EFFECT OF PARALLEL/OPPOSITE FREQUENCY MOVEMENTS ON PERCEPTUAL INTEGRATION OF INHARMONIC COMPONENTS

Ryota Miyauchi
Research Institute of Electrical Communication, Tohoku University
2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan
ryota@ais.riec.tohoku.ac.jp
Sonoko Kobayashi, Chinatsu Miyauchi, Shushin Noda, Hitoshi Osaka, Yoshitaka Nakajima
Faculty of Design, Kyushu University
4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan

Abstract

We investigated whether perceptual integration of inharmonic components is affected by parallel/opposite frequency movements (FMs) of the components. Recognition thresholds (RTs) of a target component embedded in background components were measured. In the parallel-FM condition, the target and the background always moved in the same direction on logarithmic frequency. In the opposite-FM condition, they moved in opposite directions. When all the components ascended or descended monotonically, the RTs in the opposite-FM condition were 2-4 dB lower than those in the parallel-FM condition. This benefit increased to 4-5 dB when a target consisted of ascending or descending inharmonic components with vowel-like formants. These results suggest that a target is more easily heard out from a background in the opposite-FM condition than in the parallel-FM condition because perceptual integration between the target and the background is attenuated by different frequency movements. As the shape of FM became more complex, however, this effect was limited to certain conditions and eventually disappeared.

In complex acoustic environments, there are a variety of sounds from different sources. When these sounds reach our ears, they are mixed into an overlapping sound wave. Our auditory system, however, can hear out a particular acoustic source from this mixture of sounds. Bregman (1990) proposes that these complex acoustic environments are subjected to an *auditory scene analysis*. First, the mixture of sounds is decomposed into a number of discrete frequency components, and then the components that are likely to have arisen from the individual source are integrated into one perceptual sound object based on some acoustic cues of incoming sounds, such as harmonicity, synchronization of onsets and offsets (terminations), spatial location, and common amplitude and frequency movements.

Our research question is whether or not common frequency movements (FMs) actually work as an independent cue for perceptual integration. The frequency of the components that compose many natural sounds like vocal sounds and musical instrument's sounds move up and down in parallel. This fact leads to the idea that parallelism between frequency movements can be an important cue of perceptual integration.

However, there is little evidence for independent effects of coherent FM (Furukawa and Moore, 1997) partly because it is difficult to eliminate other cues from experimental procedure. One of the most efficient cues is harmonicity. When the components are harmonic, opposite FMs cause inharmonicity whereas parallel FMs preserve harmonicity. It is difficult to determine how and to what extent common FMs and harmonicity contribute separately or correlatively (McAdams, 1989; Bregman and Doehring, 1984). In order to avoid this problem, we employed only inharmonic components.

It is also important to take account of effects of simultaneous, forward, and backward maskings between neighboring components. In order to minimize these effects, we used the sinusoidal FM for the movement pattern, and the neighboring components were always separated by at least two equivalent rectangular bandwidths (ERBs) except in Experiment 4.

Experiments 1-3

We measured recognition thresholds (RTs) of a target embedded in a background. The target and the background were moved either in the same direction (the parallel-FM condition) or in opposite directions (the opposite-FM condition). If coherent FM can be an effective cue, the RT in the opposite-FM condition is supposed to shift to a lower value than that in the parallel-FM condition.

Method

Experiments 1, 2, and 3 were conducted following the same procedure. Both the target and the background were frequency modulated sinusoidally. The carrier frequencies (f_c) were 1067.9 Hz for the target component, and 356.4, 616.9, 1848.6, and 3200.0 Hz for the background components. These frequency components were set at regular intervals of 950 cents on logarithmic frequency; the five components were inharmonic. The modulation rate was 2.5 Hz. Although another modulation rate, 7.5 Hz, was employed as well in Experiment 3, we omit such conditions from the present report due to the limitation of space. The range of modulation was ±10% of f_c. In the parallel-FM condition, the target and the background were modulated in the same phase. In the opposite-FM condition, they were modulated in opposite phases. Any neighboring frequency components were separated by more than two ERBs at any point.

