. The frequency range was between 1000 and 1500 Hz. A pair of FM tone was presented with one-second interval. The frequency was shifted up or down at one of the nodes in either of the tones. The shift amounts were 0%, 4%, and 8%. The subjects were requested to answer whether two pitch sequences were the same or different. The pitch identification accuracy was low at the initial for both types of FM tones. The accuracy at the final was highest for the up-down-up tone. The intermediate high frequency node showed the highest accuracy for the down-up-down tone. These results indicated that the local-pitch identification accuracies were trajectory-dependent.

1 Introduction

Several relevant findings have been reported on the perception of Frequency-Modulated (FM) tones. One of the authors found that perceived pitch contours were different from the frequency trajectory of the stimulus FM tones. The most important finding was that a linear sweep tone including a constant frequency portion in the middle evoked pitch ringing which was not included in the stimulus tone [1]. Also, short linear sweep tones were perceived as accelerating pitch contours [2]. A mathematical model was proposed for simulating these phenomena using a second order autoregressive filter [2,3,4]. Since the model performed as if the model chased the frequency trajectory of the stimulus tone, the model was called the FM-tone tracking model. Several typical perceptual illusions such as the bounce illusion were successfully simulated by a neural network model based on the FM-tone tracking model [5,6]. Shihara reported that the perceived local pitch for a sweep tone did not match the actual frequency excepting around the middle of the sweep range [7]. Kigure found that the pitch perception accuracy was different in the sequence of music-like discrete tones depending on the local frequency sequence [8].

The pitch tracking model [2,3] was based on the assumption that pitch was perceived with uniform accuracy along the continuous frequency trajectory. A detailed analysis on pitch perception accuracy was needed for further improving the tracking model. Preliminary experiments were conducted for analyzing the pitch perception accuracy for continuous FM tones [9]

This paper describes the local pitch perception accuracy focusing on the frequency trajectory pattern of the continuous FM tones. Chapter 2 analyzes the temporal position-dependency and FM trajectory-dependency on the perception of local pitch. Chapter 3 examines the side-effect of frequency changing rate accompanied by the change of local frequency of the FM tones. Chapter 4 concludes this study.

2 Local pitch perception experiment

2.1 Stimulus tones

The stimuli were piecewise linear FM tones which were concatenated linear sweeps of 100 ms duration. Fig. 1 shows the frequency trajectory of the standard stimulus tones. The Pattern A is the concatenation of rising, falling, and rising linear sweeps. Pattern B is the upside-down of the pattern A. The duration of the stimuli were 300 ms. A stimulus tone included four nodes, the initial node S, two joint nodes P1, P2, and final node E. The nodes S and P2 were at the same frequency. The nodes P1 and E were at the same frequency. Listeners compared pairs of standard and comparison stimuli. The local pitch identification accuracy was measured by whether a small amount of frequency shift given at one of the four nodes of the comparison stimulus could be detected or not. The shift amount was 0%, 4% and 8%. The comparison stimulus of 0% shift was the control stimulus for monitoring the correctness of the answers by subjects for the identical stimulus pair. The frequency range was decided to 500 Hz according to the preliminary experiments. Sampling frequency was 44.1 kHz. The rise and fall times were 5 ms each. Stimulus tones were created by MATLAB. The phase was continuous at each joint.

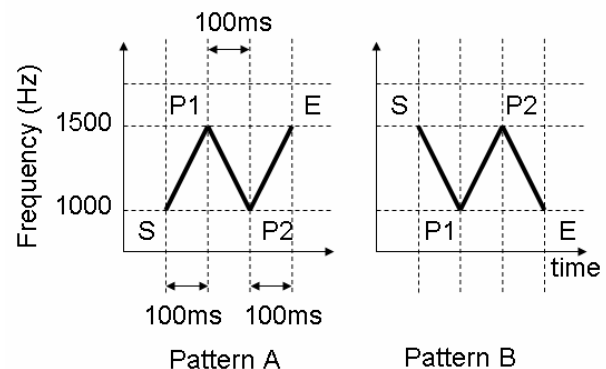


Fig. 1 Two types of frequency-modulated (FM) tones used for standard stimuli. Pattern B is the upside-down of the Pattern A. An FM tone is composed of three linear sweeps concatenated at joint nodes P1 and P2. S and E are the initial and final nodes, respectively.

