

Acoustic Features of English Phonemes Obtained with Multivariate Analyses

張, 一新

<https://hdl.handle.net/2324/7182472>

出版情報 : Kyushu University, 2023, 博士 (芸術工学) , 課程博士
バージョン :
権利関係 :



Acoustic Features of English Phonemes Obtained with Multivariate Analyses

多変量解析を用いた英語音素の音響的特徴の分析



張一新

Yixin Zhang

March, 2024

Acknowledgments

The deepest and sincerest gratitude to my supervisor Prof. Dr. Gerard B. Remijn for his continuous support, invaluable guidance, and warm encouragement throughout my doctoral study. I also would like to express special thanks to Prof. Dr. Yoshitaka Nakajima and Prof. Dr. Kazuo Ueda for reviewing and valuable comments which helped in improving the quality of my dissertation.

I would like to thank all the students in Prof. Remijn's laboratory and Prof. Ueda's laboratory, especially Clarissa Pesigan and Mira Adhani who always helped me by my side during the study. In addition, I would also like to thank those who had an essential role in my study, Prof. Dr. Mark A. Elliott at the National University of Ireland, Galway. Thanks to Dr. Takuya Kishida for helping me to solve some technical issues and Xiaoyang Yu who endeavored to create the database with me.

My greatest thanks go to my strongest backing, my grandparents Mingyi Wang and Liping Zhang, my parents Feng Zhang and Hong Wang, and my beloved Minghan Li gave me enough love, respect, and unconditional support. More special thanks to Ge Liu and Mengdi Song for our eternal friendship.

Finally, I would like to express my gratitude to everyone who helped and supported me during these seven years in Japan and witnessed me from a master's student to a PhD student, making it like a smooth and memorable journey. This journey makes me grow up and become a better myself. For the future, I want to express it with one of my favorite words "Shoot for the moon. Even if you miss, you'll land among the stars".

Abstract

The general purpose of the dissertation is to elucidate how acoustic features of English speech can be expressed using multivariate analyses. There were three research purposes. The first purpose was to investigate the suitability of factor analyses of changing spectra for studying acoustic natures of English phonemes. Here, the results of a newly developed factor analysis, “origin-shifted factor analysis”, were compared with the results of a normal factor analysis to analyze the spectral power fluctuations in English speech. The methods mainly differed in whether to use cepstral liftering and/or the origin shift. The results showed that three spectral factors were obtained in four main frequency bands, and neither the cepstral liftering nor the origin shift distorted the essential characteristics of the factors. This confirms that origin-shifted factor analysis is more recommendable for future speech analyses since it would reduce the generation of continuous noise in the data reduction procedure of factor analysis.

The second purpose was to determine the acoustic features of English phonemes in general, by analyzing the results of the aforementioned factor analyses. In the previous research on factor analysis of English phonemes by Nakajima et al. (2017), they indicated the English database that was used had poor recording quality and some phoneme labeling issues. To better explore the acoustic features of English speech, a newly recorded and labeled database with higher quality was created. The results showed that by using origin-shifted factor analysis, regardless of whether three or four spectral factors were extracted, the factors always appeared in a similar set of four frequency bands. When three spectral factors were extracted, the distributions of

the factor scores on a factor with the frequency loadings around 1200 Hz were very close to the sonority hierarchy in phonology. Sonority is a concept related to phoneme classification, and it is considered necessary to understand syllable formation. The analysis results with the new speech database corresponded to the notion in phonology that vowels have the highest sonority, followed by sonorant consonants, while obstruents are the lowest in terms of sonority. The present results revealed that, when four factors were extracted, the distributions of factor scores on two factors were highly related to the sequence of the sonority hierarchy. One factor had frequency loadings around 1200 Hz, and the other was double-peaked around 700 Hz and 2200 Hz. Furthermore, the factor with the frequency loadings around 300 Hz clearly divided English phonemes into obstruents and non-obstruents (vowels and sonorant consonants).

The third purpose was to obtain clear insight into whether and how obstruents would appear in the factor spaces. In an English syllable, an isolated obstruent always has a position as a syllable onset or a syllable end. Analyzing obstruents acoustically as in the present study helps to understand this phonological phenomenon. The current three-factor analysis provided evidence that obstruents were not only associated with the factor related to a frequency loadings around 4100 Hz, as suggested in previous research, but also with the double-peaked factor with frequencies around 300 Hz and 2300 Hz. Moreover, in the four-factor analysis, the distribution of obstruents in English speech was related to two spectral factors. One factor had frequency loadings around 4100 Hz, while the other around 300 Hz. Generally, the results of the present

dissertation clearly show that multivariate analyses can be utilized to connect English phonology and speech acoustics, even for low-sonorous phonemes such as obstruents.

Research Output

The following Chapters of this dissertation have been presented or published in the following conference and peer-reviewed journal.

Chapter 2 was published in the international, peer-reviewed journal “*Applied Sciences*” (MDPI, Switzerland).

Zhang, Y., Nakajima, Y., Ueda, K., Kishida, T., & Remijn, G. B. (2020). Comparison of Multivariate Analysis Methods as Applied to English Speech. *Applied Sciences*, 10(20), 7076. <https://www.mdpi.com/2076-3417/10/20/7076>

Chapter 3 was presented at ISSN: 2435-7030 –The Asian Conference on Language 2021 (ACL), Tokyo, Japan.

Zhang, Y., Nakajima, Y., Ueda, K., & Remijn, G. B. (2021). How the Acoustic Correlates of English Obstruents Appear in Multivariate Analysis. In ISSN: 2435-7030 *The Asian Conference on Language 2021 (ACL)*. <https://doi.org/10.22492/issn.2435-7030.2021.6>

The basic research method was presented in an oral session of the 37th Annual Meeting of the International Society for Psychophysics, Fechner Day 2021, 19-23 October.

Zhang, Y., Nakajima, Y., Ueda, K., & Remijn, G. B. (2021). Acoustic Correlates of English Consonant-vowel-consonant (CVC) Words Obtained with Multivariate Analysis. *The 37th Annual Meeting of the International Society for Psychophysics Fechner Day 2021* (pp.82).

Table of Contents

Acknowledgments	ii
Abstract	iii
Research Output	vi
Table of Contents	vii
List of Tables	ix
List of Figures	x
Chapter 1. Introduction: speech sounds	1
1.1 Phonology	2
1.1.1 Sonority	4
1.1.2 Syllable structure	7
1.2 Multivariate analysis of speech sounds	10
1.3 Purpose and contents of this dissertation	15
Chapter 2. Study 1: Comparison of multivariate analysis methods as applied to English speech	18
2.1 Purpose	18
2.2 Method	19
2.2.1 Speech samples	19
2.2.2 Procedure	19
2.3 Results	22
2.3.1 Factor loadings of the three spectral factors	22
2.3.2 Factor loadings of the speech from female and male speakers	25
2.3.3 Factor scores of the three spectral factors	27
2.3.4 General distribution of English phonemes in the factor space	33
2.3.5 Distribution of English obstruents in the factor space	34
2.4 Discussion and conclusions	39
Chapter 3. Study 2: A three-factor analysis of English speech	42
3.1 Purpose	42
3.2 Method	43
3.2.1 Speech samples	43

3.2.2 Procedure	47
3.3 Results	49
3.3.1 Factor loadings of the three spectral factors	49
3.3.2 Factor scores of the three spectral factors	51
3.3.3 General distribution of English phonemes in the factor space	51
3.3.4 Distribution of English obstruents in the factor space	54
3.4 Discussion and conclusions	58
Chapter 4. Study 3: A four-factor analysis of English speech	61
4.1 Purpose	61
4.2 Method	62
4.2.1 Speech samples	62
4.2.2 Procedure	62
4.3 Results	63
4.3.1 Factor loadings of the four spectral factors	63
4.3.2 Factor scores of the four spectral factors	66
4.3.3 General distribution of English phonemes in the factor space	70
4.3.4 Distribution of English obstruents in the factor space	83
4.4 Discussion and conclusions	84
Chapter 5. General discussion	87
References	95
Appendix A. The 100 speech sentences used in the Fukuoka-Galway English Speech Database (Campbell, 1993)	99
Appendix B. Informed consent and instructions for creating the Fukuoka-Galway English Speech Database	103
Appendix C. The expanded version of Figure 3.3 (a)-(d)	104
Appendix D. The expanded version of Figure 4.3 (a)-(e)	108

List of Tables

Table 1.1. The phonemes of British English (BBC pronunciation) by International Phonetic Alphabet (Jones, 2011). 3

Table 1.2. Previously proposed sonority (aperture) scales (from Nakajima et al., 2017). 6

Table 2.1. Division of speech signals into 19 critical-band filters, as used for the analyses. The critical band number, center frequency (Hz) and the passband of each critical band filter are shown with reference to Zwicker and Terhardt (1980) and Kishida et al. (2016). 21

Table 3.1. Critical-band filters used in Study 2. Different from the analysis in Study 1, speech signals were divided into 20 critical-band filters. The band number, center frequency (Hz), and the passband of each critical band filter are shown with reference to Zwicker and Terhardt (1980) and Kishida et al. (2016). 48

Table 5.1. An overview of the results of the acoustic correlates of English phonemes concerning sonority by using origin-shifted factor analyses of Studies 1, 2, and 3 in Chapters 2, 3, and 4, respectively. 90

List of Figures

Figure 2.1. Factor loadings of the three extracted spectral factors of 200 English speech samples from three native speakers. Four different factor analyses were performed: (a) Normal factor analysis without cepstral liftering, similar to Ueda and Nakajima (2017) and Nakajima et al. (2017); (b) Origin-shifted factor analysis without cepstral liftering; (c) Normal factor analysis with cepstral liftering; (d) Origin-shifted factor analysis with cepstral liftering, following Kishida et al. (2016). 24

Figure 2.2. Factor loadings of the three extracted spectral factors from the speech of two female native speakers in panel (a), and from one male speaker in panel (b) by origin-shifted factor analysis with cepstral liftering. 26

Figure 2.3. Factor loadings of the three extracted spectral factors from one male native speaker. Two analyses were performed: (a) Origin-shifted factor analysis with cepstral liftering, and (b) Origin-shifted factor analysis without cepstral liftering. 27

Figure 2.4. The distribution of factor scores by normal factor analysis without cepstral liftering, similar to Nakajima et al (2017). (a–c) represent all possible combinations of two factors out of the three factors, and (d) shows a two-dimensional mapping of the three factors. 29

Figure 2.5. The distribution of factor scores by origin-shifted factor analysis without cepstral liftering. (a–c) represent all possible combinations of two factors out of the three factors, and (d) shows a two-dimensional mapping of the three factors. 30

Figure 2.6. The distribution of factor scores by normal factor analysis with cepstral liftering. (a–c) represent all possible combinations of two factors out of the three factors, and (d) shows a two-dimensional mapping of the three factors. 31

Figure 2.7. The distribution of the factor scores by origin-shifted factor analysis with cepstral liftering, following Kishida et al. (2016). (a–c) represent all possible combinations of two factors out of the three factors, and (d) shows a two-dimensional mapping of the three factors. 32

Figure 2.8. The L-shaped distributions of voiced obstruents (a) and voiceless obstruents (b) in the two-dimensional space of the “Low & mid-high factor” and the “High factor”. Factor scores of the three extracted spectral factors were obtained by origin-shifted factor analysis with cepstral liftering (cf. Figure 2.7) and the English speech samples were from three native speakers. 36

Figure 2.9. The L-shaped distributions of obstruents divided into fricatives/affricates (a) and plosives (b) in the two-dimensional space of the “Low & mid-high factor” and the “High factor”. Factor scores of the three extracted spectral factors were obtained by origin-shifted factor analysis with cepstral liftering (cf. Figure 2.7) and the English speech samples were from three native speakers. 38

Figure 3.1. An example of segmented phoneme labeling in a sentence from the “Fukuoka-Galway English Speech Database”. The sentence was “The price range is smaller than any of us expected”. The first row shows the speech signal’s waveform; the second row shows the spectrogram of the spoken speech; the third row shows the phoneme labeling, and the fourth row shows the word labeling, which made up a

spoken sentence. Vowels “@” (/ə/), “ai” (/aɪ/), “ei” (/eɪ/). Sonorant consonants “r” (/r/) and “n” (/n/) are highlighted in blue. Obstruents “dh” (/ð/), “p” (/p/), “s” (/s/), and “jh” (/dʒ/) are highlighted in black. 46

Figure 3.2. Factor loadings of the three extracted spectral factors of 300 English speech samples, from three native speakers in the “Fukuoka-Galway English Speech Database”, with origin-shifted factor analysis without cepstral liftering. The first factor, the “Low & mid-high factor” (red line) was double-peaked, with high factor loadings on two center frequency ranges around 300 Hz and around 2300 Hz. The second factor, the “Mid-low factor”, was related to the frequency range around 1200 Hz (black line). The third factor, the “High factor”, was related to the frequency range around 4100 Hz (blue line). 50

Figure 3.3. The distribution of the factor scores by origin-shifted factor analysis without cepstral liftering. (a–c) represent all possible combinations of two out of the three factors, and (d) shows a two-dimensional mapping of the three factors. 53

Figure 3.4. The distributions of the factor scores of obstruents divided into fricatives/affricates (a) and plosives (b) in the two-dimensional space of the “Low & mid-high factor” and the “High factor”. 55

Figure 3.5. The distributions of the average of the factor scores of obstruents divided into fricatives/affricates (a) and plosives (b) in the two-dimensional space of the “Low & mid-high factor” and the “High factor”. 57

Figure 4.1. Factor loadings of the four extracted spectral factors of 100 English speech samples, from three native speakers in the “Fukuoka-Galway English Speech

Database”, with origin-shifted factor analysis (a) with cepstral liftering, and (b) without cepstral liftering. Four factors appeared in four critical-bands. The first factor, the “Low factor” was related to the frequency range around 300 Hz (red line). The second factor, the “Mid-low factor” was related to the frequency range around 1200 Hz (black line). The third factor, the “Mid-high factor”, was double-peaked, related to two frequency ranges around 700 Hz and 2200 Hz (green line). The fourth factor, the “High factor” was related to the frequency range around 4100 Hz (blue line). 65

