

Perceptual learning of frequency discrimination for tones with low fundamental frequency: Learning for high but not for low harmonics

Hiromitsu Miyazono^{1,*} and Brian C. J. Moore^{2,†}

¹Faculty of Administration, Prefectural University of Kumamoto, 3-1-100 Tsukide, Kumamoto, 862-8502 Japan

²Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 2EB, UK

(Received 16 February 2009, Accepted for publication 21 April 2009)

Keywords: Perceptual learning, Pitch, Dominant region, Random phase, Cosine phase

PACS number: 43.66.Hg, 43.66.Nm [doi:10.1250/ast.30.383]

1. Introduction

To determine which harmonics have the greatest effect on (dominate) the perceived pitch of a complex tone, the perception or detection of a shift in the frequencies of a group of harmonics, or a single harmonic, has been measured, leaving the other harmonics fixed in frequency. For fundamental frequencies (F0s) in the range 100 to 400 Hz, the lower harmonics (1–5) usually dominate the pitch percept.

Miyazono *et al.* [1] studied the dominant region for tones with very low F0. They measured the threshold for detecting a change in the F0 of a group of harmonics (called Group B) within a complex tone, while the other harmonics were fixed in frequency (see Fig. 1). The threshold is called the F0 discrimination limen (FODL). The components in the complex tones started either with cosine or with random phase. The relative phase of the components is assumed not to affect pitch perception when the tone contains resolved components, but can affect pitch perception when the components are all unresolved [2]. Miyazono *et al.* [1] showed that the position of the dominant region depended on the relative phases of the components. When the envelope had a low peak factor, with multiple peaks per period (random phase), the dominant region fell at low harmonic numbers. When the envelope had a high peak factor with one peak per period (cosine phase), the dominant region fell at high harmonic numbers. The temporal information derived from unresolved harmonics in the cosine-phase stimuli led to lower FODLs than the (place or temporal) information derived from the resolved harmonics in the random-phase stimuli.

These and other findings have led to the suggestion that there are two mechanisms for pitch perception, one based on low resolved harmonics and one based on high unresolved harmonics [3]. However, others have proposed that pitch may be extracted in a similar way for high and low harmonics [4]. One way of addressing this issue is by studying perceptual learning, which is the widely-observed phenomenon that subjects tend to improve with practice at basic perceptual tasks. Grimault *et al.* [5] showed that training in pure-tone

frequency discrimination resulted in significantly larger improvements in FODLs when the test complexes contained resolved harmonics than when they contained only unresolved harmonics. This was interpreted as supporting the hypothesis that the discrimination of the F0 of tones with resolved harmonics shares common underlying mechanisms with the frequency discrimination of pure tones. However, training in amplitude-modulation rate discrimination did not result in larger improvements in FODLs for tones with only unresolved harmonics than for tones with resolved harmonics, contrary to what would be expected if the FODLs for tones with only unresolved harmonics depended mainly on the use of temporal envelope cues.

In this study, we measured perceptual learning of F0 discrimination at a low (50 Hz) F0, using stimuli similar to those of Miyazono *et al.* [1]. To determine whether there are distinct perceptual mechanisms for low and high harmonics, we assessed whether perceptual learning occurred at a similar rate when there were either low or high harmonics in Group B, and we assessed whether training with one type of stimulus would result in improved performance for the other type of stimulus, i.e. whether there was transfer of learning. After initial testing using both random- and cosine-phase tones, and with a variety of harmonics in Group B, subjects were split into two groups. One group was trained using random-phase stimuli with harmonics 1–5 in Group B. The other group was trained using cosine-phase stimuli with harmonics 26–30 in Group B. Performance was compared for the two groups, both during training and using post-training tests with the same stimuli as used in the pre-training tests.

