Comparison of Factors Extracted from Power Fluctuations in Critical-Band-Filtered Homophonic Choral Music

Kazuo Ueda,[*1] Yoshitaka Nakajima[*2]

[*] *Department of Human Science and Center for Applied Perceptual Research, Kyushu University, Japan*
[1]ueda@design.kyushu-u.ac.jp, [2]nakajima@design.kyushu-u.ac.jp

ABSTRACT

A consistent pattern of three factors, which led to four common frequency bands with boundaries of about 540, 1720, and 3280 Hz, had been obtained from factor analyses of power fluctuations of critical-band-filtered spoken sentences in a variety of languages/dialects. The aim of the present investigation was to clarify whether the same factors and frequency bands could be found in homophonic choral music sung with texts in English, Japanese, or nonsense syllables, or with mono-vowel vocalization. Recordings of choral music were analyzed. Three factors and four frequency bands similar to those obtained from spoken sentences appeared in the analyses of music with ordinary texts in English and Japanese. However, no distinct structure was observed in the analysis of a tune, which was sung with no text but a mimicked buzz of bumblebees, and another tune, which was vocalized with a single vowel. Thus, it was suggested that the patterns of the first three factors could appear if there was a certain amount of syllable variety in choral music, and that basically the same frequency channels were utilized for conveying speech information both in spoken sentences and in choral music.

I. INTRODUCTION

The present authors and their colleagues found that factor analyses on power fluctuations of critical-band-filtered speech (Figure 1) showed a consistent pattern of three factors in eight languages/dialects, and the factors led to four common frequency bands with frequency boundaries of about 540, 1720, and 3280 Hz (Ueda, Nakajima, & Satsukawa, 2010). These findings signified that there is a way to combine the critical bands (Fletcher, 1940; Zwicker & Terhardt, 1980; Schneider, Morreongiello, & Trehub, 1990; Fastl & Zwicker, 2007) into fewer frequency bands to extract a small number of power fluctuations that can convey speech information. This line of investigations was inspired by the pioneering work of Plomp and his colleagues (Plomp, Pols, & van de Geer, 1967; Pols, Tromp, & Plomp, 1973; Plomp, 1976, 2002), in which they extracted principal components from the spectra of steady Dutch vowels.

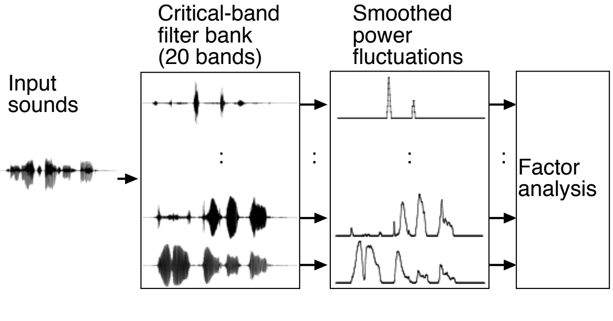
Input sounds
Critical-band filter bank (20 bands)
Smoothed power fluctuations
Factor analysis

Figure 1. A block diagram of the analyses. The analyses were repeated alternately with critical-band-filter banks A and B (see Table 1, for details). Each filter output was squared and smoothed to obtain a power fluctuation. Factor analyses were based on correlation coefficients between the concurrent power fluctuations. Principal components analyses were performed to the correlation coefficient matrices, and then varimax rotation was applied to the first three components.

It had been revealed that four frequency bands could make noise-vocoded speech reasonably intelligible (Schannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Dorman, Loizou, & Rainey, 1997; Smith, Delgutte, & Oxenham, 2002; Riquimaroux, 2006), but the frequency boundaries had not been selected on a clear basis. We synthesized four-band noise-vocoded speech utilizing the above frequency bands, and 95% accuracy in word intelligibility was attained even without feedback (Ueda, Araki, & Nakajima, 2009). The amounts of information transmitted in relation to voicing, manners of articulations, and places of articulations, were obtained from confusion

Table 1. Critical-band-filter settings. The center frequencies and the passbands are expressed in Hz. The frequencies in the bank A were basically adopted from Zwicker and Terhardt (1980). The center frequencies of the filters in the bank B were placed at the upper cutoff frequencies of the corresponding filters in the bank A.

