TR－A－0138 04

Relative pitch judgments for formant structured broadband noise

Kazuo Ueda

1992. 3.16

ATR視聴覚機構研究所
〒619-02 京都府相楽郡精華町光台2-2 ☎07749-5-1411

ATR Auditory and Visual Perception Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
Telephone: +81-7749-5-1411
Facsimile: +81-7749-5-1408

Relative pitch judgments for formant structured broadband noise

Kazuo Ueda

Hearing & Speech Perception Department
ATR Auditory and Visual Perception Research Laboratories
Sanpeidani, Inuidani, Seika-cho, Soraku-gun, Kyoto, 619-02 Japan

The effect of formant structure on pitch of broadband noise was investigated. Series of stimuli whose formant structures were modified systematically, were constructed by means of filtering. Single formant stimuli (Experiment 1) and three formant stimuli (Experiment 2) were generated. Paired comparison was used to assess the pitch of the stimulus. The data were analyzed by a multidimensional scaling technique. The results show that the spectral energy distribution (the frequency and the level of the formants) of the stimuli affect pitch, and that the second formant frequency is not the sole determinant of pitch of broadband noise. This is contrary to previous results on pitch of whispered vowels. The close relation among the psychological dimension of noise pitch, tone height, and sharpness was suggested.

PACS numbers: 43. 66. Hg, 43. 66. Jh, 43. 71. -k

* This article is under submission to the Journal of the Acoustical Society of America.

INTRODUCTION

It has already been shown that pitch sensation is affected by both periodicity of the stimulus and its spectral energy distribution. For example, a paired comparison experiment of complex tones similar to Shepard's (1964), in which listeners had to choose the higher stimulus, showed that pitch judgment was influenced by both periodicity and spectral energy distribution, and that these two physical factors corresponded to two psychological dimensions of pitch, tone chroma and tone height, respectively. However, the relative weights of the judgment of the two dimensions differed considerably among listeners (Ueda and Ohgushi, 1987). It was shown that the peak shift of the large formant, which covered the whole spectral range of the stimulus, affected pitch: when the peak went up, pitch tended to be higher even if the periodicity of the stimulus remained unchanged.

Several other studies have shown that pitch can be affected by spectral composition of sounds (e.g., Brink, 1977; Chuang and Wang, 1978; Ohgushi, 1978; Hesse, 1982; Moore and Glasberg, 1990, 1991; Patterson, 1990). In the auditory nerve fibres, the information of spectral energy distribution of a stimulus is conveyed as "place information," whereas the information of periodicity (below 5 kHz) of a stimulus is conveyed as "temporal information." Taking other phenomena of pitch perception into account (see Demany, 1989, for an extensive review), it has been suggested that both place and temporal information play important roles in pitch perception, and that neither alone is sufficient to entirely explain the phenomena of pitch perception (Noorden, 1982; Moore, 1989). In particular, it has been hypothesized that place information induces tone height, whereas temporal information induces tone chroma (Ohgushi, 1976; Noorden, 1982).

How would pitch perception be, then, if the sound had no clear periodicity: that is, if it was noise (we will restrict ourselves to monaural or diotic hearing here)? Révész (1913) considered that noise pitch consists of tone height alone. Since that time, other research has shown experimentally the existence of noise pitch, especially for band noise (e.g., Small and Daniloff, 1966; Rainbolt and Schubert, 1968; Houtsma, 1984; Zwicker and Fastl, 1990).

These studies seem to indicate that there are two ways of perceiving noise pitch, and that the pitch is subject to the experimental context. One is that the pitch corresponding frequency of a spectral edge is perceived. When the task for the subjects was pure tone matching, subjects tended to match a pure tone frequency to the frequency of a spectral edge (Zwicker and Fastl, 1990). In the other experiment, subjects attained octave matching to the standard stimulus when the cutoff frequency of the comparison stimulus was adjusted (Small and Daniloff, 1966). Other support for the correlation between this "edge pitch" and a spectral edge comes from the fact that positive correlation exists between pitch strength and steepness of spectral edges (Fastl, 1980; Zwicker and Fastl, 1990).