Figure 4.2. The distribution of the factor scores by origin-shifted factor analysis with cepstral liftering. (a–f) represent all possible combinations of two factors out of the four factors. 67

Figure 4.3. The distribution of the factor scores by origin-shifted factor analysis without cepstral liftering. (a–f) represent all possible combinations of two factors out of the four factors. 69

Figure 4.4. The distribution of the average factor scores of each English phoneme by origin-shifted factor analysis with cepstral liftering. (a–f) represent all possible combinations of two factors out of the four factors. 73

Figure 4.5. The distribution of the average factor scores of each English phoneme by origin-shifted factor analysis without cepstral liftering. (a–f) represent all possible combinations of two factors out of the four factors. 74

Figure 4.6. The enlargement of Figure 4.5 (a). Panel (a) represents the distribution of the average factor scores of each English phoneme in the two-dimensional space of the “Low factor” and the “High factor” by origin-shifted factor analysis without

cepstral liftering (cf. Figure 4.5a). Panels (b–d) represent the enlargement of panel (a) in three phoneme categories, i.e. vowels, sonorant consonants, and obstruents. 76

Figure 4.7. The enlargement of Figure 4.5 (b). Panel (a) represents the distribution of the average factor scores of each English phoneme in the two-dimensional space of the “Mid-low factor” and the “High factor” by origin-shifted factor analysis without cepstral liftering (cf. Figure 4.5b). Panels (b–d) represent the enlargement of panel (a) in three phoneme categories, i.e. vowels, sonorant consonants, and obstruents. 77

Figure 4.8. The enlargement of Figure 4.5 (c). Panel (a) represents the distribution of the average factor scores of each English phoneme in the two-dimensional space of the “Mid-high factor” and the “High factor” by origin-shifted factor analysis without cepstral liftering (cf. Figure 4.5c). Panels (b–d) represent the enlargement of panel (a) in three phoneme categories, i.e. vowels, sonorant consonants, and obstruents. 78

Figure 4.9. The enlargement of Figure 4.5 (d). Panel (a) represents the distribution of the average factor scores of each English phoneme in the two-dimensional space of the “Mid-low factor” and the “Mid-high factor” by origin-shifted factor analysis without cepstral liftering (cf. Figure 4.5d). Panels (b–d) represent the enlargement of panel (a) in three phoneme categories, i.e. vowels, sonorant consonants, and obstruents.

79

Figure 4.10. The enlargement of Figure 4.5 (e). Panel (a) represents the distribution of the average factor scores of each English phoneme in the two-dimensional space of the “Low factor” and the “Mid-high factor” by origin-shifted factor analysis without

cepstral liftering (cf. Figure 4.5e). Panels (b–d) represent the enlargement of panel (a) in three phoneme categories, i.e. vowels, sonorant consonants, and obstruents. 80

Figure 4.11. The enlargement of Figure 4.5 (f). Panel (a) represents the distribution of the average factor scores of each English phoneme in the two-dimensional space of the “Low factor” and the “Mid-low factor” by origin-shifted factor analysis without cepstral liftering (cf. Figure 4.5f). Panels (b–d) represent the enlargement of panel (a) in three phoneme categories, i.e. vowels, sonorant consonants, and obstruents. 81

Figure 4.12. The distribution of English phonemes obtained by origin-shifted factor analysis without cepstral liftering in the three-dimensional factor space. Panel (a) shows a two-dimensional mapping of the four factors, and the horizontal axis represents the combination of the “Mid-low factor” and the “Mid-high factor”. Panel (b) represents the enlargement of obstruents in panel (a). 82

Chapter 1. Introduction: speech sounds

With the development of human nature and human society, language has become the most important communication tool. From the babbling stage of infancy, we learn to use language sounds to express our needs and feelings. The human vocal tract can produce thousands of audibly distinct sounds. While some phonemes, such as the “t” sound, are found in almost every language, each language typically has its own set of speech sounds that are used to form words. The specific inventory of distinct speech sounds in a language is usually limited to a relatively small number, typically no more than a few dozen (Hayes, 2008). The smallest unit of sound that can be used in a language to distinguish different words, is called the phoneme (e.g., Spencer, 1996). For example, “big” /bɪg/ and “bag” /bæg/ are two separate English words that are distinguished by the substitution of the phoneme /ɪ/ for another phoneme /æ/. In Mandarin Chinese, due to the substitution of the phoneme /n/ for another phoneme /ŋ/, “林” /lín/ and “玲” /líng/ indicate two Chinese characters (ideograms) that distinguish meanings. In different languages, there are many phenomena in phonology where words have different meanings by changing phonemes like this.

This dissertation investigates the acoustic correlates of English phonemes using multivariate analyses. In order to provide a background to the research, basic features of English speech will be introduced below: phonology (section 1.1), sonority (section 1.1.1), and syllable structure (section 1.1.2).

1.1 Phonology

Phonology is a branch of linguistics that deals with the linguistic patterning of sounds. It explores how phoneme sounds are used and organized in particular languages, and how it contributes to the structure and meaning of the word. Phonemes are mainly divided into vowels and consonants in different languages (Jones, 2011). For example, the phonemes of British English (BBC pronunciation) are generally described as in Table 1.1 (Jones, 2011). Vowels in British English are divided into seven short vowels (/ɪ/, /e/, /æ/, /ʌ/, /ɒ/, /ʊ/, /ə/), five long vowels (/i:/, /ɑ:/, /ɔ:/, /u:/, /ɜ:/), and eight diphthongs (/eɪ/, /aɪ/, /ɔɪ/, /əʊ/, /aʊ/, /ɪə/, /eə/, /uə/). Consonants in English are divided into two liquids (/l/, /r/), two glides (/j/, /w/), three nasals (/m/, /n/, /ŋ/), two affricatives (/tʃ/, /dʒ/), nine fricatives (/f/, /v/, /θ/, /ð/, /s/, /z/, /ʃ/, /ʒ/, /h/), and six plosives (/p/, /t/, /k/, /b/, /d/, /g/). The pronunciation of vowels and consonants will be discussed in detail in section 1.1.1. Vowels are “voiced”, in that they include vocal fold vibration, while consonants can be either voiced or unvoiced.

Table 1.1. The phonemes of British English (BBC pronunciation) according to the International Phonetic Alphabet (Jones, 2011).

Vowels	Short vowels								
	ɪ	e	æ	ʌ	ɒ	ʊ	ə		
	kit	dress	trap	strut	lot	foot	other		
	Long vowels								
	iː	ɑː	ɔː	uː	ɜː				
	fleece	palm	thought	goose	nurse				
	Diphthongs								
	eɪ	aɪ	ɔɪ	əʊ	aʊ	ɪə	eə	ʊə	
	face	price	choice	goat	mouth	near	square	cure	

Consonants	Liquids								
	l	r							
	light	right							
	Glides								
	w	j							
	wet	yet							
	Nasals								
	m	n	ŋ						
	sum	sun	sung						
	Affricatives								
	tʃ	dʒ							
	chain	Jane							
	Fricatives								
	f	v	θ	ð	s	z	ʃ	ʒ	h
	fine	vine	think	this	seal	zeal	ship	measure	how
	Plosives								
	p	t	k	b	d	g			
	pin	tin	kin	bin	din	give			

1.1.1 Sonority

A linguistic concept called “sonority” plays an essential role in developing phonological patterns in speech. Sonority has been difficult to define, but it can be loosely defined as the relative loudness, perceptibility, or acoustic intensity of sounds (Harris, 2006). In spite of the fact that sonority has not yet been clearly defined, the sonority concept has been applied to various aspects of speech: as a feature of stress or accent carried by certain syllables, as a feature of voice quality, and as related to the internal structure of syllables (Price, 1980; Wedin et al., 1978; Féry, 2003). Sonority has been described mainly in terms of its assumed relation to speech sounds. Researchers have assumed a sonority hierarchy and a Sonority Sequencing Principle, as discussed below.

It is considered that the sonority of each phoneme has its inherent position in a sound hierarchy. This hierarchy arranges phonemes based on their sonority levels. In general, vowels (such as /u/, /e/, /o/) are highly sonorous, while consonants (such as /p/, /s/, /r/) are lower sonorous and less perceptible (Hayes, 2008). In English, vowels (such as /ɪ/ in the English word “big” /bɪg/) hold the highest position in the sonority hierarchy, followed by sonorant consonants (such as /w/ in the English word “well” /wel/). Obstruents (such as /t/ in the English word “ten” /ten/), which involve obstructing airflow, occupy the lowest position in the hierarchy. The ordinal hierarchy of sonority has been proposed many times in linguistics. An overview of this hierarchy (about British English) is given in Table 1.2 (Nakajima et al., 2017). One of the classic examples is the hierarchy of “aperture”, which is very similar to sonority, proposed by de Saussure (1959). It has seven aperture levels. Selkirk (1984) claims that there are

eleven sonority levels, while Harris (1994) condenses the sonority levels into six. Spencer (1996) also claims that there are six sonority levels, but points out that vowels and glides are separated in the six categories.

Table 1.2. Previously proposed sonority (aperture) scales (from Nakajima et al., 2017).

Sonority hierarchy	Proposed category	Corresponding English phonemes
de Saussure (1916/1959)		
6	a	/æ, ɑ:, ʌ/
5	e, o, ö	/e, ɔ, ɔ:/
4	i, u, ü	/ɪ, i:, ʊ, u:, j, w/
3	Liquids	/l, r/
2	Nasals	/m, n, ŋ/
1	Fricatives	/θ, ð, f, v, s, z, ʃ, ʒ/
0	Occlusives	/p, t, k, b, d, g/
Selkirk (1984)		
10	a	/æ, ɑ:, ʌ/
9	e, o	/e, ɔ, ɔ:/
8	i, u	/ɪ, i:, ʊ, u:/
7	r	/r/
6	l	/l/
5	m, n	/m, n/
4	s	/s/
3	v, z, ð	/v, z, ð/
2	f, θ	/f, θ/
1	b, d, g	/b, d, g/
0.5	p, t, k	/p, t, k/
Harris (1994)		
5	Low vowels	/æ, ɑ:, ʌ, e, ɔ, ɔ:/
4	High vowels and glides	/ɪ, i:, ʊ, u:, j, w/
3	Liquids	/l, r/
2	Nasals	/m, n, ŋ/
1	Fricatives	/θ, ð, f, v, s, z, ʃ, ʒ, h/
0	Plosives	/p, t, k, b, d, g/
Spencer (1996)		
5	Vowels	/æ, ɑ:, ʌ, e, ɔ, ɔ:, ə, ə:, ɪ, i:, ʊ, u:, aɪ, aʊ, eə, ei, oi, əʊ, ɪə, ʊə/
4	Glides	/j, w/
3	Liquids	/l, r/
2	Nasals	/m, n, ŋ/
1	Fricatives and affricates	/θ, ð, f, v, s, z, ʃ, ʒ, h, tʃ, dʒ/
0	Plosives	/p, t, k, b, d, g/

For example, the English word “sun” /sʌn/ is composed of three phonemes: /s/, /ʌ/, and /n/. The vowel /ʌ/ has a louder sound and is easier to perceive in the word. Based on the sonority hierarchy, the pattern of sonority shows a mountain-like trend, changing from a less sonorous sound /s/ to a more sonorous sound /ʌ/, and back to a less sonorous sound /n/. Note that the sonority hierarchy is closely related to the workings of the vocal organs. The difference between vowels and consonants is mainly that the vocal cords vibrate when pronouncing vowels (such as /æ/, /ə/, /e/, /aɪ/), and that the airflow goes freely through the mouth without any obstruction, making the sound often louder and more sonorant. When producing less sonorant sounds, such as consonants (e.g., /p/, /b/, /v/), the vocal organs maintain a relatively narrow space, suppress sounds, and airflow is obstructed when a certain part of the mouth vibrates (Spencer, 1996). Generally, by using sonority as a guiding principle, linguists can analyze and describe the organization of sounds or phonemes in various languages and understand how they interact within syllables and other linguistic structures.

1.1.2 Syllable structure

Syllables play a prominent role in the pronunciation and perception of speech sounds. Words are composed of syllables, and syllables are the basic building blocks of spoken language. Syllables can typically consist of a vowel sound or a vowel sound along with one or more surrounding consonant sounds. A word can have one or more syllables, and understanding syllables can assist in pronunciation and spelling (Harris, 1994). For example, in English, “a” /ə/ is a one-syllable word (the article “a” used before a noun), even if it consists only of a single vowel; “happy” /hæpi/ is a two-

syllable word, which consists of /hæp/ and /i/ (“Cambridge English Pronouncing Dictionary” as a pronunciation reference, Jones, 2011). Breaking down words into syllables can aid in pronunciation, reading, and decoding new words. In a typical syllable, there are three parts: the onset (or beginning), the nucleus (or peak), and the coda (or end). The nucleus is the most essential part of a syllable and can even exist on its own, as seen in words like “account” where the first syllable “a” consists only of a vowel as the nucleus.

The Sonority Sequencing Principle (SSP) refers to a phonological principle that governs the arrangement of sounds within a syllable and outlines the structure of a syllable in terms of sonority. It states that in a syllable, sounds with higher sonority tend to appear towards the middle (or peak), while sounds with lower sonority tend to occur towards the edges (Harris, 1994). Generally speaking, the principle shows that sounds that are higher in the sonority hierarchy (see section 1.1.1), such as vowels and sonorant consonants (glides, liquids, and nasals) are more likely to be found closer to the nucleus or peak of a syllable. Sounds that are lower in the sonority hierarchy, like fricatives, affricates, and plosives, are more likely to occur towards the edges or boundaries of a syllable. For example, in the one-syllable English word “top” /tɒp/, the nucleus is a vowel /ɒ/, while the onset /t/ and coda /p/ are typically composed of consonants. Therefore, obstruents are typically considered syllable boundaries, while vowels always have a place at a syllable nucleus (Harris, 1994).