Band	Bank A		Bank B	
number	Center fr.	Passband	Center fr.	Passband
1	75	50-100	100	50-150
2	150	100-200	200	150-250
3	250	200-300	300	250-350
4	350	300-400	400	350-450
5	450	400-510	510	450-570
6	570	510-630	630	570-700
7	700	630-770	770	700-840
8	840	770-920	920	840-1000
9	1000	920-1080	1080	1000-1170
10	1170	1080-1270	1270	1170-1370
11	1370	1270-1480	1480	1370-1600
12	1600	1480-1720	1720	1600-1850
13	1850	1720-2000	2000	1850-2150
14	2150	2000-2320	2320	2150-2500
15	2500	2320-2700	2700	2500-2900
16	2900	2700-3150	3150	2900-3400
17	3400	3150-3700	3700	3400-4000
18	4000	3700-4400	4400	4000-4800
19	4800	4400-5300	5300	4800-5800
20	5800	5300-6400	6400	5800-7000

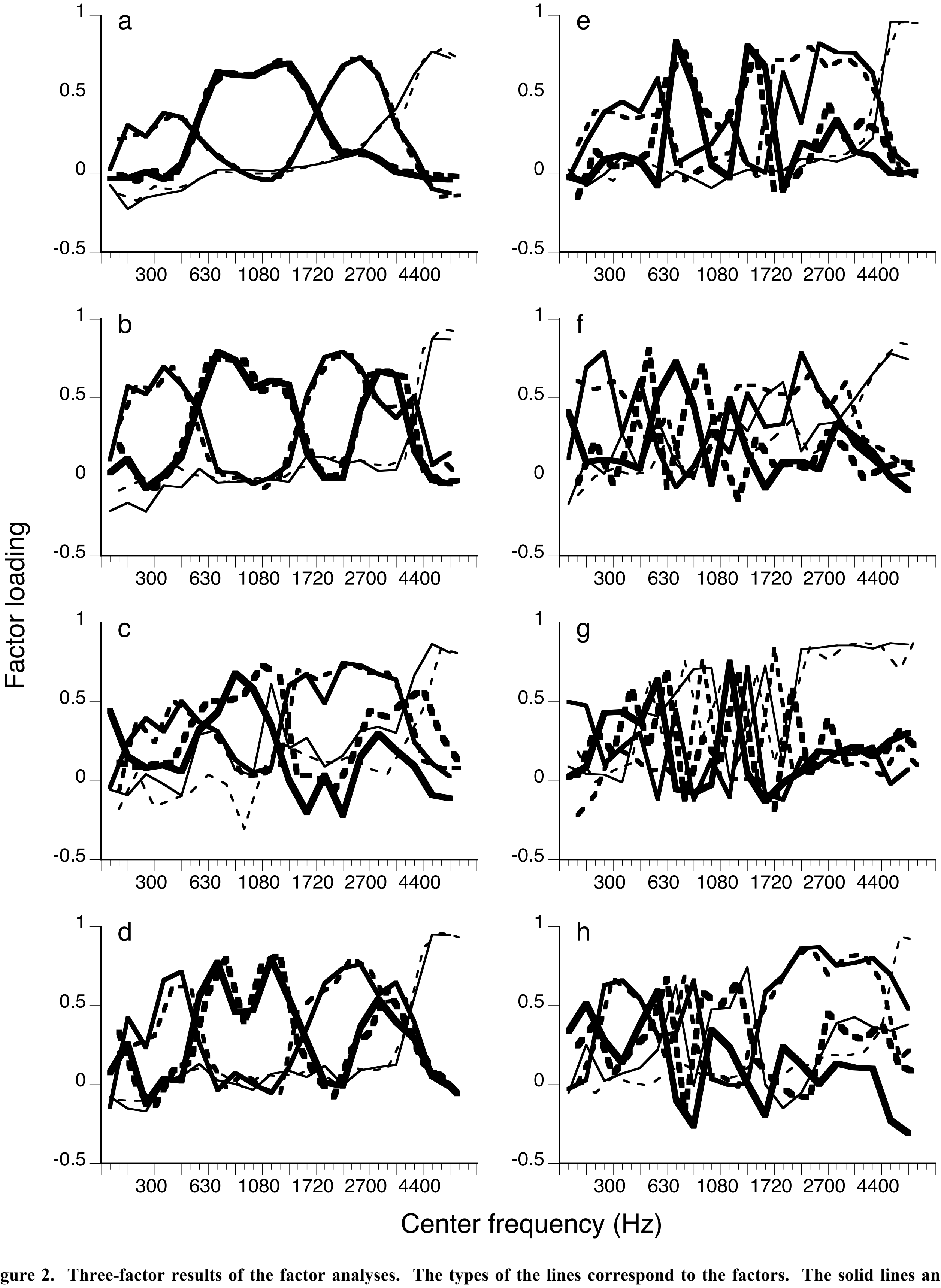

Figure 2. Three-factor results of the factor analyses. The types of the lines correspond to the factors. The solid lines and the dashed lines represent the results obtained with critical-band filter banks A and B, respectively. (a) Speech of eight languages/dialects (from Ueda, Nakajima, & Satsukawa, 2010), (b) Now is the month of maying (Morley, 1595), (c) O happy eyes (Elgar, 1890), (d) Can't buy me love (McCartney and Lennon, 1963), (e) Karatachi-no hana (Flowers of trifoliate orange) (Yamada, 1925), (f) The flight of the bumblebee (Rimsky-Korsakov, 1901), (g) To be sung of a summer night on the water, I (Delius, 1917), and (h) The splendour falls on castle walls (Delius, 1923).