2. Method

2.1. Stimuli

In each trial, subjects were presented with two successive tones, each with rise/fall times of 20 ms and a steady state duration of 500 ms. The silent interval between the two tones was 300 ms. The tones contained 57 components and had an overall level of 70 dB SPL. To avoid the perception of an edge pitch, the level of components above the 30th was progressively reduced, by 2 dB per component. The components were divided into three groups, as shown in Fig. 1: 1 to $J - 2$

*e-mail: miyazono@pu-kumamoto.ac.jp

†e-mail: bcmj@cam.ac.uk

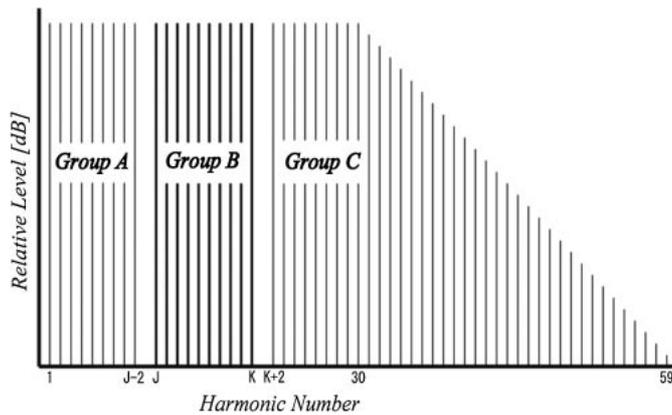


Fig. 1 Schematic spectrum of stimulus showing components in Groups A, B and C.

(Group A), J to K (Group B) and $K + 2$ to 59 (Group C). FODLs were measured for the components in Group B. These had an F_0 of $50 - \Delta F$ Hz for one tone in a trial and $50 + \Delta F$ Hz for the other. The components in Groups A and C were all harmonics of 50 Hz. The components with numbers $J - 1$ and $K + 1$ were omitted to prevent overlap between harmonics in Group B and Groups A or C. The order of the two tones was selected randomly for each trial. The value of J (corresponding to the lowest component in Group B) was 1 or 26 in the training sessions and 1, 6 or 26 in the pre- and post-training sessions. The value of K (corresponding to the highest component in Group B) was 5 or 30 in the training sessions and 5, 25 or 30 in the pre- and post-training sessions. For $J = 1$, there was no component in Group A.

In one condition, all components started in cosine phase, providing highly effective temporal information when the components were unresolved. In the other condition, the components started in random phase, which leads to a waveform with less distinct envelope peaks; the random phases varied across runs, but not within a run.

2.2. Subjects

Ten subjects were tested. Their ages ranged from 20 to 27 years. They were not musically trained. All subjects had absolute thresholds better than 20 dB HL over the range of audiometric frequencies from 250 to 8,000 Hz.

2.3. Procedure

FODLs for Group B were measured using a 2-interval, 2-alternative forced-choice procedure with feedback. Subjects were asked to indicate which of the two tones in a trial they perceived to have the higher pitch. A three-down, one-up adaptive procedure was used to estimate the 79% correct point on the psychometric function. The value of ΔF was changed by a factor of $2^{0.5}$ until four turnpoints had occurred, and was changed by a factor of $2^{0.5}$ thereafter. Twelve turnpoints were obtained and the geometric mean of the values of ΔF at the last eight turnpoints was used to estimate the FODL.

Each subject was tested over a period of 14 days, 4 for measurement of pre-training thresholds, 6 for training, and 4 for measurement of post-training thresholds. There were 2 days between each testing day. In the pre-training and post-training sessions, all ten subjects were tested in same conditions, as indicated in Table 1. In each of these sessions,

Table 1 Summary of the conditions used for pre-training, training, and post-training, showing the components in Group B and phase condition.