On the other hand, when a forced choice procedure and a noise comparison stimulus were utilized, i.e. when the subjects were asked to judge whether the comparison stimulus (1/2-oct. band noise) was lower than the standard stimuli (equal to, or broader than 1/2-oct.), no "edge pitch" was apparent in group data (Rainbolt and Schubert, 1968). Instead, the estimated frequency which corresponded to the pitch of the standard stimulus fell somewhere around the center of the band noise. However, individual data suggested that some of the subjects did perceive the edge pitch, which corresponded to the upper or the lower limit of the noise (however, no one seemed to perceive the two distinct pitches which corresponded to both the upper and the lower limit of the noise).

Thus, it can be hypothesized that it is possible to perceive the edge pitch, if the noise has a distinct spectral edge. Therefore, to exclude this possibility, the use of broadband noise should be considered. There seems to be no systematic study of the pitch of broadband noise, especially when spectral energy distribution is varied. However, since whisper can be regarded as a kind of broadband noise, the studies of whispered vowel pitch should be related to this issue.

There have been two published investigations which measured pitch of whispered vowels. Helmholtz (1954) measured the pitch of several whispered vowels by comparing with the sound of tuning-forks. In another study, the pitch of nine whispered English vowels (each produced by two speakers, one male and one female) were determined by the method of adjustment, utilizing a pure-tone oscillator (Thomas, 1969). In both cases, the determined frequencies roughly correspond to the second formant frequencies. However, three aspects of these results should be reconsidered.

First, the task of matching the pure-tone frequency to the pitch of the whispered vowels would be difficult: since the spectrum of a whisper is spread over a wide frequency range, and is complex in shape, the pitch of whisper is more ambiguous compared with a pure tone. It is doubtful whether complete matching can be attained between these two sounds. Thomas (1969) reported that the second formant frequency was not the only answer for some subjects: they could also match the frequency of the pure tone to that of the first formant of a whisper, /a/ and /ɔ/. Therefore, the approach taken by Rainbolt and Schubert

(1968), that is, to utilize a noise stimulus as a comparison stimulus may be more suitable for assessing pitch of noise.

Second, the number of speech samples seems too small compared with the large variety inherent in speech production. One way to overcome this difficulty would be to establish some basic knowledge on perception of noise pitch in general.

Third, Meyer-Eppler (1957) showed that several types of changes can be observed on spectrograms when subjects sing whispered vowels: shift of the third formant with the first two formants remaining unchanged (/a/), shift of the fifth formant (/a/), shift of the first formant (/u/), and increasing intensity of the higher spectral regions (/e, i, a, u/). Although he did not examine the dependency of pitch on these physical factors, his results do question the notion that the second formant frequencies are the sole determinant of pitch of whisper. A systematic investigation would be needed on this point, including pitch of broadband noise in general.

Therefore, the purpose of this study is to investigate how frequency and level of formants affect pitch of broadband noise. If noise pitch corresponds to the frequency of a formant peak, the level of the formant should have no effect on pitch. To exclude the possibility that the subjects perceive edge pitch of band noise, all the stimuli utilized had a broad spectrum which covered the whole audible frequency range. Instead of pure tone matching, the pitch of the noise was assessed by paired comparison of the noise stimuli, and the results were analyzed by a multidimensional scaling technique (MDS).

In Experiment 1, the pitch of a single formant broadband noise was determined relatively; a systematic operation was applied to the formant frequency and to the level of the formant. In Experiment 2, three formant stimuli were used, including the stimuli whose formant frequencies were close to the first three formant frequencies of whispered vowels, /a/ and /u/.

I. EXPERIMENT I: PITCH OF SINGLE FORMANT BROADBAND NOISE

A. Method

1. Stimuli

A total of 14 stimuli were prepared. Pink noise generated from a noise generator (Brüel & Kjær, type 1049) was passed through a digital equalizer (Yamaha, DEQ7) and a lowpass filter (NF, FV-665, 20 kHz, 96 dB / Oct.), to an AD/DA converter (Pavec, MD-8000 mkII, 16 bits, 48 kHz sampling) connected to a computer

(Masscomp, MC5600). The digital equalizer consisted of A/D, D/A converters (16 bits, 44.1 kHz sampling) and a DSP. The digital filter of the equalizer was a second order filter, IIR, and Butterworth type. A formant was created using the digital equalizer, with its parametric equalizer function. The Q value was fixed at 3.2, whereas the peak frequencies were varied between 150, 300, 600, 1.2 k, 2.4 k, 5 k, and 10 kHz, and the filter gains at the peak frequencies were both +10 and +18 dB. Since the Q value of the equalizer was fixed, the spectral "skirts" were broadened when the filter gain at the peak frequencies was raised from +10 to +18 dB.