The SSP helps to explain the natural patterns in sound segments observed in English. It provides a framework for understanding the reason that certain sound segments are preferred or disallowed within a syllable (Zsiga, 2013; Nakajima et al.,

2017). Consider the one-syllable English word “print” (following an example in Zsiga, 2013), which consists of five phonemes as /prɪnt/. Among them, the vowel /ɪ/ has the highest sonority and is the syllable nucleus (peak). The obstruents /p/ and /t/ have the lowest sonority, and are farther from the syllable nucleus than sonorant consonants /r/ and /n/. If these phonemes are rearranged and the sounds create a non-existing word like “rpɪtn” (/rpɪtn/), it would violate the SSP. In this hypothetical word, the sonority hierarchy is changed and broken, as the obstruent /p/ appears after the sonorant /r/ and the obstruent /t/ appears before the nasal /n/, so that the less sonorant phonemes are now closer to the syllable nucleus /ɪ/. This rearrangement goes against the typical sonority pattern observed in English. By following the SSP, a balance can be achieved between ease of articulation and perceptual clarity in the syllable structures.

Some syllables in English may have exceptions that deviate from the general SSP by allowing complex consonant clusters at the beginning or end of a syllable, such as the consonant cluster “st” in the English word “stop” /stɒp/. The position of “s” in the sonority hierarchy is higher than that of the second consonant /t/, but /s/ cannot be considered as a syllable nucleus. /stɒp/ is a one-syllable word, and the syllable nucleus is /ɒ/. A possible explanation is that the SSP might be concerned with the overall simplicity (or optimality) of the syllable in which the cluster is contained. Among this special phenomenon, Ohala (1999) hypothesized that a consonant omitted from a consonant cluster is the one that least conforms to the shape of an optimal syllable. She performed a study of consonant cluster reduction with children under 3 years old, with normal language development. The results supported the hypothesis that children

always reduced consonant clusters to whichever consonant would result in the least complex syllable as defined by sonority, e.g., “stig” /stɪg/ becomes /tɪg/.

1.2 Multivariate analysis of speech sounds

Before moving on to topic of this dissertation, the acoustic correlates of English phonemes using multivariate analysis, it is important to introduce how humans process the frequency spectra of speech and other sounds. The human hearing system has the ability to perceive a wide range of frequencies. Human hearing covers a range of frequencies within the audible range, of which the minimum value is typically 16 or 20 Hz, while the maximum is 16 or 20 kHz. Frequency components below 20 Hz are almost impossible to perceive (Møller & Pedersen, 2004). In order to survive and to communicate, humans must adequately process the frequency spectrum of speech and other sounds. In an early attempt to understand the frequency resolution of the hearing system, Zwicker (1961) introduced the bark scale, which divides the audible frequency range into 24 distinct frequency bands ranging from 20 to 15500 Hz. These 24 frequency bands are commonly referred to as critical bands, acting as a set of filters through which sounds are processed (Smith et al., 1999). When it comes to speech sounds, it is desirable to simulate the frequency resolution of the human ear in the normal speech frequency range within 7000 Hz by using 20 frequency bands (Ueda et al., 2017; Nakajima et al., 2018). However, it is worth noting that even fewer frequency bands appear to be sufficient for synthesizing speech that is reasonably intelligible. For example, Ueda et al. (2018) conducted three perceptual experiments with 4-band noise-vocoded speech consisting of Japanese sentences. The results showed that very

high intelligibility (92-94%) of mora recognition was achieved even without any extensive training.

Importantly, previous studies on this issue proceeded from a series of multivariate analyses. Generally speaking, multivariate analysis is being used to reduce and summarize large data sets into several easier factors (Johnson & Wichern, 2007), and has been used in a few previous studies to investigate acoustic features of speech sounds. Early research on the acoustic properties of speech sounds involved principal component analyses and listening experiments. Principal component analysis (PCA) refers to a statistical method that reduces the dimensions of the data set and extracts the main feature components from it, while retaining the maximum amount of information for some data sets with a large number of dimensions or features (Jolliffe, 2002; Jolliffe & Cadima, 2016). Plomp et al. (1967) determined the frequency spectra of 15 Dutch vowels according to 18 band-pass filters. By performing principal component analysis, they observed that the first two principal components of these vowels were closely linked to the formants of the vowels. Furthermore, these two primary principal components could account for most of the phonological features present in the 15 Dutch vowels (Plomp et al., 1967; Pols et al., 1973; Plomp, 1976; Plomp, 2001). In subsequent research using PCA, Zahorian and Rothenberg (1981) showed that English speech sounds synthesized from four principal components were found to be about 80% intelligible, so PCA could potentially reveal sufficient information to represent continuous speech.

In addition, some studies on the acoustic properties of speech also involved the use of factor analysis, which is the other common multivariate statistical analysis

method. Factor analysis is also used to simplify a complex data set to explore the basic dimensions that explain the relationships between multiple variables (Horn, 1965). One of the most recent studies using factor analysis was from Ueda and Nakajima (2017). Based on the PCA method popularized by Zahorian and Rothenberg (1981), Ueda et al. (2010) and Ueda and Nakajima (2017) expanded the analyzed speech samples to the complete spoken sentences of eight languages/dialects, i.e., American English, British English, Cantonese, French, German, Japanese, Mandarin, and Spanish (analysis samples were from NTT-AT, 2002). The power fluctuations from the 20 critical-band filters of speech sounds were submitted to a PCA, and a varimax rotation led to three or four spectral factors. Varimax rotation is used to make any variable have a high contribution rate on only one factor, while the factor loadings on other factors are almost zero (Kaiser, 1958), if the data fit this assumption. For example, if the eigenvectors of the first three or four principal components are rotated, three or four power fluctuation factors will be produced. The results of Ueda et al. (2010) and Ueda and Nakajima (2017) showed that three or four spectral factors universally appeared in speech samples of eight languages/dialects. The factors were related to four frequency ranges: around 50–540, 540–1700, 1700–3300, and above 3300 Hz. Three spectral factors appeared: the “Low & mid-high factor”, related to the two separate frequency ranges of 50–540 and 1700–3300 Hz, the “Mid-low factor”, related to the frequency range of 540–1700 Hz, and the “High factor”, related to the frequency range above 3300 Hz.

Proceeding from the findings of the analysis above, Kishida et al. (2016) investigated the contributions of the number of spectral factors (power-fluctuation

factors) to speech intelligibility through a listening experiment. British-English sentences (200 sentences spoken by 5 male native speakers of English), Japanese sentences (200 sentences spoken by 5 male native speakers of Japanese), and Mandarin-Chinese sentences (78 sentences spoken by 5 male native speakers of Chinese) were selected to check the consistency of a new analysis method, and some Japanese sentences were resynthesized for a listening experiment to measure intelligibility. They developed a new multivariate analysis method, which is called the “origin-shifted principal component analysis”. This analysis method was the first step of factor analysis; the new PCA was followed by varimax rotation. They further added cepstral analysis. The origin shift referred to the fact that the origin to calculate variances was shifted from the gravity center of the data points to the silent point, at which variates (powers) are zero. Without such an origin shift, a continuous noise would be generated, in most cases, from the non-zero point onto which the original silent point would be mapped. This non-zero point is usually close to the silent point, but not to the gravity center. It is thus desirable to include the silent point in the subspace of the principal components. This analysis method ensured that, if speech sounds were to be resynthesized, they would not be accompanied by a steady background noise caused by the data reduction procedure. Furthermore, to extract the spectral envelopes of the speech signals, cepstral liftering was applied. Cepstral liftering is a smoothing technique for signal spectra applied to speech analysis (Rabiner & Schafer, 1978; Awan et al., 2012). Kishida et al. (2016) thus modified the first stage of the method of factor analysis to obtain 1 to 9 principal components suitable for resynthesizing speech sounds as 20-critical-band noise-vocoded speech. They

confirmed that there were indeed three or four common factors in spoken British English, Japanese, and Mandarin Chinese. Using three or more spectral factors for speech resynthesis was sufficient to retain fundamental language information in the noise-vocoded speech (in Japanese).

Nakajima et al. (2017) analyzed spectral changes for 200 sentences that were spoken by three English native speakers (two female and one male) by utilizing factor analysis. The speech samples were taken from an English speech database, “The ATR British English Speech Database” (Campbell, 1993). In this research, all English phonemes were divided into three general categories to be discriminable in the factor space, which were the same as in phonological categories, i.e., vowels, sonorant consonants, and obstruents. As in Ueda and Nakajima (2017), three spectral factors appeared: the “Low & mid-high factor”, related to the two separate frequency ranges around 300 Hz and 2200 Hz, the “Mid-low factor”, related to the frequency range around 1100 Hz, and the “High factor”, related to the frequency range above 3300 Hz. The results showed that the distribution of factor scores of English obstruents got the highest position on the “High factor”, which means it was associated closely with obstruents. The distributions of factor scores of English vowels, sonorant consonants, and obstruents, respectively, went from a high to a low position on the “Mid-low factor”. This tendency was highly related to the sonority hierarchy in phonology (for more detailed explanation see section 1.1.1), ranking vowels as highest, followed by sonorant consonants, and obstruents as lowest. Together, this made the study by Nakajima et al. (2017) one of the rare studies to connect acoustic features in speech to sonority. The authors suggested that the phonological implications of the spectral

factors seem important to understanding how English speech sounds carry linguistic information. Overall, these preceding studies have shown that multivariate analyses can offer a viable and potent method for analyzing speech phonology, and this research formed the basis for the present studies on speech sonority described in this dissertation.

It is worth noting that Nakajima et al. (2017) utilized the ATR English speech database, with a total number of 31663 phonemes labeled. However, to maintain the correct phoneme labels for the samples in factor analysis, they had to discard some labels. For example, they discarded unnecessary labels of closure periods, which were nearly silent, in some plosives. The labels of some diphthongs had to be modified. Careful inspection of this database revealed 7523 mislabeled phonemes, which were consequently left out of the analysis of Nakajima et al. (2017). The unsuitable recording quality of the speech sounds of the database was also a problem. To further explore the relationship between acoustic characteristics and phonology in English, it was necessary to use a high-quality English database with correct phoneme labels.

1.3 Purpose and contents of this dissertation

The general purpose of the dissertation is to elucidate how acoustic features of English speech can be expressed using multivariate analyses of changing spectra. Following this Chapter 1, which describes basic terminology on phonological aspects of speech sounds and previous research on multivariate analysis of speech spectra, there will be four more chapters.

There are three research purposes of this dissertation. The first purpose is to investigate the suitability of factor analyses as applied to English phonemes. To investigate this, Chapter 2 describes a study in which factor analysis methods were compared. Here, it was determined in detail whether the results of the normal factor analysis (Nakajima et al., 2017) and origin-shifted factor analysis (Kishida et al., 2016) would differ or not when applied to English speech. Furthermore, the acoustic correlates of English phonemes, especially English obstruents, were investigated.

The second purpose is to determine the acoustic features of English phonemes in general, by analyzing the results of the aforementioned factor analyses. Chapter 3 describes a study in which a three-factor analysis of English speech was performed. To examine the acoustic features of the spectral power fluctuations in English phonemes more precisely than before, a high-quality English speech database was created. The database was named the “Fukuoka-Galway English database”, with 100 sentences spoken by three native speakers of English. In Chapter 3, following the results of Chapter 2, origin-shifted factor analysis was used. The results showed three categories of English phonemes (vowels, sonorant consonants, obstruents) that exhibited the acoustic features of the three factors. Following the results of Chapter 3, Chapter 4 describes a study in which a four-factor analysis of English speech was performed. By using origin-shifted factor analysis, the English speech sentences from the “Fukuoka-Galway English Database” were further analyzed, and the features of the four spectral factors were observed. The purpose was to investigate how the three categories of English phonemes (vowels, sonorant consonants, obstruents) exhibited the acoustic features of the spectral power fluctuations on these four factors.

The third purpose of this dissertation is to obtain clear insight into whether and how obstruents would appear in the factor spaces. The details of the acoustic features of English obstruents were also described in Chapter 2 to Chapter 4. Chapter 5 describes the general discussion and conclusions of this dissertation. The results of Chapter 2 to Chapter 4 were summarized and future potential developments of the present research were discussed.

Chapter 2. Study 1: Comparison of multivariate analysis methods as applied to English speech

2.1 Purpose

The main purpose of Study 1 was to determine in detail whether the results of the normal factor analysis (Ueda and Nakajima, 2017; Nakajima et al., 2017) and origin-shifted factor analysis (Kishida et al., 2016) differ or not when applied to English speech. Four different factor analyses were used for this purpose: (a) Normal factor analysis without cepstral liftering, similar to Ueda and Nakajima (2017) and Nakajima et al (2017); (b) Origin-shifted factor analysis without cepstral liftering; (c) Normal factor analysis with cepstral liftering; (d) Origin-shifted factor analysis with cepstral liftering, following Kishida et al. (2016). These factor analyses were used for analyzing 200 sentences spoken by three native speakers of English. The sentences were taken from the “The ATR British English Speech Database” (Campbell, 1993).

Another research purpose in this Study 1 was to determine the acoustic correlates of English phonemes by using the aforementioned factor analyses, specifically focusing on how English vowels, consonants, and obstruents appear in the obtained factor score spaces, in order to compare the results with the previous research of Nakajima et al. (2017).

2.2 Method

2.2.1 Speech samples

To compare the results of origin-shifted factor analysis with the results of the normal factor analysis, the same 200 English sentences as in Nakajima et al. (2017) were used as analysis samples. The sentences were spoken by one male and two female native speakers of English, taken from “The ATR British English Speech Database” (Campbell, 1993). The speech sentences were recorded with a sampling frequency of 12000 Hz and 16-bit linear quantization. The speech signals of all the spoken sentences in the database were manually segmented into individual phonemes and were labeled utilizing the Machine Readable Phonetic Alphabet (MRPA) (Lennig and Brassard, 1984). A total number of 31663 “phonemes” were labeled, but to maintain consistency with Nakajima et al.’s study (2017), in which the phoneme labeling of this database had to be reexamined, 7523 were omitted, and the same 24140 English phonemes were taken up as analysis samples.