matrices under various experimental conditions; we emplyed original Japanese speech (monosyllables), 20-band noise-vocoded speech, 4-band noise-vocoded speech, and degraded versions of 4-band noise-vocoded speech, in which one of the bands was eliminated. The power fluctuations in the lowest band of the 4-band conditions played a crucial role for identifying voicing of speech, whereas the power fluctuations in the second lowest band was vital for identifying vowels. Thus, the frequency bands yielded from purely acoustic analyses of speech were closely related to speech perception, at least in Japanese (Noguchi, Satsukawa, Ueda, & Nakajima, 2011).

We were interested in whether spoken sentences and sung texts were perceived from the same acoustic cues or not (e.g., Nakajima, Takeichi, Kidera, & Ueda, 2012). Thus, the aim of the present investigation was to clarify whether the same or similar factors and frequency bands would be found in homophonic choral music as in spoken sentences.

II. METHOD

A. Chorus Samples

Recordings of *a cappella* choral music, mainly obtained from commercially available compact discs, were analyzed. They were recordings by the King's Singers, the Finzi Singers, the CBSO Chorus, and the Kyoto Academy Choir.

The list of the choral music analyzed was the following: (a) Now is the month of maying (Morley, 1595, track 4), (b) O happy eyes (Elgar, 1890, track 1), (c) Can't buy me love (McCartney & Lennon, 1963, track 16), (d) Karatachi-no hana (Flowers of trifoliate orange) (Yamada, 1925), (e) The flight of the bumble bee (Rimsky-Korsakov, 1901, track 11), (f) To be sung of a summer night on the water, I (Delius, 1917, track 3), and (g) The splendour falls on castle walls (Delius, 1923, track 5). (e) and (f) were not entirely homophonic, but we included them in the list because we needed examples of tunes of nonsense syllables or a single vowel.

B. Signal Processing and Analyses

The stereo signals were mixed into mono signals, and the sampling frequency was reduced from 44.1 to 16 kHz, keeping the amplitude resolution of 16-bit quantization. The same method of analysis was applied as in our previous investigation on speech (Figure 1). The whole analysis procedure was repeated alternately with two banks of 20-critical-band filters, A and B (Table 1), for each input. Each filter output was squared, smoothed with a Gaussian window of $\sigma = 5$ (ms), and sampled at every 1 ms to obtain power fluctuations. Factor analyses with varimax rotation were performed on the correlation matrices of these power fluctuations.

III. Results

We obtained 3-factor results (Figures 2), including previous results obtained from speech analyses [Figure 2a; from Ueda et al. (2010)], with cumulative contributions of (a) 34 and 35%, (b) 48 and 51%, (c) 42 and 44%, (d) 46 and 49%, (e) 41 and 44%, (f) 42 and 42%, (g) 54 and 55%, and (h) 52 and 51%, utilizing the filter banks A and B, respectively. In the analyses of music with ordinary texts in English or Japanese (Figures 2b-e, h), three factors, which were similar to those obtained from spoken sentences, were observed. By taking the crossover frequencies of the curves as boundaries, we observed four frequency bands similar to those obtained from spoken sentences. Figure 2f shows the results of the analyses of a tune sung with no text but a mimicked buzz of bumblebees; it was very difficult to observe a distinct structure, except a peak of a factor appeared above 3280 Hz. Figure 2g shows the results obtained from the analyses of a tune vocalized with a single vowel; a flat shape of a factor above 2000 Hz, and irregular patterns of all the factors below that frequency, were characteristic.

IV. DISCUSSION

Factor analyses were performed on correlation matrices of power fluctuations obtained from critical-band-filtered choral music. Music with normal texts yielded similar factors and frequency bands to those ordinary speech had yielded. However, music with limited variations in syllables yielded unstructured factors.

A third factor, which usually exhibits a peak above 3280 Hz, did not appear in Figures 2g, probably because the music did not contain fricatives: only a vowel was sung throughout the tune. In contrast, a peak above 3280 Hz appeared in Figure 2f, probably because the tune contained plenty of a fricative, i.e., /z/.

Although lyrics, written by Lord Alfred Tennyson, were sung in one of the tunes analyzed (Figure 2h), the results looked unstable compared with the results of other tunes with normal texts. At present, no clear explanation of these results can be provided: However, a slow tempo, a short text with partial repetition, and rhyming, might have reduced appearances and variety of syllables, hence, a shortage of syllables to be analyzed might have led to these in-between results.