To estimate how much the stochastic characteristics of the pink noise influence the stimuli produced, the FFT spectrum was obtained of the pink noise generated from the generator. An FFT analyzer (Advantest, TR9403) was used for this purpose. The spectrum of the noise was averaged over 320 ms, and this yielded one measurement. The measurement was repeated 10 times, each for different portion of the noise. The results showed that the deviation of the spectrum from the "ideal" spectrum was a maximum of plus or minus 3 dB. Thus, it was concluded that the effect of the spectrum fluctuation caused by the stochastic characteristics of the noise on the stimuli would be negligible, at least for stimuli longer than 320 ms.

The stimuli were excised on the computer, so that the duration of each stimulus became 400 ms, including 20 ms rise- and fall-time. During the rise- and fall-time, the amplitude envelope was tapered by a second order function.

A total of 182 pairs were made by pairing each stimulus with every other stimulus. The pairs were randomly ordered, to yield a stimulus series of 182 trials. Four series of different order were made. The time interval between a pair of stimuli was 300 ms, and the interval between trials was 2 s. After DA conversion, the stimulus series were recorded on a tape with a digital audio tape (DAT) recorder (Sony, DTC-1000 ES). In the experiment, the stimuli were played back to the subjects from the DAT.

2. Procedure

A total of 11 subjects participated in the experiment. All were students at Dosisha University and reported no hearing difficulty and no previous history of auditory pathology. The subjects were paid for their service.

They had to judge the higher stimulus in each pair. Even if they found it difficult, they were forced to decide. Each stimulus combination was judged eight times. The stimuli were presented diotically through headphones (STAX SR Lambda Professional with STAX SRM-1/MK 2 adaptors). The sound pressure level of a stimulus (formant frequency: 1.2 kHz, filter gain at the formant frequencies: +18 dB) measured with a coupler (Brüel & Kjær, type 4153) and a precision sound-level meter (Brüel & Kjær, type 2231), was about 66 dB SPL.

To provide the listeners with a perceptual frame of reference upon which to base their judgments, they were presented with all 14 stimuli prior to making any specific judgments. After that, to familiarize them with the task, they were given 30 trials for practice. They took a few minutes rest after every 182 trials.

B. Results

The subject responses were arranged in the form of a matrix, which showed the frequencies with which one stimulus was judged to be higher than another. Profile distances among the stimuli (Kruskal and Wish, 1978) were calculated to yield a psychological distance matrix of pitch. The psychological distance matrix was analyzed by a non-metric MDS technique. The program used was ALSCAL. Judged from stress values and the configuration obtained, the one dimensional solution was adopted (stress value: 0.058). Fig. 1 shows the configuration obtained. The stimuli in the right side were judged to be higher than the left side. Note that this analysis was based on the judgments of all possible combination of the stimuli, even when the configuration was one dimensional.

On the whole, the higher the formant frequency, the higher the pitch of the stimulus is judged to be, although the stimuli of formant frequencies 5 and 10 kHz (stimuli 6 and 7 in the figure) are exceptions. Concerning the level of the formant, for the stimuli with formant frequency below 300 Hz, a larger formant induced a lower pitch, whereas for the stimuli with formant frequency above 600 Hz, a larger formant induced a higher pitch. Moreover, the higher the formant frequency, the larger is the pitch difference caused by the level difference.

The small distances among stimuli 1, 2 and 3 reflect the fact that the pitch differences among these are very subtle, although, one subject responded fairly convincingly (e.g., responded in the same way for six trials out of eight, for the combination of the stimulus 1L and 1H). The larger distances among other stimuli imply that the subjects perceived more distinct pitch differences.

C. Discussion

Both formant frequency and level of formant affect pitch, and there is an interaction between them. That is, the pitch relation among the stimuli cannot be directly predicted only from formant frequencies: the frequency region where a formant exists and the level of the formant, i.e. spectral energy distribution of the stimuli, must be taken into account.