Session	The lowest (J) and highest (K) component numbers of Group B: (J, K), and phase of components			
Pre-Post	(1, 5)	(1, 25)	(6, 30)	(26, 30)
	Random/Cosine phase			
Training	(1, 5) Random phase		(26, 30) Cosine phase	

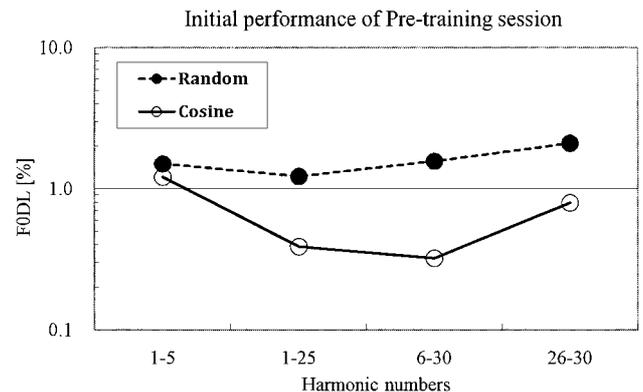


Fig. 2 Mean pre-training results across all subjects. The filled and open circles indicate FODLs for the random-phase and cosine-phase stimuli, respectively. The horizontal axis indicates the numbers of the harmonics in Group B.

the FODL was estimated at least 3 times for each condition. Subjects were divided into two groups of five subjects for the training sessions: Resolved-Random (RR) and Unresolved-Cosine (UC). Subject group RR was trained using random-phase tones with $J = 1$ and $K = 5$, i.e. Group B contained at least some resolved harmonics. This condition was chosen so as to promote performance of the task based on the use of resolved harmonics; temporal envelope information from unresolved harmonics was minimal because of the low numbers of the harmonics in Group B and because of the random phase. Subject group UC was trained using cosine-phase tones with $J = 26$ and $K = 30$, i.e., Group B contained unresolved components. This condition was chosen so as to promote performance of the task based on the use of temporal envelope information. On each training day, there were two sessions. Three threshold estimates were obtained in each session, giving six estimates per day and 36 estimates in total for the training.

3. Results and discussion

Figure 2 shows geometric mean FODLs, expressed as a percentage, for the pre-training sessions. The horizontal axis shows the numbers of the harmonics in Group B. The filled and open circles show FODLs for random- and cosine-phase stimuli, respectively. When Group B contained harmonics 1–5, FODLs were similar for random- and cosine-phase stimuli, which is consistent with the expectation that the lower

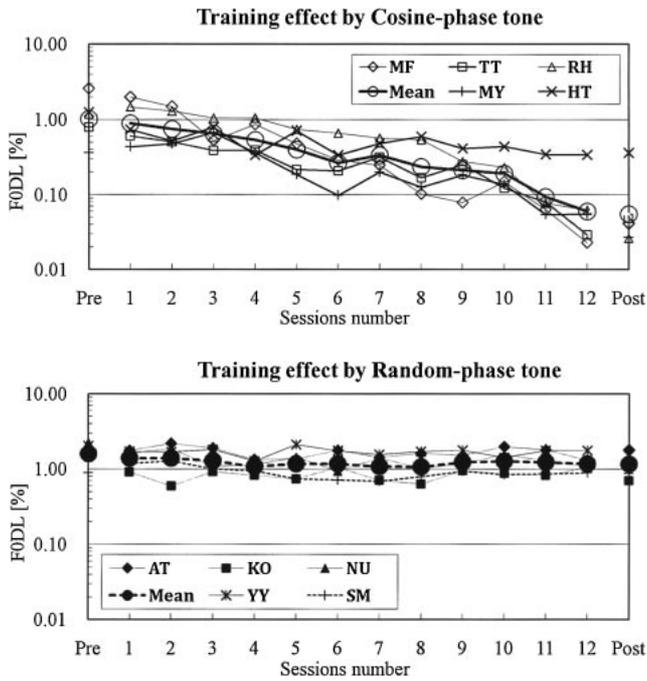


Fig. 3 Individual and mean (large symbols) results for pre-training (extreme left), the training sessions, and post-training (extreme right). The upper panel shows FODLs for subjects trained using cosine-phase stimuli with harmonics 26–30 in Group B (Subject group UC). The lower panel shows FODLs for subjects trained using random-phase stimuli with harmonics 1–5 in Group B (Subject group RR).