2.2 Experimental conditions

A pair of standard and comparison stimuli was presented with a 1000 ms interval between them as shown in Fig.2. Subjects were forced to choose "same" or "different" for each stimulus pair (2AFC). Stimuli were diotically presented by a electro-static headphones (STAX SR-404 Signature) at 65 dB(SPL) in a sound-proof chamber. Subjects were 10 normal hearing students (eight males and two females). A session included 66 stimulus pairs. Although the standard stimulus is drawn at the left hand side in Fig. 2, a session includes the same numbers of stimulus pairs of the reverse order for canceling time order effect. A Subject listened to two sessions including a short intermission between them. The presentation order of the stimulus pairs was randomized for each session. A session included control stimulus of identical tone pairs.

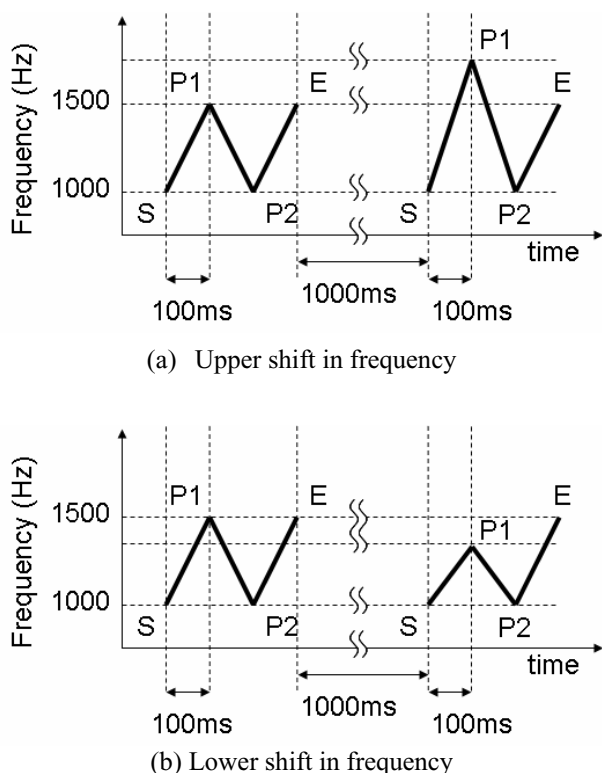


Fig. 2 A stimulus sequence. A standard and comparison FM tones are presented with 1000 ms interval. The frequency of a node in the comparison stimulus is shifted. A session of stimuli includes a couple of stimulus pair of the same shift condition, the standard stimulus first and comparison stimulus first.

2.3 Results for Pattern A

The frequency-shift detection rate was obtained as the ratio of the number of answers "different" to the number of stimulus pairs. The control stimuli were excluded for obtaining the detection rate.

Figure 3 shows the frequency-shift detection rates for FM pattern A. Fig. 3 (a) and (b) show the results when the frequency of one of the four nodes was shifted 4% upper and 4% lower, respectively. The detection rate was high at the final node (E) and the second node (P1). The detection

rate was low at the initial (S) and the third node (P2). The detection rates were significantly different among four nodes ($p < 0.0008$).

Almost the same results were obtained when the frequency shift was 8%. The detection rates were higher for the case of 8% shift than that of 4% shift.