The existence of a balance point along the frequency axis, where the level variation of a formant has no effect on pitch, is assumed. In the frequency region lower than the balance point, an increase of spectral energy results in lower pitch, whereas in the frequency region higher than the balance point, an increase of spectral energy results in higher pitch.

The pitch difference between stimuli 6 and 7 was very small. One reason may be that ear sensitivity gets worse at high frequency. If this is the case, the fact that the pitch difference between the stimulus 6H and 7H was smaller than the difference between the stimulus 6L and 7L can be explained as a result of recruitment. Another reason may be that pitch sensation at high frequency generally becomes more ambiguous. The subjects may not have been used to judging pitch differences between noise stimuli which differed in formant frequency in the high frequency region.

No attempt was made to equalize loudness among the stimuli in this experiment. However, the loudness differences seem to have little effect on the results. No psychological dimension which corresponded to loudness was found. If the subject response was based on loudness, the level effect of low formant frequency stimuli and that of high formant frequency stimuli should not be contradictory. Moreover, the level effect should be maximum for the stimuli whose formant lies in the middle frequency region, since loudness difference would be maximum at middle frequencies. However, the opposite is true in both cases.

The phenomena that pitch goes down when the level of a low frequency formant increases, and that the opposite happens for a high frequency formant, may be associated with "Stevens rules" which describe pitch-intensity effects of a pure tone (Stevens, 1935). It was confirmed that Stevens rules hold true for the averaged results over a number of subjects, although the effect was smaller than shown by Stevens (Walliser, 1969).

Moreover, Miyazaki (1977) suggested that the effect is largely dependent on the pitch concept, which is apt to vary according to the experimental context. He supposed that timbre is influenced by intensity, not pitch which is related to the matching of musical intervals. Since the stimuli used in the present experiment have no clear periodicity, the pitch sensation elicited would be very similar to a kind of timbre, depending on its spectrum. Thus the association between Stevens rules and the present results appears to be justified, as long as the averaged results are concerned.

However some difficulties have recently been pointed out for Stevens rules. It has been shown that there are large individual differences and large ear differences observed in the pitch-intensity relation of pure tone (Verschuure and Meeteren, 1975; Burns, 1982). For a given frequency, an increment of intensity causes an up-ward shift of pitch for some subjects, while the same thing causes down-ward shift for others. For some subjects, the shapes of the pitch-intensity function are different between the ears, and are sometimes nonmonotonic.

Thus, to determine the relation between Stevens rules and the pitch of broadband noise, further experiments would be needed which would allow precise comparison between pure tone pitch and noise pitch, for each ear of each individual. However, this is beyond the scope of the present investigation.

II. EXPERIMENT II: PITCH OF THREE FORMANT BROADBAND NOISE

This experiment investigated the issue: is pitch determined by the peak frequency of a specific formant or not, if the stimuli have more than one formant?

A. Experimental method

1. Stimuli

A total of three stimulus groups (A-C) were constructed. Each group consisted of 7 stimuli. Table I shows the parameters of these stimuli. The formant frequencies of stimulus ST in the stimulus group B and C were close to the values of the first three formants of whispered vowels by males, /a/ and /u/ respectively, determined by Kallail and Emanuel (1984). Stimuli other than ST in each group were constructed by modifying one of the formants of stimulus ST: the modification was either an upward shift of a formant frequency or an increase in level of a formant. The degree of frequency shift was an octave in the group A, and 100 to 400 Hz in groups B and C. The degree of formant frequency shifts in groups B and C were in accordance with to the results of an analysis made by Meyer-Eppler (1957).

The experimental apparatus used in stimulus generation and presentation, and Q value of the filter were the same as in experiment 1. The stimuli were paired and then randomly ordered within each stimulus group. Each group was presented with a series of 168 trials. Each series consisted of four blocks of 42 trials which were different from each other in stimulus order.

2. *Procedure*

A total of 11 subjects participated in the experiment. None of them participated in Experiment 1. All were students at Dosisha University and reported no hearing difficulty and no previous history of auditory pathology. The subjects were paid for their service.

The experimental procedure was basically the same as that of Experiment 1. However, since it was predicted that pitch variation in this experiment might be smaller than in the previous one, additional trials were included at the end of the experiment in order to check the ability of the subjects concerning pitch judgment. The task in these trials was the same as in the previous trials. However, the stimuli were different: they were pure tones, whose frequencies in each pair corresponded to one of the formant frequencies of stimulus ST and the shifted frequency of that formant. A total of 28 trials were carried out.