harmonics were at least partially resolved. However, when Group B contained harmonics above the 5th, FODLs were lower for cosine- than for random-phase stimuli. This is consistent with the idea that F0 discrimination can be based on the temporal structure of stimuli with unresolved harmonics, and that the greater peak factor of the envelope for the cosine-phase stimuli led to better performance for those stimuli. For the cosine-phase stimuli, performance was best when Group B contained many harmonics (1–25 and 6–30). For these conditions, temporal information would have been available to the auditory system over a wide range of centre frequencies. For the random-phase stimuli, performance was best when low harmonics were present (1–5 and 1–25).

Figure 3 shows individual (small symbols) and mean (large symbols) FODLs across training sessions. For Subject group UC (upper panel), performance improved steadily across training sessions, for all subjects. The mean FODL after training was very small ($< 0.1\%$), and was more than a factor of 10 lower than the pre-training FODL of about 1%. In contrast, for Subject group RR (lower panel), there was no clear effect of training. The mean FODL after training was only very slightly lower than the mean FODL before training; both FODLs were a little above 1%. The improvement in performance with training for Subject group UC cannot be attributed to learning of the procedure, since if this were the case, both groups would have shown an effect of training. The failure to find any effect of procedural learning may have resulted from the relatively long pre-training used here.

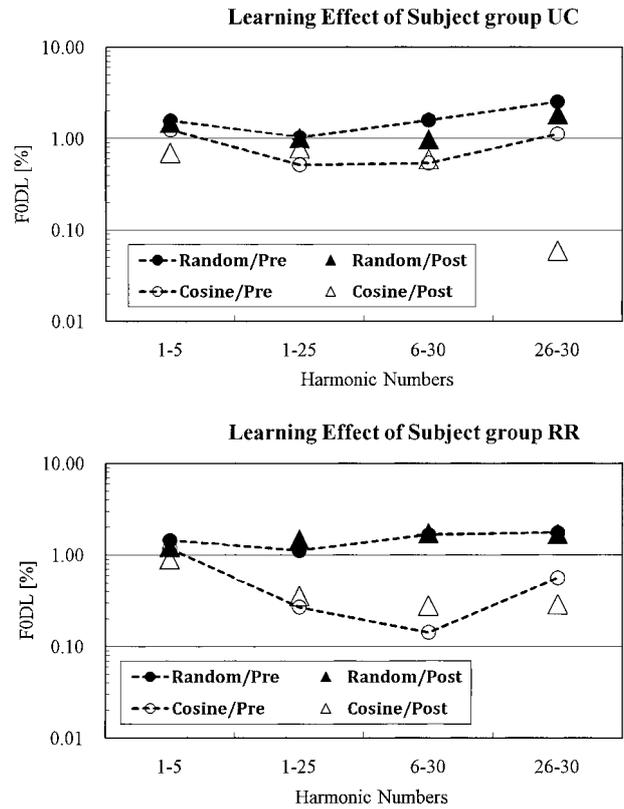


Fig. 4 Mean pre- and post-training results averaged across 5 subjects for each training condition. The upper panel shows FODLs for Subject group UC. The lower panel shows FODLs for Subject group RR.

Figure 4 compares the FODLs obtained pre- and post-training. The layout is similar to that for Fig. 2. For Subject group UC (upper panel), there was a large difference in FODLs between pre- and post-training for the condition on which the group had been trained (compare the right-most open circle and open triangle). However, the improvement did not generalize to the other conditions; all other differences between pre- and post-training FODLs were small. For Subject group RR (lower panel), there was no substantial difference in pre- and post-training FODLs for any condition.

The very small post-training FODLs obtained for Subject group UC for the condition on which they had been trained (cosine-phase, harmonics 26–30) appear to be mysterious at first sight. The FODLs are smaller than have been measured in other studies using a small group of very high harmonics without any fixed “background” harmonics [3,6]. This leads us to suspect that the very small FODLs found here may be the result of subjects using a cue that does not directly involve F0 discrimination of the harmonics in Group B; rather the cue used may depend on an interaction or comparison of the harmonics in Group B with those in Groups A and C.