V. CONCLUSION

The same method of analysis could be applied to both spoken sentences and choral music. Choral music with ordinary texts exhibited a similar structure of factors and frequency boundaries as in spoken sentences, whereas choral music without such texts exhibited disorganized structures of factors. It was suggested that similar frequency channels were utilized for conveying speech information both in spoken sentences and in choral music.

ACKNOWLEDGMENTS

This research was supported by Grants-in-Aid for Scientific Research Nos. 14101001, 19103003, 20330152, and 23653227 from the Japan Society for the Promotion of Science, Kyushu University Interdisciplinary Programs in Education and Projects in Research Development, and by a Grant-in-Aid for the 21st Century COE program from the Ministry of Education, Culture, Sports, Science and Technology.

REFERENCES

Delius, F. (1917). To be sung of a summer night on the water, I [Recorded by the CBSO Chorus]. On *Delius: Part songs; Grainger: Folk songs for Chorus, Songs from the Jungle Book* [CD]. London: Conifer. (1988)

Delius, F. (1923). The splendour falls on castle walls, by A. Tennyson (Lyrics) [Recorded by the CBSO Chorus]. On *Delius: Part songs; Grainger: Folk songs for Chorus, Songs from the Jungle Book* [CD]. London: Conifer. (1988)

Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America*, 102, 2403-2411.

Elgar, E. (1890). O happy eyes, by C. A. Elgar (Lyrics) [Recorded by the Finzi Singers]. On *Elgar: Part-Songs* [CD]. Colchester: Chandos. (1994)

Fletcher, H. (1940). Auditory patterns. Review of Modern Physics, 12, 47–65.

McCartney, P., & Lennon, J. (1963). Can't buy me love, by K. Abbs (Arrangement) [Recorded by the King's Singers]. On *The Beatles' Collection* [CD]. Tokyo: JVC. (1986)

Morley, T. (1595). Now is the month of maying [Recorded by the King's Singers]. On *All at Once Well Met: English Madrigals* [CD]. Hayes: EMI. (1974)

Nakajima, Y., Takeichi, H., Kidera, S., & Ueda, K. (2012). Multivariate analyses of speech signals in singing and non-singing voices. In E. Cambouropoulos (Ed.), *Proceedings of the 12th International Conference on Music Perception and Cognition* (in press). Greece: Aristotle University of Thessaloniki.

Noguchi, K., Satsukawa, Y., Ueda, K., & Nakajima, Y. (2012). Effects of frequency-band elimination on syllable identification in Japanese noise-vocoded speech: Analyses with two speakers. *Proceedings of Auditory Research Meeting, the Acoustical Society of Japan, 42,* 231-236.

Plomp, R. (1976). *Aspects of Tone Sensation: A Psychophysical Study*. London: Academic Press.

Plomp, R. (2002). *The Intelligent Ear: On the Nature of Sound Perception*. Mahwah, New Jersey: Lawrence Erlbaum.

Plomp, R., Pols, L. C. W., & van de Geer, J. P. (1967). Dimensional analysis of vowel spectra. *Journal of the Acoustical Society of America*, 41, 707–712.

Pols, L. C. W., Tromp, H. R. C., & Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *Journal of the Acoustical Society of America,* 53, 1093–1101.

Rimsky-Korsakov, N. A. (1901). The flight of the bumblebee, by D. Runswick (Arrangement) [Recorded by the King's Singers]. On *A Tribute to the Comedian Harmonists* [CD]. Tokyo: EMI Angel. (1984)

Riquimaroux, H. (2006). Perception of noise-vocoded speech sounds: Sentences, words, accents and melodies. *Acoustical Science & Technology,* 27, 325–331.

Schneider, B. A., Morreongiello, B. A., & Trehub, S. E. (1990). Size of critical band in infants, children, and adults. *Journal of Experimetal Psychology: Human Perception and Performance*, 16, 642– 652.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303–304.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416, 87-90.

Ueda, K., Araki, T., & Nakajima, Y. (2009). The effect of amplitude envelope coherence across frequency bands on the quality of noise-vocoded speech, EURONOISE 2009, Edinburgh, Scotland, *Acta Acustica united with Acustica*, 95, Suppl. 1, S107.

Ueda, K., Nakajima, Y., & Satsukawa, Y. (2010). Effects of frequency-band elimination on syllable identification of Japanese noise-vocoded speech: Analysis of confusion matrices. *Proceedings of the 26th Annual Meeting of the International Society for Psychophysics*, Padova, Italy, 39-44.

Yamada, K. (1925). *Karatachi-no hana (Flowers of trifoliate orange),* by H. Kitahara (Lyrics) [Recorded by the Kyoto Academy Choir]. [Cassette recording]. (1981)

Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bnadwidth as a function of frequency. *Journal of the Acoustical Society of America,* 68, 1523-1525.