The sound pressure level of the stimulus F2H in group B was about 69 dB SPL.

B. Results

The data were omitted for three subjects who committed errors more than twice in the test of pure tone pitch judgment. The common error of these three subjects was for the comparison of 5 and 10 kHz stimuli, although these subjects also made mistakes at much lower frequencies, e.g. 2.8 and 2.4 kHz, or 950 and 850 Hz. The frequency differences in this test were well above normal frequency DLs. Therefore, there should be some other reasons for the errors than the subjects' frequency discriminability. In the case of errors at high frequencies, one reason may be that the subjects were not used to judging pitch of high frequency. Another possible reason, independent of frequency, may be that the subjects failed to assign their responses correctly, even though they could discriminate the pitch.

The data of the other eight subjects were analyzed in the same way as in Experiment 1, and one dimensional solutions were adopted. The configurations obtained are shown in Figure 2(a)-(c). The stress values were 0.007, 0.025, and 0.021, respectively. The stimuli on the right of Fig. 2 were judged to be higher than those on the left.

The results can be summarized as follows.

(1) Shift of formant frequency induced pitch shift, regardless of formant number.

(2) The effects of formant level increase were different from formant to formant, and from stimulus group to stimulus group.

For all the stimulus groups, an increase of the F_1 level induced a lower pitch, but an increase of the F_3 level resulted in the opposite effect. For the stimulus group A, an increase of the F_2 level resulted in a slightly lower pitch, whereas for the stimulus group B and C, it resulted in moderately higher pitch. Generally, the level effect was smallest for F_2, and largest for F_3.

C. Discussion

The results of this experiment showed that the frequency of a second formant is not the sole determinant for the pitch of broadband noise, but frequency and level of the other formants also affect pitch. Shift of the formant frequency induced higher pitch. The increase in level of a low frequency formant induced lower pitch, whereas the increase in level of a high frequency formant induced higher pitch. These trends were essentially the same for all the stimulus groups, which differed from each other in the combination of the formant frequencies.

The balance point discussed for the results of the previous experiment, can also be assumed for the present results. If the balance point is defined as the frequency of a formant whose level variation has no effect on the pitch of the stimulus, the exact position of the balance point will depend on the frequencies and levels of other formants. In the stimulus group A, the formant frequency of F_2 was very near to the balance point, since the level variation of F_2 had almost no effect on pitch. However, in the stimulus group B and C, a frequency somewhat lower than the F_2 would be the balance point.

The discussion above implicitly hypothesized that the subjects perceived a single pitch for each stimulus. In other words, it was hypothesized that they did not focus their attention on each individual formant, and did not perceive several pitches at the same time. Support for this hypothesis comes from the fact that the Q of the filter was rather broad (3.2), and that the order of stimulus presentation was unpredictable, for the subjects who would like to focus their attention to the individual formant. However, further investigation is needed to gain stronger support for this hypothesis.

Although the stimuli utilized in this experiment had complex formant structures, it was very hard to hear them as speech, probably because they were steady sounds without any of the fluctuations which are characteristic of speech. However, the argument that the comparison of the results of this study with those of previous studies on whisper is not fair, seems inadequate, at least concerning the study by Thomas (1969). The stimuli he used were excised whisper vowels, and they were in steady state.

These sounds would sound less like speech. Moreover, he told his subjects that the stimuli were "noise." Thus it is hard to believe that his results were deeply concerned with the speech specific mode of perception.

Instead, the discrepancy between the present results and the previous results probably arise mainly from the difference in the measurement method: pure tone matching would be an inadequate method to assess pitch of whisper and broadband noise. The reason why this method is thought to be inadequate was explained in the Introduction. On the other hand, the method utilized in the present study, i.e., relative pitch judgment among noise stimuli, is much better and more natural than pure tone matching because when we perceive intonation in whisper, we judge relative pitch of noise stimuli, and we do not match pure tone.

III. GENERAL DISCUSSION

The previous studies of pitch of band noise were variously affected by edge pitch, partly because the stimuli had definite spectral edges well within the hearing frequency range, and partly because of their experimental paradigms which were typically pure tone matching. The studies of pitch of whisper, a kind of broadband noise, also suffered from the problem of experimental method.