The duration of the target, that started and ended simultaneously with the background, was 200, 400, and 2000 ms including 20-ms cosine shaped rise and fall times in Experiments 1, 2, and 3, respectively. When the duration was 200 ms (a half period of the 2.5-Hz modulation), four initial modulation phases, 0, $\pi/2$, π, and $3\pi/2$, were employed. Thus, there were four FM patterns: Peak (P), Descending (D), Trough (T), and Ascending (A). When the durations were 400 and 2000 ms (a period and five periods of the 2.5-Hz modulation), two initial modulation phases, 0 and π, were employed. Thus, there were two FM patterns: +sine (2.5+) and –sine (2.5-). All combinations of these FM patterns employed in Experiments 1-3 are indicated in Figure 1.

There were two conditions for the presentation of the stimulus pattern, the *target condition* and the *no-target condition*. In both conditions, the target was presented alone at first. Two seconds after that, in the target condition, the same target was presented with the background. In the no-target condition, the target was absent when the background was presented. The stimuli were digitally generated at a 44.1-kHz sampling frequency, and presented monaurally via headphones (STAX Lambda Nova Basic System) through a D/A converter, a low-pass filter (cut-off frequency = 3.5 kHz in Experiment 2 and 8.3 kHz in the other experiments), and a driver unit (STAX SRM-Xh) in a soundproof room. The sound pressure level was calibrated with a precision sound level meter mounted with an artificial ear and a microphone (Brüel & kjær Type 2209, Type 4153, and Type 4144).

We used the constant method. In Experiments 1 and 2, the sound pressure level of each background was fixed at 74 dB SPL. The target level was changed in steps of 2 dB from -47 to +3 dB in relation to the background level. In Experiment 3, the background level was 71.5 dB SPL, and the target level was changed in steps of 4 dB from -44 to 4 dB. All target levels