2.2.2 Procedure

First, the same analysis procedure was performed as in Kishida et al. (2016). As the initial step, the speech samples (at the sampling frequency of 12000 Hz) were resampled every 1 ms with a 30-ms Hamming window. Every 30-ms-long segment of the power spectra was obtained by the Fastest Fourier Transform in the West (FFTW, Frigo & Johnson, 2020). The frequency resolution of FFTW was around 33 Hz ($1/0.03 = \sim 33.3$ Hz). The power spectrum was smoothed with a 5-ms short-pass lifter by cepstral liftering (for more detailed explanation see section 1.2). Different from

Kishida et al. (2016), in the present analysis, the smoothed power spectra were divided into 19 critical bands, and averaged power was calculated for each band. Following the Nyquist theorem (Nyquist, 1928), only when the sampling frequency was higher than double that of the highest frequency of the speech signal, can we correctly process the original signal bands from the sample signal. The speech sounds of the database had a 12000-Hz sampling frequency, and the highest possible frequency that can be processed was 6000 Hz. It was decided to set the highest frequency range of the analysis to 5300 Hz. Nineteen critical-band filters were constructed to cover the frequency range of 50–5300 Hz (Table 2.1). Their center frequencies ranged from 75–4800 Hz. The frequency bands were adopted from Zwicker and Terhardt (1980) and Kishida et al. (2016). The frequency range below 50 Hz was not utilized because it should have been less relevant to the speech signal. The first frequency band was from 50 to 100 Hz, and the center frequency was 75 Hz.

Table 2.1. Division of speech signals into 19 critical-band filters, as used for the analyses. The critical band number, center frequency (Hz) and the passband of each critical band filter are shown with reference to Zwicker and Terhardt (1980) and Kishida et al. (2016).

Band No.	Center frequency [Hz]	Passband [Hz]
1	75	50 - 100
2	150	100 - 200
3	250	200 - 300
4	350	300 - 400
5	450	400 - 510
6	570	510 - 630
7	700	630 - 770
8	840	770 - 920
9	1000	920 - 1080
10	1170	1080 - 1270
11	1370	1270 - 1480
12	1600	1480 - 1720
13	1850	1720 - 2000
14	2150	2000 - 2320
15	2500	2320 - 2700
16	2900	2700 - 3150
17	3400	3150 - 3700
18	4000	3700 - 4400
19	4800	4400 - 5300

Thus, 19 powers for the 19 filters were sampled at every 1 ms to perform PCA followed by varimax rotation (Kaiser, 1958). In the origin-shifted factor analysis, the origin of the factors, which were represented as orthogonal vectors in the 19-dimensional variate space, was shifted from the gravity center of the data points to the silent point, at which all variates (powers) are zero. This means that a long enough silent part of the speech signal was represented always at the silent point. This should be suitable for observing the nature of relatively weak sounds, which are closer to the silent point. This was not the case in Nakajima et al. (2017), in which a normal factor analysis was used. Finally, only the temporal middle point of each uttered phoneme was left in the obtained factor space.

2.3 Results

2.3.1 Factor loadings of the three spectral factors

From the perspective of the analysis processes, the major differences between the two previous methods, i.e., the factor analyses in Nakajima et al. (2017) and in Kishida et al. (2016), were whether to use the cepstral liftering and whether to use the origin shift to the silent point. Figure 2.1 shows the factor loadings for all the English speech samples spoken by the three native speakers of English, obtained with each of the four analysis methods. Panels (a) and (b) show the factor loadings between origin-shifted factor analysis and the normal factor analysis without cepstral liftering. The cumulative contributions of the three spectral factors were around 45%. Panels (c) and (d) show the factor loadings between origin-shifted factor analysis and the normal factor analysis with cepstral liftering. The cumulative contributions of the three

spectral factors were around 43%. One division of the horizontal axis corresponds to 0.5 critical bandwidth.

Four main frequency bands were obtained. The first constituted a low-frequency band, approximately from 50 to 600 Hz. The second was a mid-low frequency band, approximately from 600 to 1700 Hz. The third was a mid-high-frequency band, approximately from 1700 to 3000 Hz. Finally, the fourth range was a high-frequency band, approximately above 3000 Hz. Confirming the results of Ueda and Nakajima (2017), Nakajima et al. (2017), and Kishida et al. (2016), these four frequency bands were related to the three spectral factors in Figure 2.1. One factor, the “Low & mid-high factor” (red line) was double-peaked in that the frequencies of relatively high loadings were around 300 Hz and around 2300 Hz. The second factor, the “Mid-low factor”, was located around 1100 Hz (black line). The third factor, the “High factor”, was located around 4100 Hz (blue line).

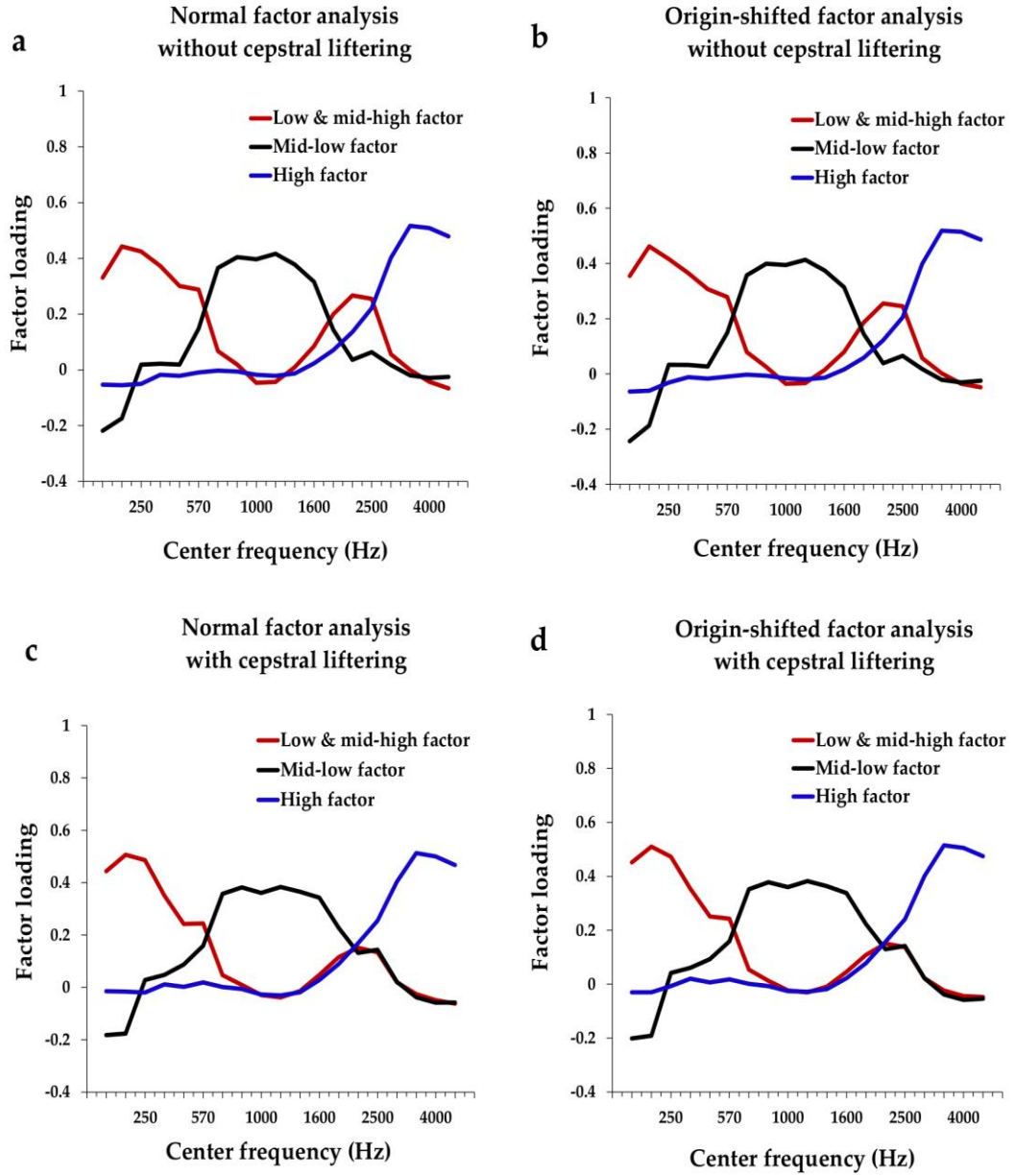


Figure 2.1. Factor loadings of the three extracted spectral factors of 200 English speech samples from three native speakers. Four different factor analyses were performed: (a) Normal factor analysis without cepstral liftering, similar to Ueda and Nakajima (2017) and Nakajima et al. (2017); (b) Origin-shifted factor analysis without cepstral liftering; (c) Normal factor analysis with cepstral liftering; (d) Origin-shifted factor analysis with cepstral liftering, following Kishida et al. (2016).

To observe whether using only the origin shift would affect the results of the factor loadings, by comparing Figures 2.1 (a) and (b), or by comparing Figures 2.1 (c) and (d), it can be concluded that the origin shift did not greatly affect the factor loadings. The similarity in the factor loadings between Figures 2.1 (a) and (b) or between Figures 2.1 (c) and (d) confirmed that the origin shift keeps the essential features of the factors in the analyses.

As to whether the use of cepstral liftering alone would affect the results of the factor loadings, by comparing Figures 2.1 (a) and (c), or by comparing Figures 2.1 (b) and (d), it can be concluded that the cepstral liftering affected the factor loadings only for the “Low & mid-high factor”. In particular, the second peak (around 2300 Hz) of the double-peaked “Low & mid-high factor” was more prominent when the cepstral liftering was not used, as can be seen in Figures 2.1 (a) and (b).

2.3.2 Factor loadings of the speech from female and male speakers

To investigate whether factor loading patterns are affected by the gender of the speakers, the speech of the two female speakers and that of the male speaker were analyzed separately. Figure 2.2 shows that the same three spectral factors were obtained from the two female speakers as in panel (a), as well as from the male speaker as in panel (b), utilizing origin-shifted factor analysis with cepstral liftering. The factor loadings for the male speaker tended to be higher on the “Low & mid-high factor” and the “Mid-low factor”.

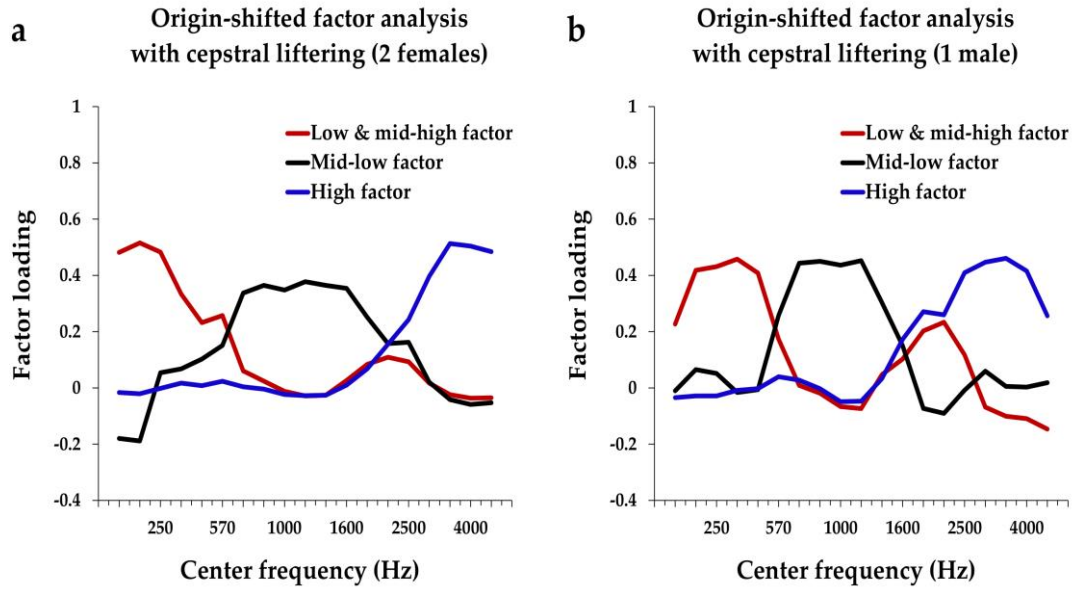


Figure 2.2. Factor loadings of the three extracted spectral factors from the speech of two female native speakers in panel (a), and from one male speaker in panel (b) by origin-shifted factor analysis with cepstral liftering.

To examine the influence of cepstral liftering on the factor loadings for the male speaker, the speech of the speaker was analyzed with the origin-shifted factor analysis with and without cepstral lifting. Figure 2.3 shows factor loadings of the three spectral factors for this speaker. The factor loadings without cepstral liftering were higher at the second peak of the “Low & mid-high factor”. This result was consistent with the results of the factor loadings of all three speakers (Figure 2.1). Thus, whether or not cepstral liftering was used affected the factor loadings of the second peak of the “Low & mid-high factor”, which was located around 2300 Hz.

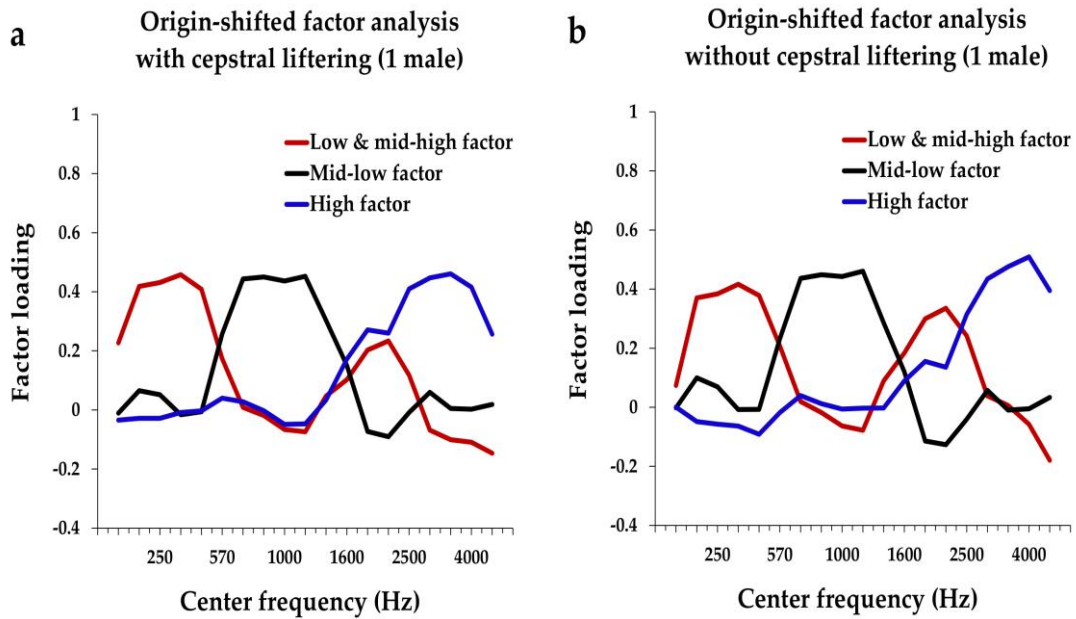


Figure 2.3. Factor loadings of the three extracted spectral factors from one male native speaker. Two analyses were performed: (a) Origin-shifted factor analysis with cepstral liftering (the same as Figure 2.2b), and (b) Origin-shifted factor analysis without cepstral liftering.