To eliminate these undesired effects, the present investigation utilized broadband noise stimuli which had formant structures, and the pitch of the stimuli were assessed by a paired comparison method. The results showed that both frequency and level of a formant affect pitch, and that the F_2 frequency is not the sole determinant of the noise pitch, but any formant frequency affects pitch. Upward shift of the formant frequencies induced higher pitch. Regarding the level of the formants, generally an increase in level of low frequency formants induced lower pitch, whereas an increase in level of high frequency formants induced higher pitch. A balance point where level variation of a formant has no effect on pitch, was assumed. However further exploration is needed for estimation of the exact position of this point, precise analysis of the individual differences in noise pitch perception, examination of the relation between the present results and Stevens rules, and the control of the subjects' attention.

The psychological dimension of noise pitch appears to be similar to that of timbre "sharpness" (Bismarck, 1974a,b), since both of them are closely related to spectral energy distribution of a stimulus. Clear distinction between this dimension and sharpness would be very difficult (cf. Miyazaki, 1977); pitch itself is highly correlated with sharpness (Bismarck, 1974a), and differences in timbre (whether upper harmonics are in common or are absent) impair frequency difference limens (Moore and Glasberg, 1990, 1991), although the effect was not found under other experimental conditions (Demany and Semal, 1992). This dimension may be strongly related to "tone height" (Ohgushi, 1976; Noorden, 1982) which are hypothesized to be closely related to sharpness and place information in the auditory nerve.

ACKNOWLEDGMENTS

The author would like to thank Dr. Laurent Demany, Dr. Carl Müller, and Professor Hitoshi Okada for their helpful comments on the earlier version of this article. He also thanks Dr. Yoh'ichi Tohkura for his support with this work.

Bismarck, G. von (1974a). "Timbre of steady sounds: A factorial investigation of its verbal attributes," Acustica, 30, 146-159.

Bismarck, G. von (1974b). "Sharpness as an attribute of the timbre of steady sounds," Acustica, 30, 159-172.

Brink, G. van den (1977). "Octave and fifth settings for pure tones and residue sounds," in *Psychophysics and Physiology of Hearing,* edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 373-379.

Burns, E. M. (1982). "Pure-tone pitch anomalies. I. Pitch-intensity effects and diplacusis in normal ears," J. Acoust. Soc. Am. 72, 1394-1402.

Chuang, C. K., and Wang, W. S. (1978). "Psychophysical pitch biases related to vowel quality, intensity difference, and sequential order," J. Acoust. Soc. Am. 64, 1004-1014.

Demany, L. (1989). "Perception de la hauteur tonale," in *Psychoacoustique et perception auditive,* by M. C. Botte, G. Canévet, L. Demany, and C. Sorin (INSERM, Paris), Chap. 2. (Japanese translation by K. Ueda (1989). ATR Technical Report, TR-A-0050.)

Demany, L., and Semal, C. (1992). "Detecting combined differences in pitch and sharpness of timbre," Proc. 2nd ICMPC, C8-3, pp. 45-46.

Fastl, H. (1980). "Pitch strength and masking patterns of low-pass noise," in *Psychophysical, Physiological and Behavioural Studies in Hearing,* edited by G. van den Brink, and F. A. Bilsen (Delft University Press, Delft), pp. 334-340.

Helmholtz, H. L. F. von (1954). *On the Sensations of Tone,* translated by A. J. Ellis (Dover, New York), p. 108.

Hesse, H. P. (1982). "The judgment of musical intervals," in *Music, Mind, and Brain,* edited by M. Clynes (Plenum, New York), Chap. 11.

Houtsma, A. J. M. (1984). "Pitch salience of various complex sounds," Music Perception, 1, 296-307.

Kallail, K. J., and Emanuel, F. W. (1984). "An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects," J. Phon. 12, 175-186.

Kruskal, J. B., and Wish, M. (1978). *Multidimensional Scaling* (Sage, London), pp. 77-82.

Meyer-Eppler, W. (1957). "Realization of prosodic features in whispered speech," J. Acoust. Soc. Am. 29, 104-106.

Miyazaki, K. (1977). "Pitch-intensity dependence and its implications for pitch perception," Tohoku Psychologica Folia, 36, 75-88.