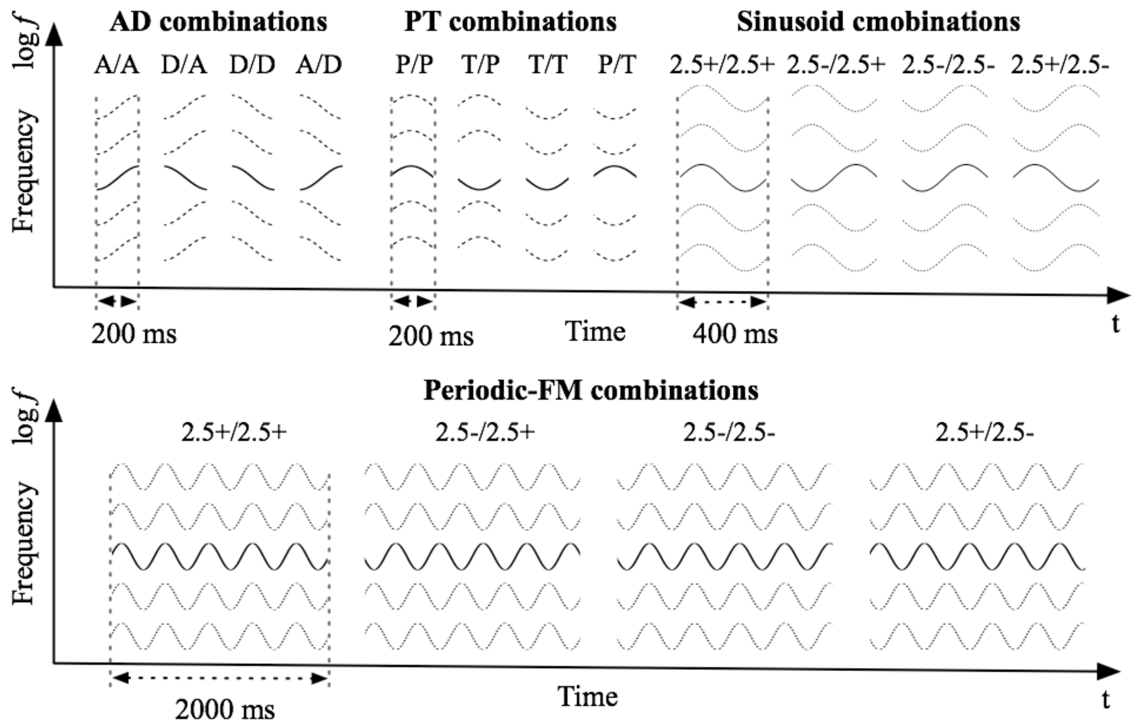

Figure 1. Schematic illustrations of FM combinations employed in Experiments 1-3. Frequency modulations are exaggerated.

in the target condition and the no-target condition appeared in random order six times for each participant, and the last five replications were used as data.

The participants were instructed to press "1" when the target was heard completely, "2" when the target was heard partially, or it was difficult to decide whether the target was heard or not, and "3" when the target was not heard. When the participant failed in hearing the stimulus pattern, it was allowed to present the same stimulus pattern once more, but no more after that.

Seven, nine, and seven university students with normal hearing participated in Experiments 1, 2, and 3, respectively. Except for two, the participant had received basic training in music and technical listening for acoustic engineers (Iwamiya et al., 2003).

Results and Discussion

The proportion of each response category was calculated for all target levels. The proportion of responses of "1" was plotted as a function of the relative target level, and a sigmoid function was fitted to this relationship (as a psychometric function). The experimental data for all FM patterns were approximated in this way, and the coefficients of goodness of fit (r^2) for the fitted curves were always greater than 0.98. Thus, the RT was defined as the relative target level corresponding to the 0.5 proportion of the fitted curve. The RTs of each FM combination are indicated in Table 1.

In the AD combinations, the RT in the opposite-FM condition was 2.1 dB (in D/A) and 4.3 dB (in A/D) lower than that in the parallel-FM condition (A/A and D/D, respectively). The results clearly showed that the target was more easily heard out from the background in the opposite-FM condition than in the parallel-FM condition even when no harmonicity existed, and, therefore, no component could stand out as being mistuned from harmonicity. In the PT and the sinusoid combinations, however, this effect appeared only in P/T and in

Table 1. The RTs in Experiments 1-3, expressed in dB. The benefit of the FM indicates the difference between the RTs in the parallel-FM condition and in the opposite-FM condition where the FM pattern of the background remains the same.

	Movement pattern of the background							
	AD		PT		Sinusoid		Periodic-FM	
FM condition	A	D	P	T	2.5+	2.5−	2.5+	2.5−
Parallel FM	−18.7	−16.5	−17.9	−18.0	−19.9	−18.4	−20.4	−19.7
Opposite FM	−20.8	−20.8	−17.5	−21.5	−25.2	−19.4	−20.5	−21.1
Benefit of the FM	2.1	4.3	-0.4	3.5	5.3	1.0	0.1	1.4

2.5+/2.5-, that is, the effect disappeared in T/P and reduced in 2.5+/2.5-. In the periodic-FM combinations, the effect completely disappeared.

Experiment 4

Experiments 1-3 offered evidence that parallel/opposite FMs affected perceptual integration of inharmonic components in certain FM combinations. Next question is whether such as effect of parallel/opposite FMs appears as well for speech-like signals.