2.3.3 Factor scores of the three spectral factors

Following the analysis of the factor loadings, the factor scores were analyzed. In Figure 2.4, Figure 2.5, Figure 2.6, and Figure 2.7 below, panels (a), (b), and (c) show the distributions of combinations of the three spectral factor scores. The distributions were between the “Low & mid-high factor” (with the frequency loadings around 300 Hz and 2300 Hz), the “Mid-low factor” (with the frequency loadings around 1100 Hz), and the “High factor” (with the frequency loadings around 4100 Hz). Panel (d) in Figure 2.4 to Figure 2.7 shows the distribution in the three-dimensional configuration as viewed in a direction from above-right to below-left in panel (a). In panel (d), the horizontal axis represents the combination of the “Mid-low factor” and the “High factor”, by calculating $(x-y)/\sqrt{2}$, where “x” indicates the factor score of the

“Mid-low factor”, and “y” the factor score of the “High factor”, following Nakajima et al. (2017). All the factor scores of English phonemes were separately distributed in three categories, i.e., vowels (orange circles), sonorant consonants (blue circles), and obstruents (black crosses).

Normal factor analysis without cepstral liftering

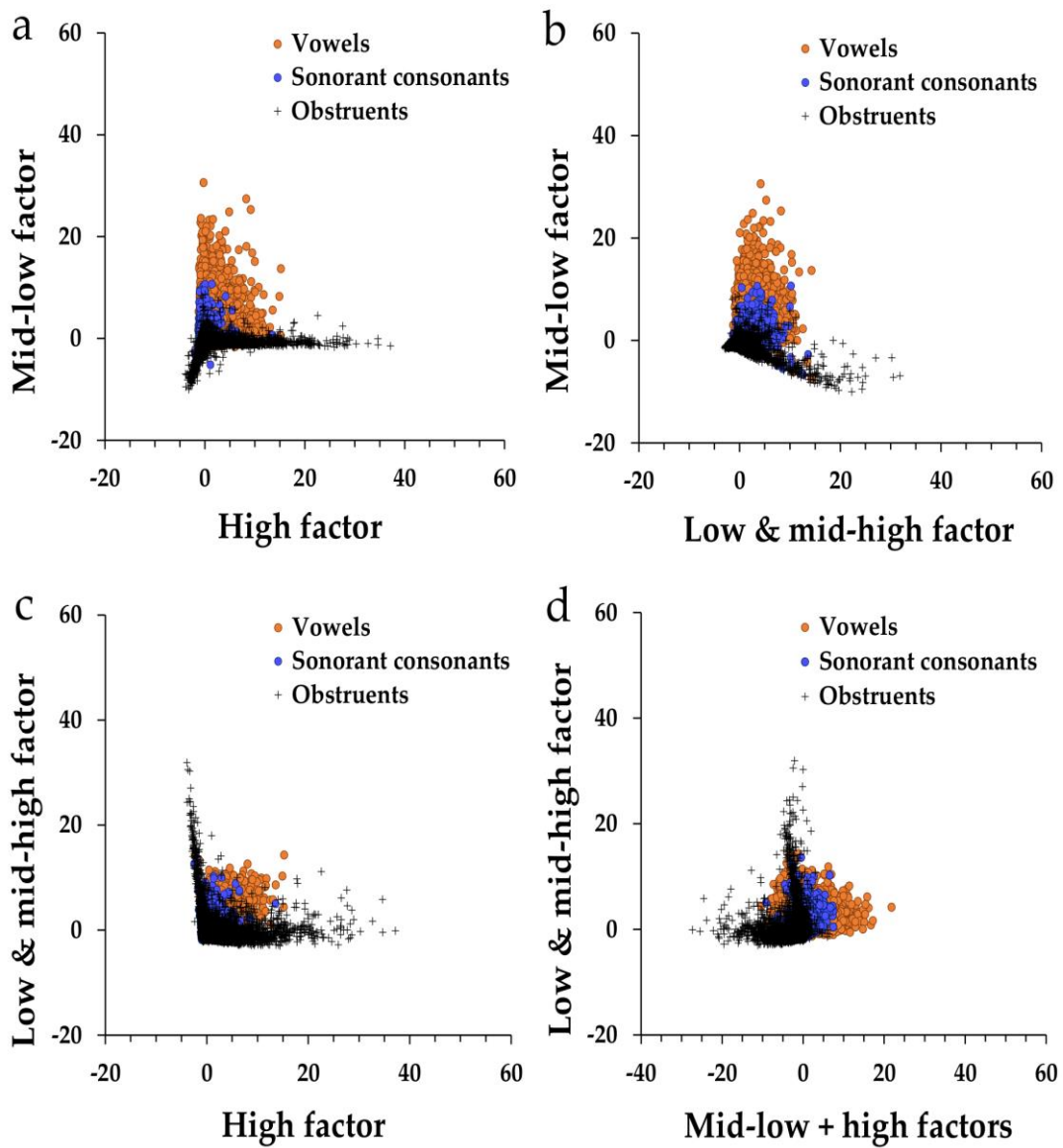


Figure 2.4. The distribution of factor scores by normal factor analysis without cepstral liftering, similar to Nakajima et al (2017). (a–c) represent all possible combinations of two factors out of the three factors, and (d) shows a two-dimensional mapping of the three factors.

Origin-shifted factor analysis without cepstral liftering

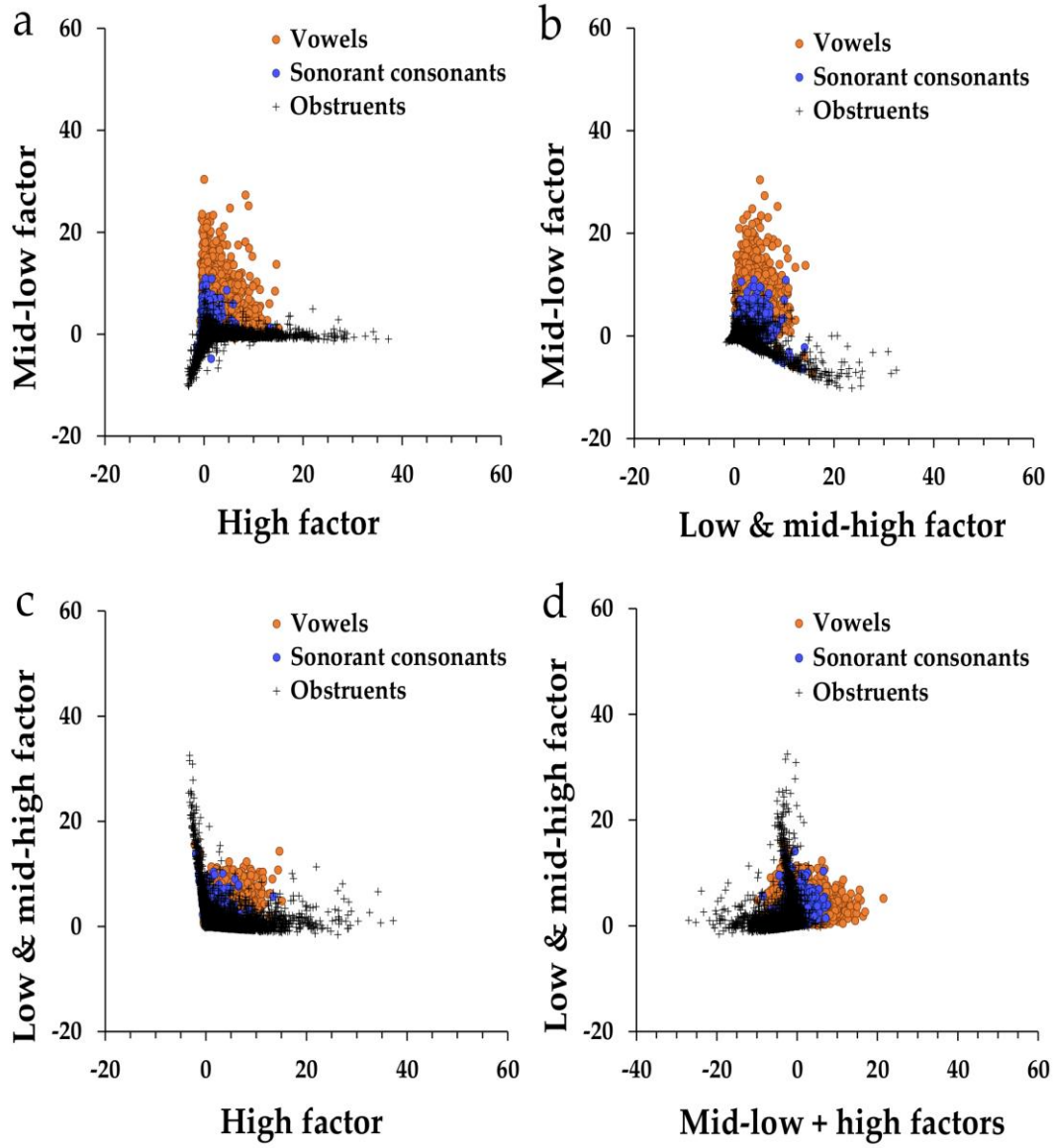


Figure 2.5. The distribution of factor scores by origin-shifted factor analysis without cepstral liftering. (a–c) represent all possible combinations of two factors out of the three factors, and (d) shows a two-dimensional mapping of the three factors.

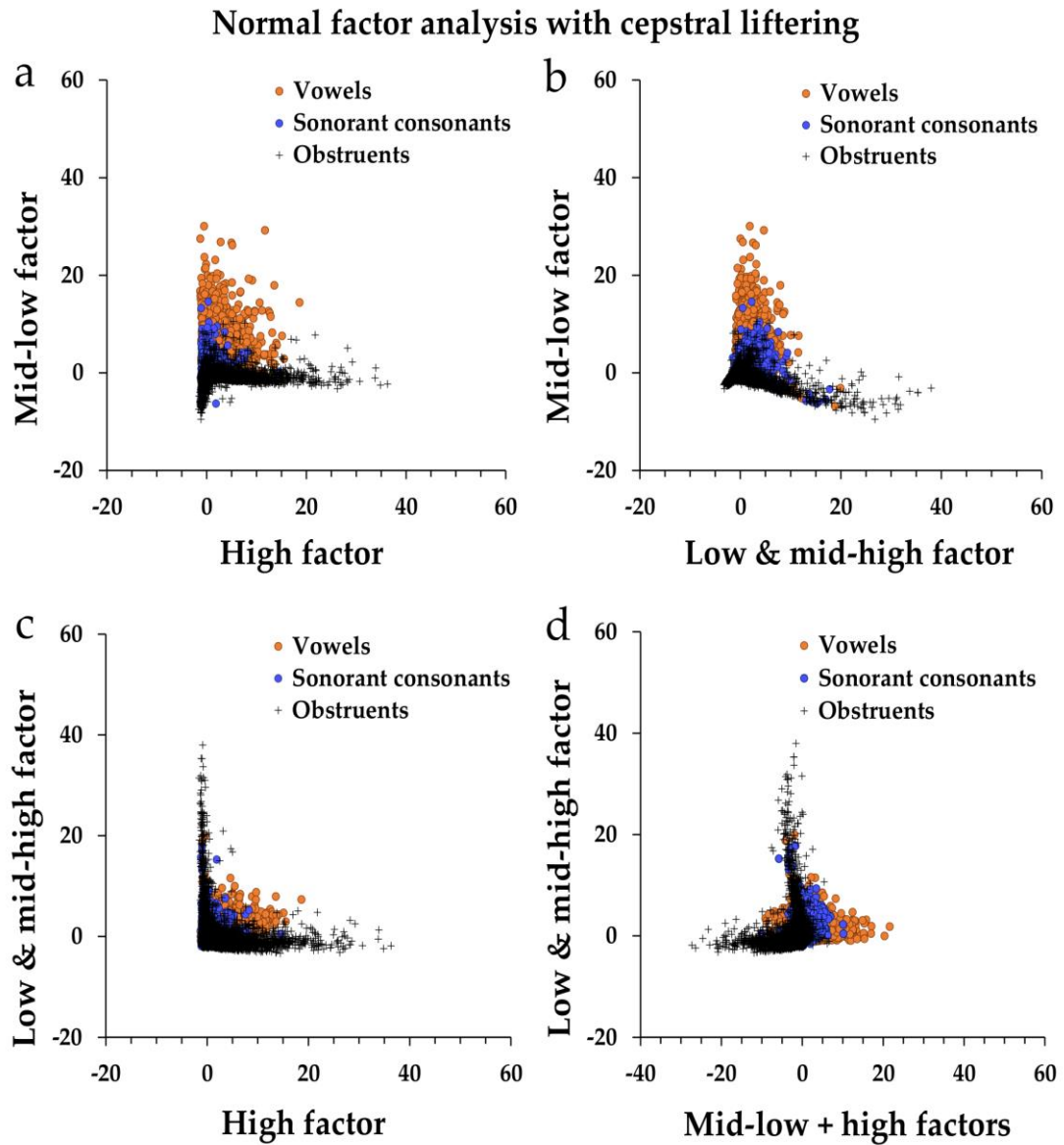


Figure 2.6. The distribution of factor scores by normal factor analysis with cepstral liftering. (a–c) represent all possible combinations of two factors out of the three factors, and (d) shows a two-dimensional mapping of the three factors.