2.4 Results for Pattern B

Figure 4 shows the frequency-shift detection rates for the FM Pattern B. Fig. 4 (a) and (b) show the results when the frequency of one of the four nodes was shifted 4% upper and 4% lower, respectively. Fig. 4 (c) shows the results when the frequency of one of the four nodes was shifted 8% upper. The frequency-shift detection rate was high at the third node (P2). The detection rate was low at the other three nodes. The detection rates were significantly different among four nodes ($p < 0.0006$).

Almost the same results were obtained when the frequency shift was 8% lower. The detection rates were higher for the case of 8% shift than that of 4% shift.

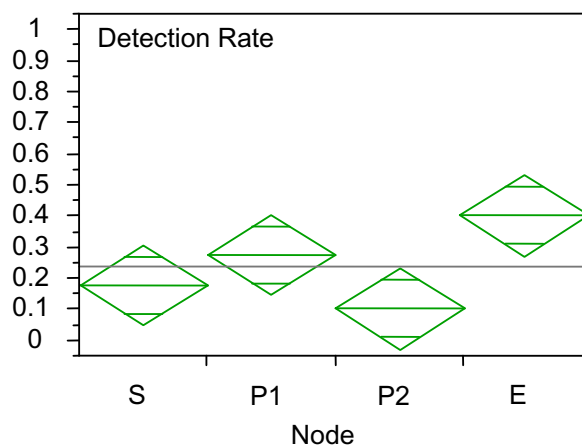
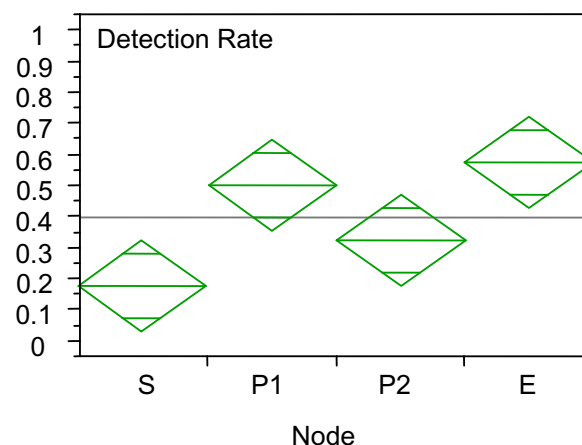
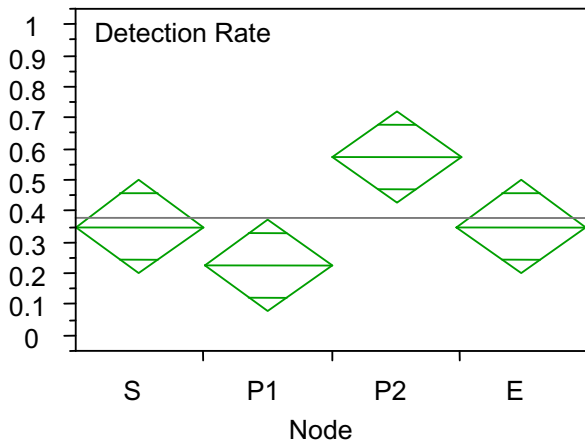
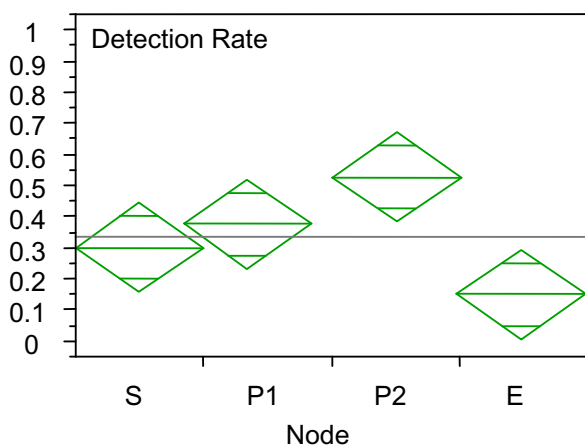