Method

In most cases, a target and a background of the same duration 400 ms included 10-ms rise and fall times were presented simultaneously. The target consisted of 13 inharmonic components were set at regular intervals of 500 cents from 102.9 to 3293.8 Hz. The background consisted of 18 inharmonic components were set at regular intervals of 500 cents from 50.0 to 6780.6 Hz. The components of the target and the background were arranged at regular intervals of 250 cents alternately at the onset. The target was given vowel-like formants corresponding to Japanese vowels /a/ or /i/; the spectral patterns and the F1-F2 formant frequencies are indicated in Figure 2. In the ascending-FM pattern, all components linearly increased by 500 cents on logarithmic frequency from the frequency values mentioned above. The descending-FM pattern was the temporal mirror image of the ascending-FM pattern.

In each trial, two targets embedded in the same background were presented with 200-ms inter-stimulus interval either in the order of /a/ and /i/ or in the order of /i/ and /a/. There were three experimental sessions: (i) The target was ascending and the background was ascending or descending. (ii) The target was descending and the background was ascending or descending. (iii) The target was ascending or descending and the background was constant.

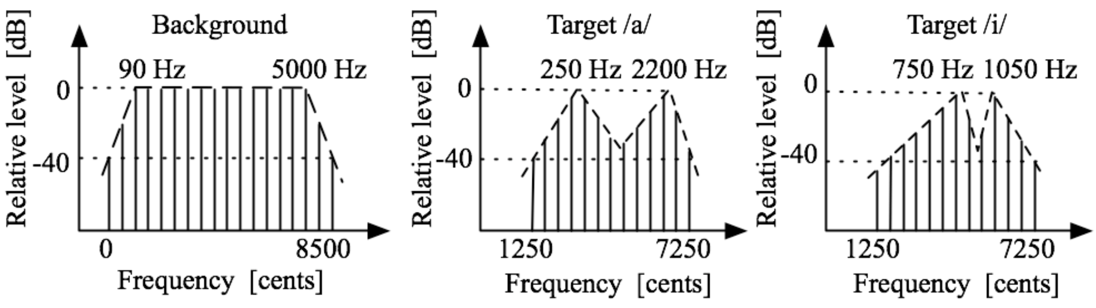

Figure 2. Spectral structures of the target and the background employed in Experiment 4.

A target-only pattern and a background-only pattern were employed in each experimental session. The stimuli were digitally generated at a 20-kHz sampling frequency, recorded by a DAT recorder through a low-pass filter (cut-off frequency = 9 kHz), and presented monaurally via headphones (RION AD-02) through a DAT player and an amplifier. The sound pressure level was calibrated in the same way as in Experiments 1-3.

The constant method was used. The sound level of the background was fixed at 77.2 dBA. The target level was changed in steps of 2 dB from -20 to 6 dB in relation to the background level. The relative levels were calculated utilizing the peaks of the spectral envelopes. All target levels appeared in random order seven times for each participant, and the last six replications were used as data.

The participants were instructed to press "1" when they heard /ai/ embedded in the background, "3" when they did not hear both two Japanese vowels, and "5" when they heard /ia/ embedded in the background. When they head only one vowel, or it was difficult to decide whether the targets were heard or not, they pressed "2" or "4."

Four university students with normal hearing participated. They were native speakers of Japanese and had received basic training in music and technical listening for acoustic engineers (Iwamiya, et al., 2003).

Results and discussion

A sigmoid function was fitted to the proportion of the correct identification of both two target vowels, responses of "1" in the /a/-/i/ condition or responses of "5" in the /i/-/a/ condition, plotted as a function of the relative target level as well in Experiments 1-3 ($r^2 > 0.98$). The RT was defined as the relative target level corresponding to the 0.5 proportion on the fitted curve. The RTs of each FM combination are indicated in Table 2.

The results showed that the RT in the opposite-FM condition was lower than that in the parallel-FM condition, and the benefit of the FM was greater than that in the AD combinations in Experiment 1, except in one condition (/ai/, A/A vs. D/A). These results indicated that the effect of the FM appeared for speech-like signals.