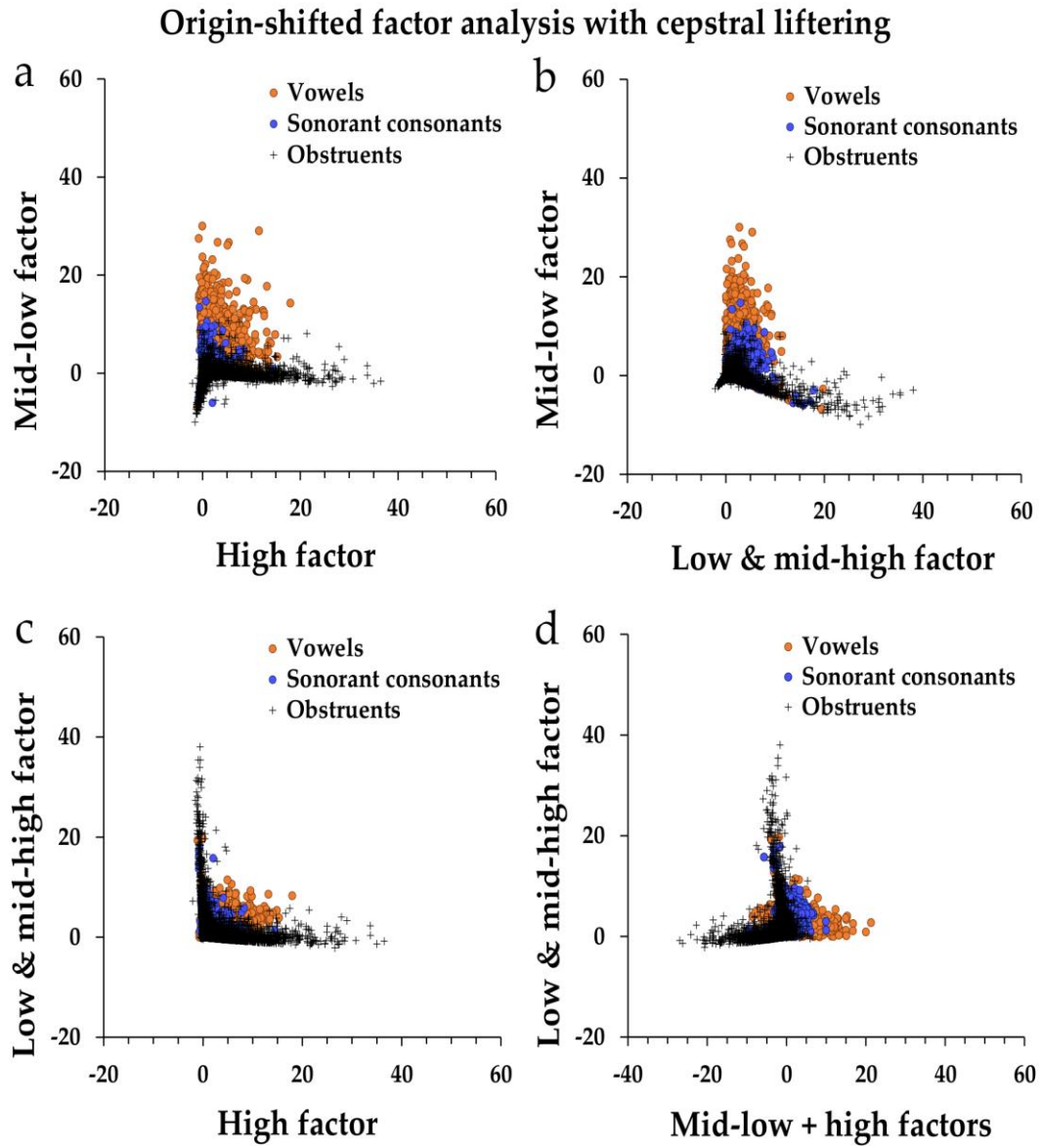


Figure 2.7. The distribution of the factor scores by origin-shifted factor analysis with cepstral liftering, following Kishida et al. (2016). (a–c) represent all possible combinations of two factors out of the three factors, and (d) shows a two-dimensional mapping of the three factors.

The distribution of the factor scores showed a close resemblance between the four factor analyses in Figure 2.4 to Figure 2.7. Although the cepstral liftering had some influence on the factor loadings of the second peak of the double-peaked “Low & mid-high factor”, the factor scores showed similar distributions in the factor space. The origin shift did not greatly impact either the factor loadings or the factor scores. Whether the origin shift was applied or not, the factor loadings (Figure 2.1) and the factor scores (Figure 2.4 to Figure 2.7) showed very similar tendencies.

2.3.4 General distribution of English phonemes in the factor space

As can be seen, all English phonemes were distributed into three distinctive areas in the three-dimensional space. Results showed that on the “Mid-low factor”, the highest factor scores were obtained by vowels, followed by sonorant consonants, while obstruents obtained the lowest factor scores. Most of the obstruents occupied a position very near to or below zero on the “Mid-low factor”. By contrast, they occupied a position above zero on the “Low & mid-high factor”, while on the “High factor” they even reached the highest position of factor scores.

In the three-factor factor analysis, Nakajima et al. (2017) showed that the distribution of the three English-phoneme categories (i.e., vowels, sonorant consonants, and obstruents) constituted an L-shape in the two-dimensional factor space of the “High factor” and the “Mid-low factor”. Following the L-shape, the distribution of the phonemes linearly extended from the origin in the positive direction along the “Mid-low factor” axis, and the distribution of the obstruents linearly extended from the origin in the positive direction along the “High factor” axis. Although the present results were not as clearly distributed and narrowly shaped as in Nakajima et al. (2017), the same

kind of L-shaped distribution was found in Figure 2.4 to Figure 2.7(a) in the present results. Interestingly, the distribution of the obstruents also showed an L-shape in the two-dimensional factor space of panel (c) in Figure 2.4 to Figure 2.7. The factor scores of obstruents were mutually exclusive between the “Low & mid-high factor” and the “High factor”: If the factor score of the “Low & mid-high factor” was higher, then the factor score of the “High factor” was lower. The L-shaped distribution of the factor scores of the obstruents across the “Low & mid-high factor” and the “High factor” as analyzed here also appeared in Nakajima et al. (2017), but hardly in a noticeable way. The present analyses have shown the L-shaped distribution far more prominently.

2.3.5 Distribution of English obstruents in the factor space

In the present results of Study 1, cepstral liftering did not affect the first peak of the “Low & mid-high factor”, which had a frequency range around 300 Hz, and occupied a low-frequency band, approximately from 50 to 600 Hz. Given the prominence of the L-shape for obstruents, more detailed analyses were performed by using origin-shifted factor analysis with cepstral liftering in the present study, while in the following studies obstruents will be analyzed without cepstral liftering.

The distributions of voiced obstruents (Figure 2.8a) and voiceless obstruents (Figure 2.8b) both showed a distinctive L-shape curving at the silent point. Here eight obstruents were categorized as voiced obstruents, with the number of analyzed data points in parentheses: /b/ (609), /d/ (838), /g/ (274), /v/ (467), /ð/ (968), /dʒ/ (53), /ʒ/ (54) and /z/ (827). Nine obstruents were categorized as voiceless obstruents: /p/ (563), /t/ (1682), /k/ (786), /ʃ/ (226), /h/ (340), /tʃ/ (195), /s/ (1389), /f/ (552), and /θ/ (205). For the voiced obstruents (Figure 2.8a), three obstruents occupied a relatively high

position on the “Low & mid-high factor”, and factor scores near zero for the “High factor”: /b/, /v/, and /ð/, for which the “Low & mid-high factor” scores of /v/ and /ð/ were around half of those of /b/. By contrast, four other voiced obstruents occupied a relatively high position on the “High factor”, and factor scores near zero for the “Low & mid-high factor”: /g/, /ʒ/, /dʒ/ and /z/, for which the “High factor” scores of /g/ and /z/ were around half of those of /ʒ/ and /dʒ/. Only /d/ indicated relatively high scores for both the “Low & mid-high factor” and the “High factor”, yet again in an L-shape.

The distributions of voiceless obstruents (Figure 2.8b), /p/, /k/, /h/, and /f/, displayed the L-shape, but the factor scores were relatively high mainly for the “Low & mid-high factor”. The distributions of /t/ and /θ/ also followed the L-shape, but the factor scores were relatively high mainly for the “High factor”. The distributions of three obstruents /ʃ/, /tʃ/, and /s/ occupied a relatively high position only for the “High factor”. The factor scores of /s/ were about half of those of /ʃ/ and /tʃ/.

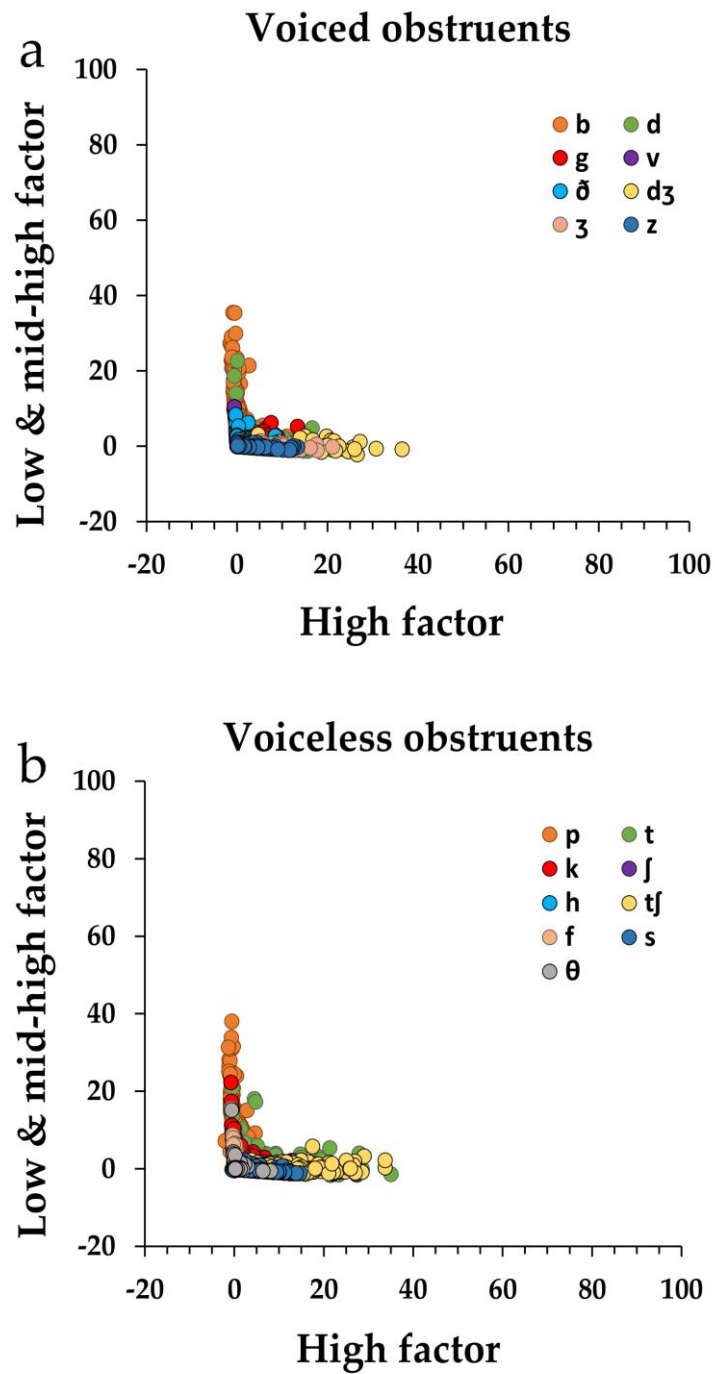


Figure 2.8. The L-shaped distributions of voiced obstruents (a) and voiceless obstruents (b) in the two-dimensional space of the “Low & mid-high factor” and the “High factor”. Factor scores of the three extracted spectral factors were obtained by origin-shifted factor analysis with cepstral liftering (cf. Figure 2.7) and the English speech samples were from three native speakers.

Figure 2.9 shows the distributions of the same set of obstruents divided into fricatives/affricates and plosives on the two-dimensional configuration, again clearly showing the distinctive L-shape. Here, eleven obstruents were categorized as fricatives/affricates, with the number of analyzed data points in parentheses: /θ/ (205), /ð/ (968), /f/ (552), /v/ (467), /s/ (1389), /z/ (827), /ʃ/ (226), /ʒ/ (54), /h/ (340), /tʃ/ (195) and /dʒ/ (53). Six obstruents were categorized as plosives: /p/ (563), /t/ (1682), /k/ (786), /b/ (609), /d/ (838) and /g/ (274). For the fricatives/affricates (Figure 2.9a), the distributions of six obstruents occupied a relatively high position on the “High factor”: /s/, /z/, /ʃ/, /ʒ/, /tʃ/ and /dʒ/. The distributions of five obstruents were both located on the “Low & mid-high factor” and the “High factor” in the L-shape: /θ/, /ð/, /f/, /v/, /h/, but they were higher mainly on the “Low & mid-high factor”. As for the distributions of plosives (Figure 2.9b), all of them displayed the L-shape, but /p/, /b/ and /k/ were higher mainly on the “Low & mid-high factor”, and /t/, /d/ and /g/ were higher mainly on the “High factor”.