One possible cue is “pitch-pulse asynchrony” (PPA) [7]. Subjects may compare the timing of envelope peaks across different auditory filters. This would apply only in the cosine-phase condition. Consider, for example, an auditory filter centred in Group A, on harmonic #20. For the cosine-phase stimuli, this would produce envelope peaks every 20 ms. For an auditory filter centred in Group B, on harmonic #28, the

envelope peaks would initially be synchronized to those of Group A. However, in the interval where the F0 was shifted upwards, the period would be shorter, and towards the end of the stimulus the envelope peaks at the output of the filter centred in Group B would occur earlier in time than those for the filter centred in Group A. In the interval where the F0 was shifted downwards, a PPA in the opposite direction would occur. Thus, there would be a PPA across auditory filters, which would differ for the two intervals of a trial.

The PPA cue is subtle, and the strong training effect found for Subject group UC probably reflected learning to use this cue. Further experiments are required to assess whether subjects did use the PPA cue and to assess the effects of training when the use of the cue is disrupted, for example by the introduction of a random temporal offset between the envelope peaks in Group A and the envelope peaks in Groups A and C.

Because of the likely influence of the PPA cue on the results for the cosine-phase stimuli, we cannot use the results to make inferences about whether there are two pitch mechanisms, one for resolved harmonics and one for unresolved harmonics. If a training effect were found for cosine-phase stimuli with high harmonics in Group B when the PPA cue was disrupted, but there were no learning effect for random-phase stimuli with low harmonics in Group B (as found in the present experiment), this would support the idea of two pitch mechanisms.

4. Conclusions

We assessed the effect of training on performance of a task which required subjects to discriminate the F0 of a group of harmonics (Group B) embedded in other harmonics whose frequencies were fixed (Groups A and C). A low F0 of 50 Hz was used. Subjects trained using cosine-phase complex tones

with harmonics 26–30 in Group B (Subject group UC) showed large improvements in training, which did not generalize to the other conditions tested. Subjects trained using random-phase complex tones with harmonics 1–5 in Group B (Subject group RR) did not show any clear improvement with training. The large training effect for Subject group UC may reflect learning to use a cue based on the detection of PPA across different centre frequencies.

References

- [1] H. Miyazono, B. R. Glasberg and B. C. J. Moore, "Dominant region for pitch at low fundamental frequencies (F0): the effect of fundamental frequency, phase and temporal structure," *Acoust. Sci. & Tech.*, **30**, 161–169 (2009).
- [2] B. C. J. Moore, "Effects of relative phase of the components on the pitch of three-component complex tones," *Psychophysics and Physiology of Hearing*, E. F. Evans and J. P. Wilson, Eds. (Academic Press, London, 1977), pp. 349–358.
- [3] T. M. Shackleton and R. P. Carlyon, "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," *J. Acoust. Soc. Am.*, **95**, 3529–3540 (1994).
- [4] R. Meddis and L. O'Mard, "A unitary model of pitch perception," *J. Acoust. Soc. Am.*, **102**, 1811–1820 (1997).
- [5] N. Grimault, C. Micheyl, R. P. Carlyon, S. P. Bacon and L. Collet, "Learning in discrimination of frequency or modulation rate: generalization to fundamental frequency discrimination," *Hear. Res.*, **184**, 41–50 (2003).
- [6] B. C. J. Moore, B. R. Glasberg, H. J. Flanagan and J. Adams, "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," *J. Acoust. Soc. Am.*, **119**, 480–490 (2006).
- [7] H. Gockel, R. P. Carlyon and B. C. J. Moore, "Pitch discrimination interference: The role of pitch pulse asynchrony," *J. Acoust. Soc. Am.*, **117**, 3860–3866 (2005).