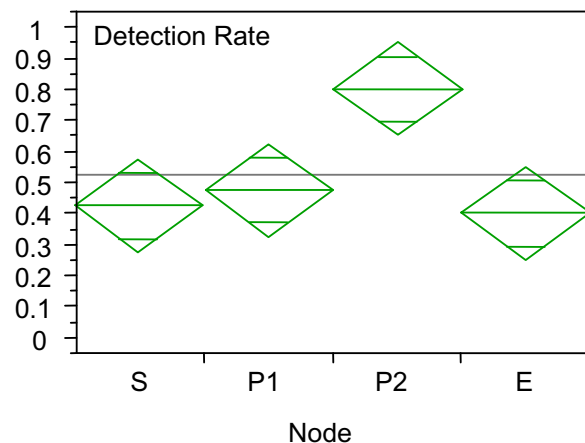
Fig. 3 Frequency-shift detection rates for FM pattern A.



(a) 4% upper shift in frequency



(b) 4% lower shift in frequency



(c) 8% upper shift in frequency

Fig. 4 Frequency-shift detection rates for FM pattern B.

2.5 Discussions

The frequency trajectory of the FM pattern B was the upside-down of the pattern A. If the frequency-shift detection rate depended on the frequency, the nodes S and P2 should show high identification rate. However, the node S showed low detection rate for both FM Pattern A and B. These results indicated that the accuracy did not simply depend on the frequency of the node but also the FM trajectory itself.

The common result for the Pattern A and B was that the local pitch identification accuracy was significantly different among four nodes of piecewise linear FM tones. Another common result was that the pitch identification accuracy was low at the initial (S) of the FM tones. The frequency-shift detection rates were not the same between the corresponding nodes of Pattern A and B (i.e. S vs. E, P1 vs. P2). The difference could not be simply explained only by the frequency of the node. It was considered that the difference was dependent on the temporal position and trajectory pattern of the FM tones.

3 Slope change side-effect

3.1 Problem

The previous experiment used stimulus pair depicted in Fig. 2. A problem arose that the node frequency shift resulted in the change of the slope of the linear sweeps in the stimulus tones. This chapter examines the side-effect of frequency slope change on the analysis of frequency trajectory-dependent pitch identification accuracy.

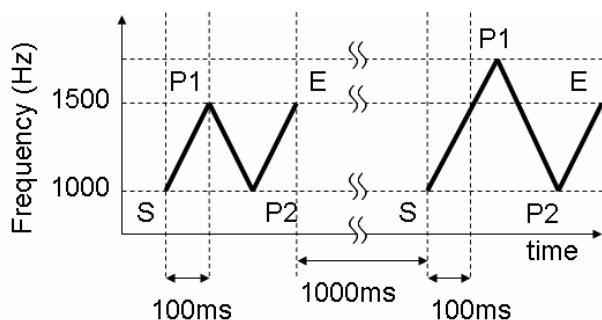
3.2 Stimulus tones

The standard frequency trajectories were the same as the previous experiment. Two types of comparison stimuli were compared. One was the stimulus preserving the frequency change rate (R) shown in Fig. 5 (a). When the frequency at the node P1 was shifted upper, the duration between nodes S and P1 was lengthened in proportion to the frequency shift amount. The other was the stimulus preserving the duration time of each linear portions of the piecewise linear FM tones (T) shown in Fig. 5 (b). This was identical to the stimulus used in the previous experiments. The sampling frequency, rise and fall times were the same as the previous experiment.