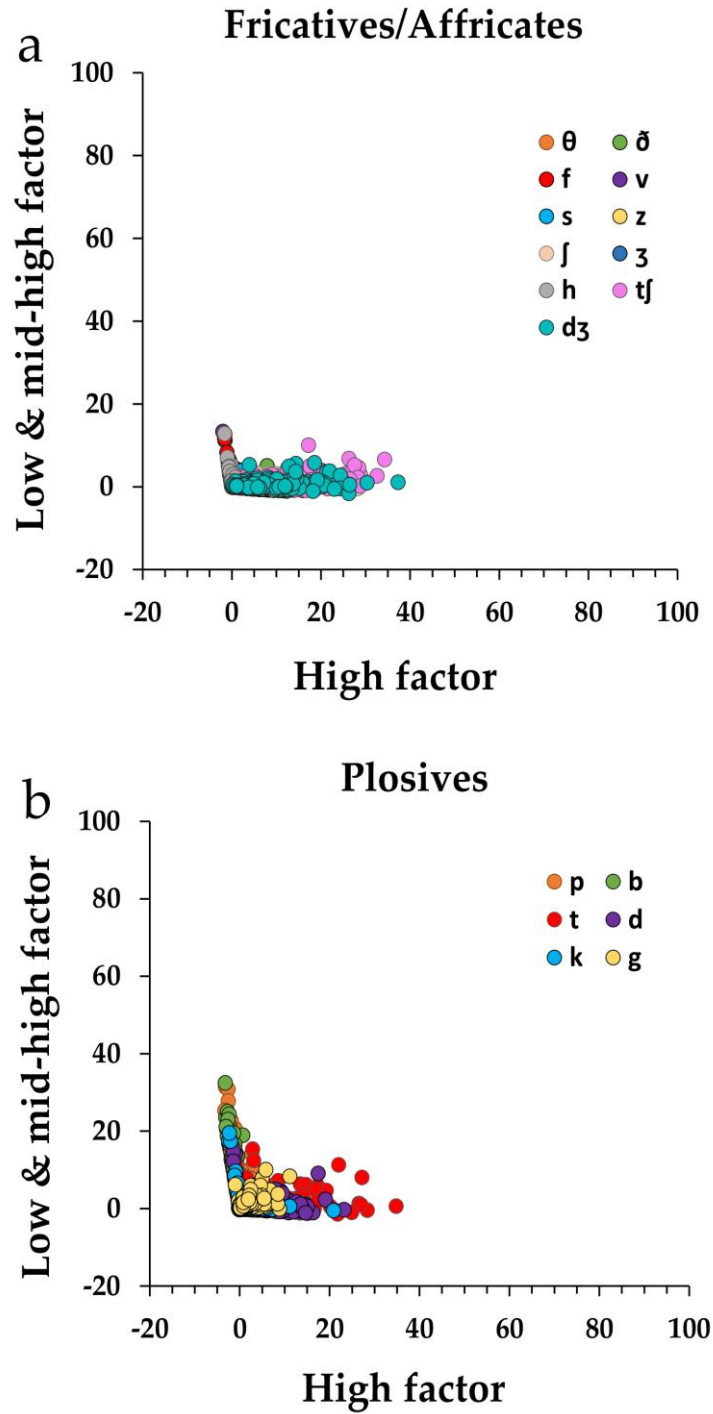


Figure 2.9. The L-shaped distributions of obstruents divided into fricatives/affricates (a) and plosives (b) in the two-dimensional space of the “Low & mid-high factor” and the “High factor”. Factor scores of the three extracted spectral factors were obtained by origin-shifted factor analysis with cepstral liftering (cf. Figure 2.7) and the English speech samples were from three native speakers.

2.4 Discussion and conclusions

Comparative analyses were performed between four different factor analysis methods applied to English speech. The difference between these four methods was manifested in two processing aspects: (1) whether cepstral liftering was applied or not, and (2) whether the origin shift was used or not. The present results showed that neither cepstral liftering nor the origin shift had a great effect on the factor scores (Figure 2.2 to Figure 2.7), whose distributions were similar. Furthermore, the factor loadings of one factor, the “Low & mid-high factor” were more pronounced when the analysis was without cepstral liftering. In particular, the second peak (around 2300 Hz) of the double-peaked factor was prominent, as can be seen in the results of the three speakers in Figures 2.1 (c) and (d) or the results of the male speaker in Figure 2.3 (b). Cepstral liftering is a process to smooth spectra of speech signals. The present results in Study 1 show that cepstral liftering did not have a great effect on the first peak of the “Low & mid-high factor”, which was in the low-frequency bands, as shown in Figure 2.1 and Figure 2.3. The 5-ms cepstral liftering had little impact on the factor loadings for the male and the two female speakers. Ideally, though, the parameter should be tailored to each specific talker, because the results might change if a more suitable parameter for the female talkers of higher voices corresponding to lower quefrequencies were employed.

Second, a potential effect of cepstral liftering may appear for a larger number of factors because the Supplementary Information of Ueda and Nakajima (2017) showed that the fourth and fifth factors represented resolvable harmonics in the low-frequency region when more than five factors were extracted. Therefore, in the present

research, it can be concluded that the cepstral liftering will not affect the final impression of the analyses, however, it was likely to polish the spectral content of the speech signal. The current results show that there is little difference between the analyses with and without cepstral liftering, and it is recommendable not to use cepstral liftering (not to change the original signal information). Given the second and third purpose of this dissertation, that is, to determine the acoustic features of English phonemes in general, including obstruents that tend to have low energy, to analyze speech in a natural manner might be preferred, i.e., without cepstral liftering.

Although the origin shift did not have a large impact on the factor scores, origin-shifted factor analysis is highly recommendable for future research on speech resynthesis and subsequent listening experiments, because it can be difficult to find any features around silent parts in speech by utilizing the normal factor analysis. If the quality of the resynthesized speech is better, it is more likely to be related closely to the real auditory signal, and the resynthesized speech is considered more useful for listening experiments. It was thus concluded that the origin shifted factor analysis is more recommendable for speech analyses in the present context, and it is used in the analyses described in Studies 2 and 3.

The second purpose of Study 1 was to determine the acoustic correlates of English phonemes utilizing different factor analyses. New insight was obtained with regard to obstruents. In an English syllable, an isolated obstruent always has a position as a syllable-onset or a syllable-end. If two or more obstruents are next to one another, one of them begins or ends the syllable. Analyzing obstruents acoustically as in the present Study 1 helps to understand this phonological phenomenon. The analyses

showed a factor with a frequency range around 1100 Hz (the “Mid-low factor”). According to Nakajima et al. (2017), this factor seems closely related to vowels and sonorant consonants, and thus to syllable nuclei, most of which are vowels but a few are sonorant consonants. The current analyses provide evidence that obstruents are not only associated with the factor related to a frequency range above 3000 Hz (the “High factor”), as suggested by Nakajima et al. (2017), but also with the double-peaked factor with frequencies around 300 Hz and 2300 Hz (the “Low & mid-high factor”).

The present findings thus suggest that these two extracted factors (the “Low & mid-high factor” and the “High factor”) of the acoustic natures of obstruents reflect strong cues as to the beginning or the end of a syllable. It is important to note that the distributions of subsets of obstruents (voiced and voiceless, Figure 2.8; fricatives/affricates and plosives, Figure 2.9) indeed occupied high positions on the “Low & mid-high factor” and the “High factor”, but not on the “Mid-low factor”. The obstruents were often far from the origin, the silent point in origin-shifted factor analysis, but they never went into the positive direction of the “Mid-low factor”. This confirms that obstruents do not constitute the syllable nucleus, but rather delimit the syllable, corroborating the typical sonority hierarchy in phonology on which obstruents have the lowest position.

Chapter 3. Study 2: A three-factor analysis of English speech

3.1 Purpose

The results of Study 1 showed that origin-shifted factor analysis is more recommendable for speech analyses, and therefore it is used in the analyses in Study 2. Moreover, the results also showed that indeed multivariate analysis of acoustic characteristics of English speech sounds can be related to phonological features (also see Kishida et al., 2016; Ueda and Nakajima, 2017; Nakajima et al., 2017). In Study 1, three spectral factors appeared in four frequency bands. The distributions of the factor scores of English vowels, sonorant consonants, and obstruents separately went from a high to a low position on a spectral factor with high factor loadings on a frequency range around 1100 Hz, which was called the “Mid-low factor”. This tendency was similar to the finding of Nakajima et al. (2017). However, in the analysis of Study 1 and in Nakajima et al. (2017), an English speech database, “The ATR British English Speech Database” (Campbell, 1993), was used in which the labeling of a large number of phonemes was problematic and the quality of the speech was unsuitable. For this reason, 7523 out of 31663 phonemic labels were omitted.

The purpose of Study 2 is therefore to further investigate the acoustic features of English phonemes by using a new English speech database, with sentences spoken by three native speakers of English, specifically recorded for the present purpose. Furthermore, the cepstral liftering as used in the analyses in Study 1 appeared to have had an influence on the “Low & mid-high factor”, though not on the first “Low” peak.

Following this, origin-shifted factor analysis without cepstral liftering was used to analyze the speech sounds of the new database in the present study.

3.2 Method

3.2.1 Speech samples

To create a high-quality English speech database, spoken sentences of native speakers of English were recorded and all recorded sentences were segmented with English phoneme labels. The recording of the English speech was performed at the National University of Ireland, Galway (NUIG), and the labeling and analysis were performed at Kyushu University (Fukuoka, Japan). The name of the database is therefore the “Fukuoka-Galway English Speech Database”. The database contained 100 sentences uttered by one male and two female native speakers of Irish English at the National University of Ireland, Galway (NUIG). The sentences were the first 100 sentences of “The ATR British English Speech Database” (Campbell, 1993). The speech samples were recorded in a soundproof booth with a background noise not exceeding ~25 dB (A), utilizing a handy linear PCM audio recorder (TASCAM DR-07), with a sampling frequency of 44100 Hz, and a linear quantization of 16-bit. The speakers were 15 cm away from the stereo microphone with a pop guard (TASCAM WS-11) and they read the sentences naturally.

Following the recording, the speech sound corresponding to each uttered English phoneme in each sentence of the 300 utterances (100 sentences from 3 speakers) was labeled by using the software “Praat” (Boersma & Weenink, 2013). A total number of 11816 phonemes were labelled. The speech signals of all the spoken

sentences were segmented into individual phonemes and were labeled utilizing the Machine Readable Phonetic Alphabet (MRPA) (Lennig and Brassard, 1984). The pronunciation of phonemes referred to the International Phonetic Alphabet, with the “Cambridge English Pronouncing Dictionary” as a pronunciation reference (Jones, 2011). All the labeling was segmented by two persons: a Chinese national, who was a student at Kyushu University (female, 33 years old), and the author of this dissertation, also a native-Chinese speaker. Further, to confirm the correctness and consistency of the labeling, two native speakers of English, who were students at Kyushu University (one male and one female from the United States, 30 years old), each checked the labeling of 50 randomly-selected sentences from the database. If the pronunciation of the phoneme they heard and the labeling were different, it was played and checked again by them and the author to re-identify the controversial phoneme until an agreement was reached on the labeling of the phoneme. Moreover, these disputed phonemes were also randomly checked in other sentences. If there were any labeling issues, they would be checked with the native speakers of English again until an agreement was reached. About 5% of the phonemes had been modified because some phonemes seemed to contain a bit of the sound of the phoneme before or after it, which made it not sound like a single phoneme.

An example of segmented phonemes is shown in Figure 3.1. The segments were from the sentence “The price range is smaller than any of us expected”. A “phoneme” row and a “word” row were separately created corresponding to the speech sounds. In the “word” row (the fourth row in Figure 3.1), the boundaries between each word were determined while listening to the audio signal. The hashtag mark (#) was

added to represent the space between words, and a double hashtag (##) was used to represent the space at the beginning and the end of the sentences. In the “phoneme” row (the third row in Figure 3.1), each phoneme was labeled into an individual sound that contained the information of that phoneme. If the phoneme could not be distinguished or seemed to contain more than one phoneme, a “+” (as the “+” between /ai/ and /s/ in Figure 3.1) or the segment of these two phonemes (as the labeling “dh@” between phonemes /dh/ and /@/ in Figure 3.1) were used to represent this part. In summary, 300 English utterances from the “Fukuoka-Galway English Speech Database” were used in the present analysis. A total of 11816 phonemes, with 4526 vowels, 2467 sonorant consonants, and 4823 obstruents, were taken up as analysis samples.

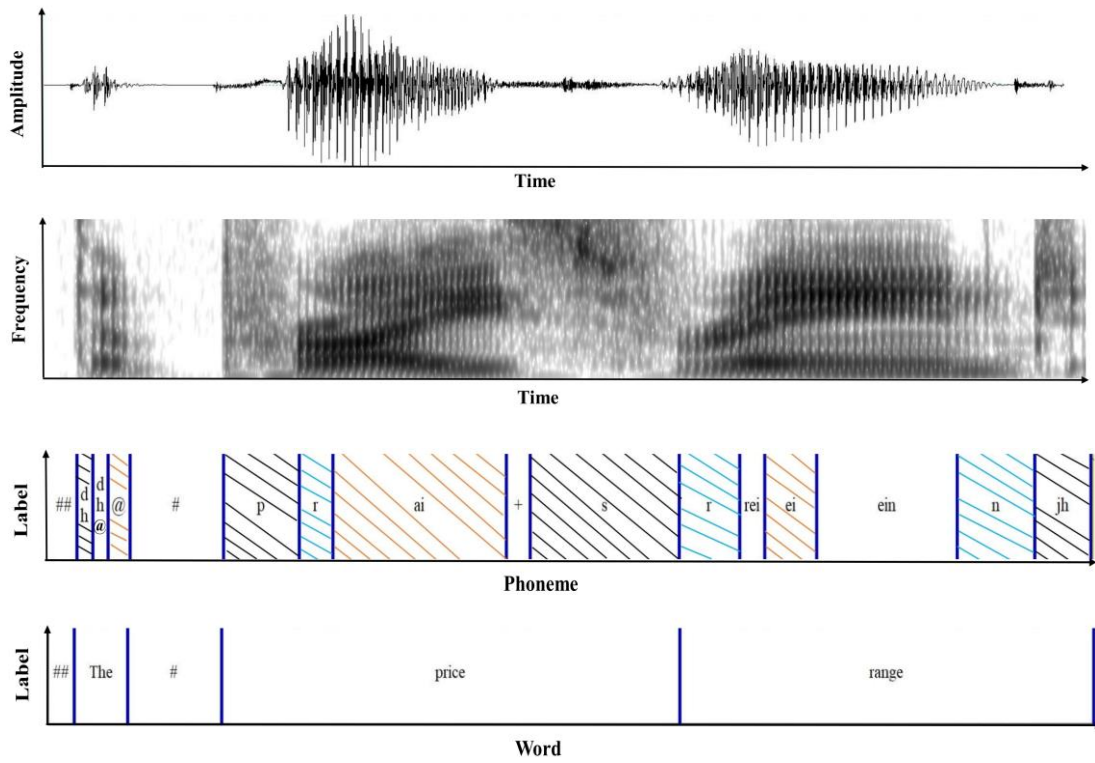


Figure 3.1. An example of segmented phoneme labeling in a sentence from the “Fukuoka-Galway English Speech Database”. The sentence was “The price range is smaller than any of us expected”. The first row shows the speech signal’s waveform; the second row shows the spectrogram of the spoken speech; the third row shows the phoneme labeling, and the fourth row shows the word labeling, which made up a spoken sentence. Vowels “@” (/ə/), “ai” (/aɪ/), “ei” (/eɪ/). Sonorant consonants “r” (/r/) and “n” (/n/) are highlighted in blue. Obstruents “dh” (/ð/), “p” (/p/), “s” (/s/), and “jh” (/dʒ/) are highlighted in black.