Moore, B. C. J. (1989). *An Introduction to the Psychology of Hearing* (Academic, London), 3rd ed., Chap. 5.

Moore, B. C. J., and Glasberg, B. R. (1990). "Frequency discrimination of complex tones with overlapping and non-overlapping harmonics," J. Acoust. Soc. Am. 87, 2163-2177.

Moore, B. C. J., and Glasberg, B. R. (1991). "Effects of signal-to noise ratio on the frequency discrimination of complex tones with overlapping or nonoverlapping harmonics," J. Acoust. Soc. Am. 89, 2858-2865.

Noorden, L. van (1982). "Two channel pitch perception," in *Music, Mind, and Brain,* edited by M. Clynes (Plenum, New York), Chap. 13.

Ohgushi, K. (1976). "Mechanism of pitch perception of complex tones," J. Acoust. Soc. Jpn. 32, 300-309 (in Japanese).

Ohgushi, K. (1978). "On the role of spatial and temporal cues in the perception of the pitch of complex tones," J. Acoust. Soc. Am. 64, 764-771.

Patterson, R. D. (1990). "The tone height of multiharmonic sounds," Music Perception, 8, 203-214.

Rainbolt, H. R., and Schubert, E. D. (1968). "Use of noise bands to establish noise pitch," J. Acoust. Soc. Am. 43, 316-323.

Révész, G. (1913). *Zur Grundlegung der Tonpsychologie* (Vert & Comp, Leipzig).

Shepard, R. N. (1964). "Circularity in judgments of relative pitch," J. Acoust. Soc. Am. 36, 2346-2353.

Small and Daniloff (1966). "Pitch of noise bands," J. Acoust. Soc. Am. 41, 506-512.

Stevens, S. S. (1935). "The relation of pitch to intensity," J. Acoust. Soc. Am. 6, 150-154.

Thomas, I. B. (1969). "Perceived pitch of whispered vowels," J. Acoust. Soc. Am. 46, 468-470.

Ueda, K., and Ohgushi, K. (1987). "Perceptual components of pitch: Spatial representation using a multidimensional scaling technique," J. Acoust. Soc. Am. 82, 1193-1200.

Verschuure, J., and Meeteren, van A. A. (1975). "The effect of intensity on pitch," Acustica, 32, 33-44.

Walliser, von K. (1969). "Über die Abhängigkeiten der Tonhöhenempfindung von Sinustönen vom Schallpegel, von überlagertem drosselndem Störschall und von der Darbietungsdauer," Acustica, 21, 211-221.

Zwicker, E., and Fastl, H. (1990). *Psychoacoustics: Facts and Models* (Springer-Verlag, Berlin), Chap. 5., pp. 120-130.

TABLE I Stimulus parameters (frequencies (Hz) and levels (L or H) of the formants) used in Experiment 2. ST: standard. "L" and "H" signify that the filter gain at the peak frequencies of the filter were set to +10 dB and +18 dB, respectively. "-" signifies that the same parameter as the standard was used. The formant frequencies of STs in the stimulus group B and C are close to those of the whispered vowels, /a/ and /u/, respectively, which were measured by Kallail and Emanuel (1984).

(a) Stimulus group A.

Stimulus	F1		F2		F3	
ST	150	L	1.2 k	L	5 k	L
F1H	-	H	-	-	-	-
F2H	-	-	-	H	-	-
F3H	-	-	-	-	-	H
F1S	300	-	-	-	-	-
F2S	-	-	2.4 k	-	-	-
F3S	-	-	-	-	10 k	-

(b) Stimulus group B.

Stimulus	F1		F2		F3	
ST	850	L	1.12 k	L	2.4 k	L
F1H	-	H	-	-	-	-
F2H	-	-	-	H	-	-
F3H	-	-	-	-	-	H
F1S	950	-	-	-	-	-
F2S	-	-	1.33 k	-	-	-
F3S	-	-	-	-	2.8 k	-

(c) Stimulus group C.

Stimulus	F1		F2		F3	
ST	400	L	1.12 k	L	2.4 k	L
F1H	-	H	-	-	-	-
F2H	-	-	-	H	-	-
F3H	-	-	-	-	-	H
F1S	500	-	-	-	-	-
F2S	-	-	1.33 k	-	-	-
F3S	-	-	-	-	2.8 k	-

FIGURE CAPTION

FIG. 1. Stimulus configuration obtained from Experiment 1. The stimuli on the right were judged to be higher than those on the left. Points are numbered from the lowest formant frequency (150 Hz) to the highest (10 kHz). 'L' and 'H' signify that the filter gain at the peak frequencies of the filter were set to +10 and +18 dB, respectively. Arrows show the correspondence between the stimuli of the same formant frequency.