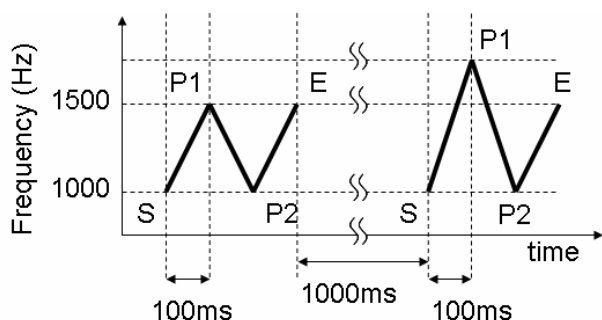
3.3 Experimental conditions

The experimental setup was the same as the previous experiment. The standard and comparison stimuli were separated by 1000 ms interval. The frequency shift amount was 0% and 4%. A session included 66 pairs of stimuli. The presentation order was randomized for each subject. Subjects were 10 normal hearing students (six males and four females). A subject listened two sessions with a short intermission. A session included control stimulus pairs composed of two identical stimuli (frequency shift amount

of 0%). The control stimuli were used for the detection of incorrect answers.



(a) Constant-rate condition "R"



(b) Constant-time condition "T"

Fig. 5 Difference between constant-rate and constant-time stimulus conditions.

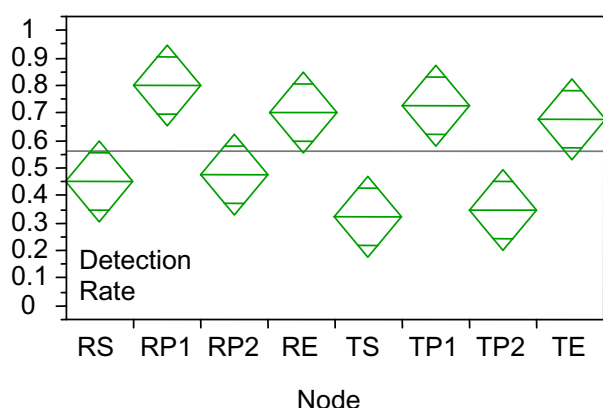


Fig. 6 Comparison of frequency-shift detection rates between constant-rate "R" and constant-time "T" conditions for Pattern A ($p < 0.0001$). The frequency-shift was 4% upper.

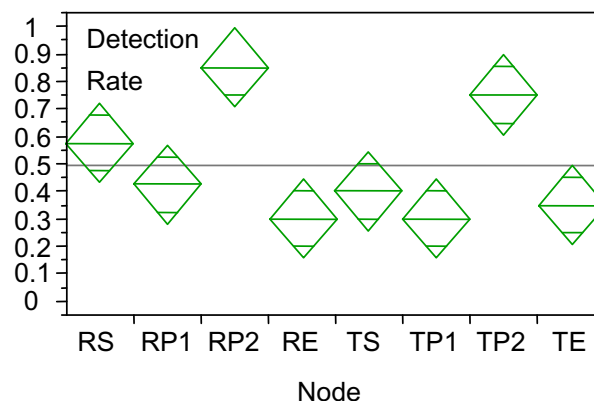


Fig. 7 Comparison of frequency-shift detection rates between constant-rate "R" and constant-time "T" conditions for Pattern B ($p < 0.0001$). The frequency-shift was 4% upper.

3.4 Results for Pattern A

Figure 6 shows the comparison of frequency-shift detection rates between constant-rate and constant-time conditions for Pattern A in case that the frequency-shift is 4% upper. The symbol "R" and "T" denote constant-rate and constant-time conditions, respectively. Fig. 6 indicates that almost same results were obtained for the constant-rate and constant-time conditions. The detection rate was high for the nodes P1 and E. The rates were low for the nodes S and P2. The difference of detection rates between corresponding nodes of constant-rate and constant-time conditions (e.g. RS vs. TS) was not significant for each of four nodes. Almost the same results were obtained in case that the frequency shift was 4% lower.

3.5 Results for Pattern B

Figure 7 shows the comparison of frequency-shift detection rates between constant-rate and constant-time conditions for Pattern B in case that the frequency-shift is 4% upper. Fig. 7 indicates that almost same results were obtained for the constant-rate and constant-time conditions. The detection rate was high for the node P2. The rates were low for the other three nodes. The difference of detection rates between corresponding nodes of constant-rate and constant-time conditions (e.g. RS vs. TS) was not significant for each of four nodes. Almost same results were obtained in case that the frequency shift was 4% lower.