General discussion

The present results clearly showed that the RT in the opposite-FM condition was lower than that in the parallel-FM condition, when the components ascended or descended monotonically. This effect was not caused by the FM differences of the target itself, because when the frequencies of the background's components were constant in Experiment 4, the RTs of the ascending and descending targets were the same. The *coherence* between the FMs of the target and the background must be an important cue of perceptual integration.

The parallel/opposite FM between the target and the background affected their perceptual integration even when the components were *inharmonic* and therefore opposite

Table 2. The RTs and the benefits of the FM in Experiment 4, expressed in dB.

	Movement pattern of the background					
Movement pattern	/ai/			/ia/		
of the target	A	D	C	A	D	C
A	−3.0	−6.0	−5.0	−1.7	−6.6	−4.4
D	−5.1	−1.4	−4.7	−5.7	−1.4	−4.4
Benefit of the FM	2.1	4.6		4.0	5.2	

FM did not cause of mistuning from harmonicity. Common FMs seemed to facilitate perceptual integration of acoustic components independently. This effect, however, was limited to certain conditions.

One possible explanation of the asymmetry that appeared in the PT and the Sinusoid combinations is based on the sensibility of the FM detection for peak/trough FMs and ascending/descending FMs. Demany and McAnally (1994) reported that peak-shaped FMs were better detected than trough-shaped FMs. If this asymmetry of FM detection works in the PT combinations, the movement of a P-shaped target should be traced more easily than the movement of a T-shaped target. Thus, the target may have been heard out more easily in T/P than in P/T. Similar asymmetry was reported for ascending/descending FMs. For example, Carlyon and Stubbs (1989) found that ascending FMs were more detectable than descending FMs. The 2.5+ pattern has a descending-FM excursion in the middle part, which produces a percept of a descending glide. Similarly, the 2.5- pattern produces a percept of an ascending glide. This can explain why the target was detected more easily in 2.5-/2.5+ than in 2.5+/2.5-. In some studies, however, thresholds were lower for descending glides than for ascending glides (Tsumura et al., 1973). This result indicates an opposite tendency from the assumption discussed above, and we have to compare seemingly different results in detail as a next step.

Acknowledgments

We would like to thank Albert S. Bregman for his valuable comments to Experiment 2 and Yôiti Suzuki for his support. This research was supported by grants to Y. Nakajima from the Japan Society for the Promotion of Science (JSPS) (19103003 in the fiscal year 2007) and by a 'Center of Excellence (COE)' program of the Japanese government entitled 'Design of Artificial Environments on the Basis of Human Sensibility.'

References

Bregman, A.S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Bregman, A.S. & Doehring, P. (1984). Fusion of simultaneous tonal glides: The role of parallelness and simple frequency relations. *Perception & Psychophysics*, *36*, 251-256.

Carlyon, R.P. & Subbs, R.J. (1989). Detecting single-cycle frequency modulation imposed on sinusoidal, harmonic, and inharmonic carriers. *Journal of the Acoustical Society of America*, *85*, 2583-2574.

Demany, L. & McAnally, K.I. (1994). The perception of frequency peaks and troughs in wide frequency modulations. *Journal of the Acoustical Society of America*, *96*, 706-715.

Furukawa, S., & Moore, B.C.J. (1997). Dependence of frequency modulation detection on frequency modulation coherence across carriers: Effects of modulation rate, harmonicity, and roving of the carrier frequencies. *Journal of the Acoustical Society of America*, *101*, 1632-1643.

Iwamiya, S., Nakajima, Y., Ueda, K., Kawahara, K., & Takada, M. (2003). Technical listening training: Improvement of sound sensitivity for acoustic engineers and sound designers. *Acoustical Science & Technology*, *24*, 27-31.

McAdams, S. (1989). Segregation of concurrent sounds I: Effects of frequency modulation coherence. *Journal of the Acoustical Society of America*, *86*, 2148-2159.

Tsumura, T., Sone, T., and Nimura, T. (1973). Auditory Detection of frequency transition. *Journal of the Acoustical Society of America*, *53*, 17-25.