FIG. 2. Stimulus configuration obtained from Experiment 2. The stimuli on the right were judged to be higher than those on the left. (a)-(c) represent the results for stimulus group A-C, respectively. ST: standard. F: formant. H: increased formant level. S: shifted formant frequency.

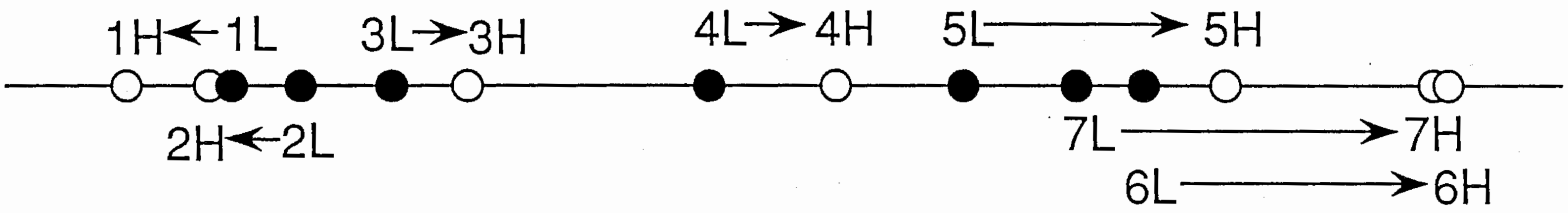

Fig. 1

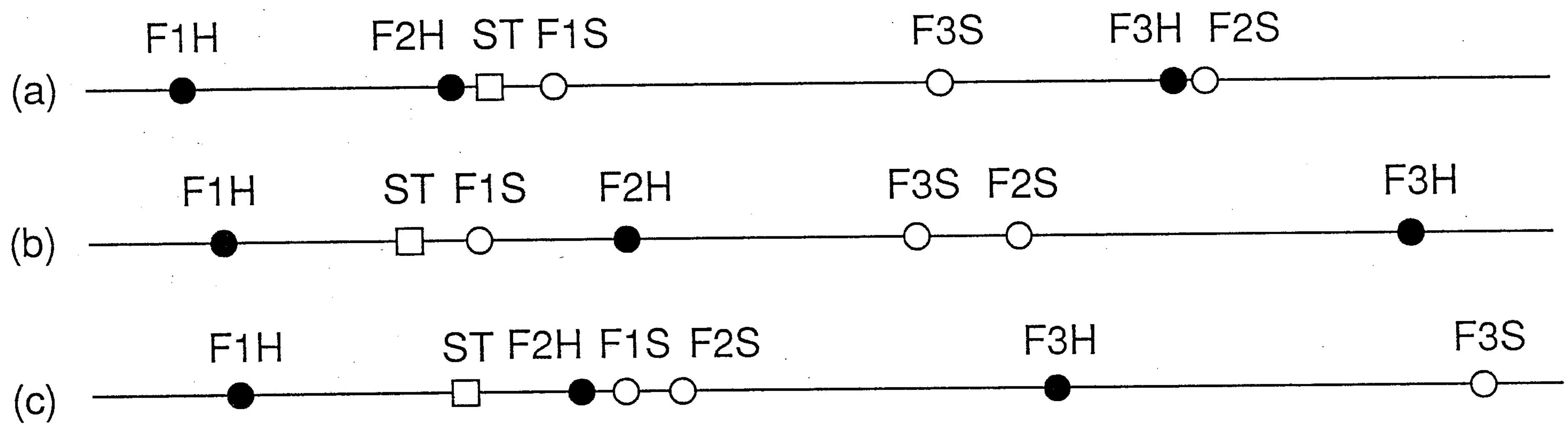
F1H
F2H
ST
F1S
F3S
F3H
F2S
(a)
F1H
ST
F1S
F2H
F3S
F2S
F3H
(b)
F1H
ST
F2H
F1S
F2S
F3H
F3S
(c)

Fig. 2