3.6 Discussion

The experimental results showed that the frequency-shift detection rate was significantly different among four nodes in the piecewise linear FM tones. The significant difference was obtained for both FM pattern A and B. A little bit high frequency-shift detection rates were obtained for the constant-rate condition compared to the constant-time

condition, however, the detection rates were not significantly different between the corresponding nodes of constant-rate and constant-time conditions (e.g. RS vs. TS).

The experimental results indicated that the side-effects of frequency slope change were small enough on analyzing the pitch identification rates, if the duration of a linear sweep portion was around 100 ms and the frequency change was less than 500 Hz.

4 Conclusions

This paper described a new finding on the local pitch identification for continuous frequency-modulated tones. The stimuli were concatenated linear sweeps of up-down-up and down-up-down sequences. The pitch identification accuracy was analyzed using the frequency-shift detection rate for the initial, two intermediate, and the final nodes in the FM tone trajectories. The frequency-shift detection rate was different depending on the temporal position of the node in the frequency trajectory of the FM tones. The rate also depended on the trajectory pattern of the stimulus tones. The pitch identification accuracy was always low for the initial of the continuous FM tone regardless of the node frequency. The rate was high for the other high frequency nodes in the FM tones.

This paper also examined a problem that the node frequency shift was accompanied by the slope change side-effect on analyzing the detection rates of the frequency-shift at each node. The experimental results indicated that the side-effect of the change of slope was small enough.

References

- [1] Kiyooki Aikawa, Minoru Tsuzaki, Hideki Kawahara, Yoh'ichi Tohkura, "Pitch ringing induced by frequency-modulated tones", *J. Acoust. Soc. Am.*, Vol.98, no.5, Pt.2, p.2926, (1995-11).
- [2] Kiyooki Aikawa, Minoru Tsuzaki, Hideki Kawahara, Yoh'ichi Tohkura, "On the dynamic characteristics of the tracking of frequency-modulated tones", *J. Acoust. Soc. of Japan*, Vol.52, no.10, pp.741-751, (1996-10) (in Japanese).
- [3] Kiyooki Aikawa, Minoru Tsuzaki, Hideki Kawahara, "Dynamic perception of frequency-modulated tones", *J. Acoust. Soc. Am.*, Vol.100, no.4 Pt.2, p.2750, (1996-12).
- [4] Kiyooki Aikawa, Minoru Tsuzaki, Hideki Kawahara, "Dynamic perception of frequency-modulated tones", Proceedings of ASA and ASJ third joint meeting, pp.691-694, (1996-12).
- [5] Kiyooki Aikawa, Hideki Kawahara, Minoru Tsuzaki, "A neural matrix model for active tracking of frequency-modulated tones", *ICSLP96*, Vol.1, pp.578-581, (1996-10).
- [6] Kiyooki Aikawa, Hideki Kawahara, "A neural computational model for tracking of multiple frequency-modulated tones", *J. Acoust. Soc. Am.*, Vol.99, no.4, Pt.2, p.2490, (1996-05).
- [7] Tomonori Shihara, Hiromitsu Miyazono, Tsuyoshi Usagawa, Masanao Ebata, "Dynamic properties of pitch perception of sweep tone", *J. Acoust. Soc. Am.*, Vol.100, no.4, p.2752, (1996-10).
- [8] Yuka Kigure, Kiyooki Aikawa, "Pitch identification accuracy depending on the note location in a melody", *2007 meeting of Acoust. Soc. of Japan*, pp.391-392, (2007-03) (in Japanese).
- [9] Yuki Hiruma, Kiyooki Aikawa, "Pitch identification accuracy depending on the position in continuous frequency-modulated tones", *2007 meeting of Acoust. Soc. of Japan*, pp.577-578, (2007-09) (in Japanese).