3.2.2 Procedure

The speech samples were analyzed with origin-shifted factor analysis without cepstral liftering, which was found to be the most suitable method from the analysis in Study 1. The two-dimensional axes derived from the analysis were rotated by varimax rotation (Kaiser, 1958) in the same way as described in Study 1, section 2.2.2. The present analysis was different from the analysis in Study 1, in that instead of dividing the power spectrum into 19 critical bands, it here was divided into 20 critical bands, and averaged power was calculated for each band (Table 3.1). The frequency bands were adopted from Zwicker and Terhardt (1980) and Kishida et al. (2016), and the center frequencies ranged from 75–5800 Hz. The frequency range below 50 Hz was not utilized because it should have been less relevant to the speech signal. The first frequency band was from 50 to 100 Hz, and the center frequency was 75 Hz. Thus 20 critical-band filters were constructed to cover the frequency range of 50 - 6400 Hz (Table 3.1).

Table 3.1. Critical-band filters used in Study 2. Different from the analysis in Study 1, speech signals were divided into 20 critical-band filters. The band number, center frequency (Hz), and the passband of each critical band filter are shown with reference to Zwicker and Terhardt (1980) and Kishida, et al. (2016).

Band No.	Center frequency [Hz]	Passband [Hz]
1	75	50 - 100
2	150	100 - 200
3	250	200 - 300
4	350	300 - 400
5	450	400 - 510
6	570	510 - 630
7	700	630 - 770
8	840	770 - 920
9	1000	920 - 1080
10	1170	1080 - 1270
11	1370	1270 - 1480
12	1600	1480 - 1720
13	1850	1720 - 2000
14	2150	2000 - 2320
15	2500	2320 - 2700
16	2900	2700 - 3150
17	3400	3150 - 3700
18	4000	3700 - 4400
19	4800	4400 - 5300
20	5800	5300 – 6400

To compare the present analysis with the analysis of Study 1, in the present analysis three spectral factors were extracted and their factor loadings were obtained. Following Study 1, all analysis samples of English phonemes were divided into three categories: vowels, sonorant consonants, and obstruents. The factor scores of the central midpoints of time for all the labeled phoneme samples were calculated. To observe the acoustic correlates in more detail, the distributions of all English phonemes were analyzed.

3.3 Results

3.3.1 Factor loadings of the three spectral factors

Figure 3.3 shows the factor loadings obtained by origin-shifted factor analysis without cepstral liftering for all English speech samples spoken by three native speakers for the “Fukuoka-Galway English Speech Database”. The cumulative contributions of the three spectral factors were around 47%. Four main frequency bands were obtained. Their frequency ranges are indicated by the center frequencies on the horizontal axis in Figure 3.2. One division of the horizontal axis corresponds to 0.5 critical bandwidth. The first band was a low-frequency band, from about 50 to 600 Hz. The second band was a mid-low frequency band, from about 600 to 1700 Hz. The third band was a mid-high-frequency band, from about 1700 to 3000 Hz. The fourth constituted a high-frequency band, which was above 3000 Hz. These four frequency bands were related to three spectral factors. One factor, the “Low & mid-high factor” (red line) was double-peaked, because it showed high factor loadings on two center frequency ranges around 300 Hz and around 2300 Hz. Although the second peak of

this factor did not appear very prominently, we considered it to be similar to the shape of the “Low & mid-high factor” as found in Study 1. The second factor, the “Mid-low factor”, was related to the frequency range around 1200 Hz (black line). The third factor, the “High factor”, was related to the frequency range around 4100 Hz (blue line).

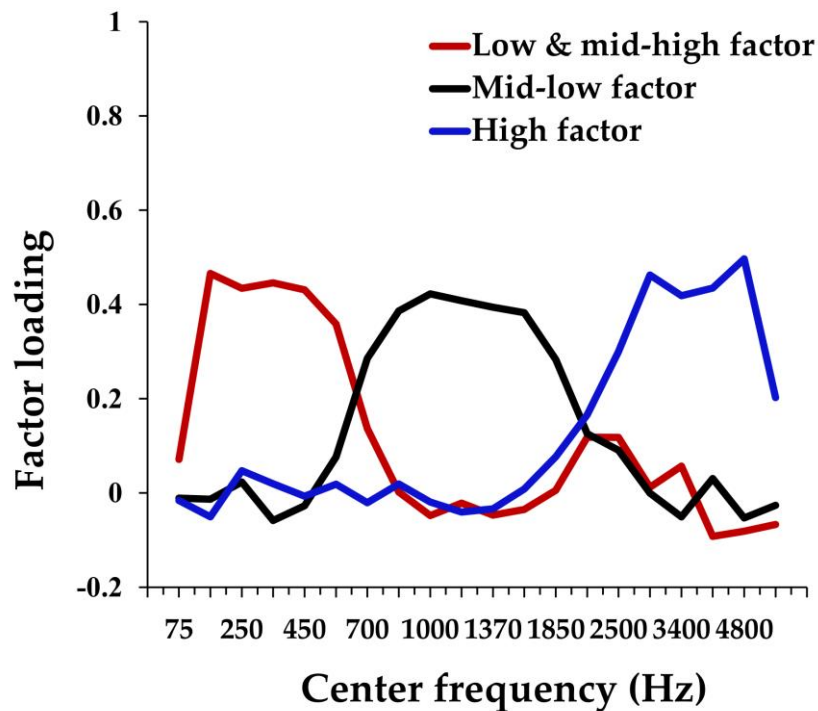


Figure 3.2. Factor loadings of the three extracted spectral factors of 300 English speech samples, from three native speakers in the “Fukuoka-Galway English Speech Database”, with origin-shifted factor analysis without cepstral liftering. The first factor, the “Low & mid-high factor” (red line) was double-peaked, with high factor loadings on two center frequency ranges around 300 Hz and around 2300 Hz. The second factor, the “Mid-low factor”, was related to the frequency range around 1200 Hz (black line). The third factor, the “High factor”, was related to the frequency range around 4100 Hz (blue line).

3.3.2 Factor scores of the three spectral factors

The factor scores were analyzed, following the analysis of factor loadings. Except for obstruents, the results of the other two categories of phonemes, vowels and sonorant consonants, were similar to those obtained in Study 1 with the “The ATR British English Speech Database” (Campbell, 1993). The results are shown in Figure 3.3 below. Panels (a), (b), and (c) show the distributions of combinations of the three spectral factor scores. Combinations were made between the “Low & mid-high factor” (the frequencies of high loadings were around 300 Hz and 2300 Hz), the “Mid-low factor” (the frequency of high loadings was around 1200 Hz), and the “High factor” (the frequency of high loadings was around 4100 Hz). Panel (d) in Figure 3.3 shows the distribution in the three-dimensional configuration as viewed in a direction from above-right to below-left in panel (a). In panel (d), the horizontal axis represents the combination of the “Mid-low factor” and the “High factor”, by calculating $(x-y)/\sqrt{2}$, where “x” indicates the factor score of the “Mid-low factor”, and “y” the factor score of the “High factor”, following Nakajima et al. (2017) and the results of Study 1, section 2.3.2.

3.3.3 General distribution of English phonemes in the factor space

The factor scores of all English phonemes were obtained and distributed into three categorical areas in the factor space, i.e., vowels (orange circles), sonorant consonants (blue circles), and obstruents (black crosses), as shown in Figure 3.3 and Appendix C. On the “Mid-low factor” with a center frequency range of around 1200 Hz, the highest factor scores were obtained by vowels, followed by sonorant

consonants, and obstruents. Actually, the same tendency also appeared on the “Low & mid-high factor”. By comparing the vertical axis of Figure 3.3 (a) and (c), it can be clearly seen that the classification of the three categories of English phonemes on the “Low & mid-high factor” was not as widely distributed as on the “Mid-low factor”. They were concentrated in the middle position on the “Low & mid-high factor”. Moreover, most of the obstruents occupied a position very near to or below zero on the “Mid-low factor”. Instead, obstruents were associated with the two other spectral factors: the “Low & mid-high factor” and the “High factor”, although the distribution was not as clear as the previous results for the “Low & mid-high factor” as described in Study 1, section 2.3.3.

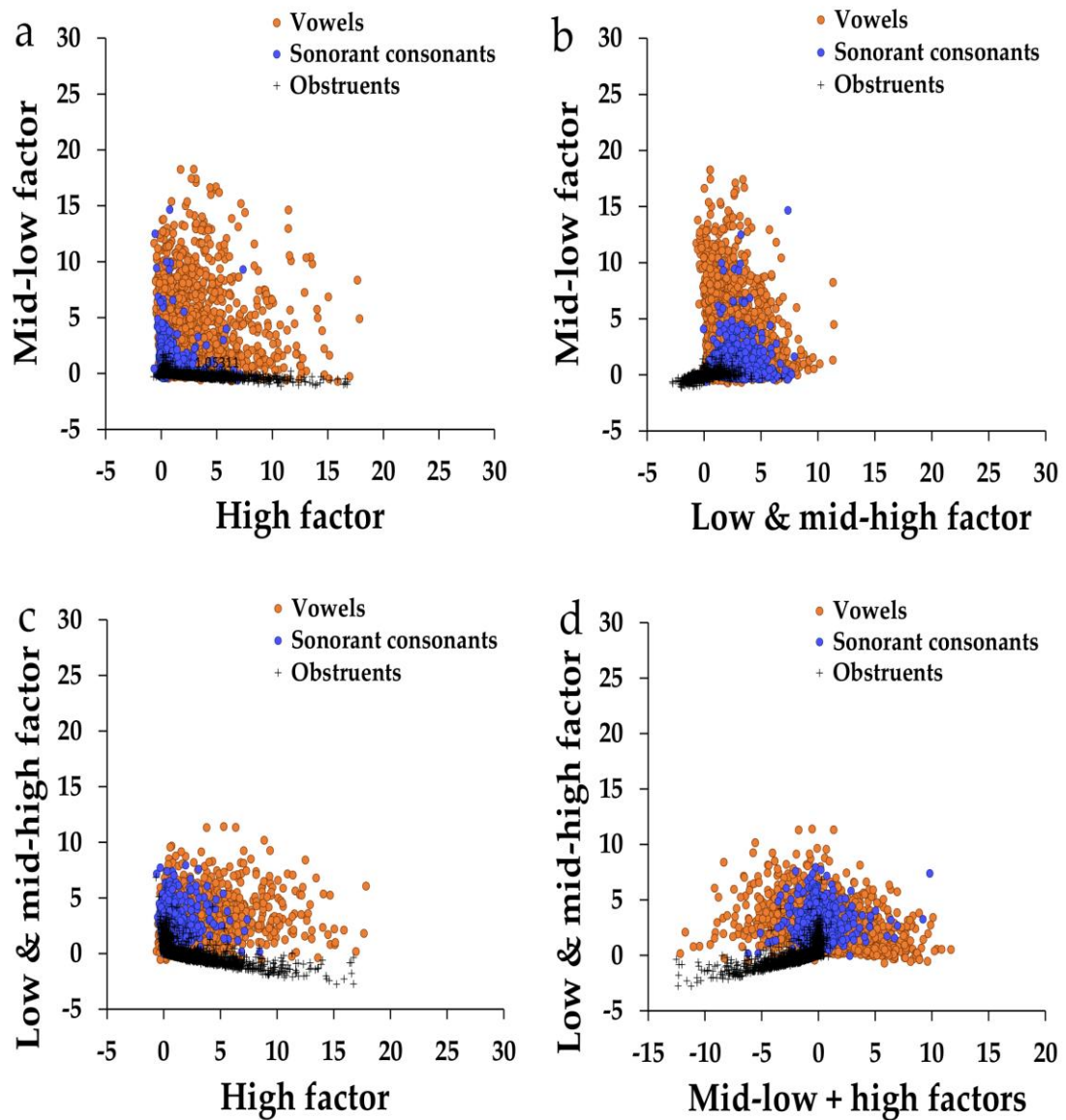


Figure 3.3. The distribution of the factor scores by origin-shifted factor analysis without cepstral ltering. (a–c) represent all possible combinations of two out of the three factors, and (d) shows a two-dimensional mapping of the three factors.

3.3.4 Distribution of English obstruents in the factor space

To get more insight into the distributions of individual obstruents in the factor space, more detailed analyses of obstruents on the “Low & mid-high factor” and the “High factor” were performed. Figure 3.4 shows the distributions of the factor scores of individual obstruents divided into eleven fricatives/affricates in Figure 3.4 (a), and six plosives in Figure 3.4 (b). The distributions show that five fricatives/affricates (/θ/, /ð/, /f/, /v/, /h/) were close to the origin and both located on the “Low & mid-high factor” and the “High factor”, while six fricatives/affricates (/s/, /z/, /ʃ/, /ʒ/, /tʃ/, /dʒ/) were distributed on only the “High factor”, occupying a relatively wide distribution and high position. Furthermore, all six plosives (/p/, /t/, /k/, /b/, /d/, /g/) occupied a position above zero on the “Low & mid-high factor” and the “High factor”. Two plosives (/t/, /g/) occupied a wider distribution and reached the highest positions near fricatives/affricates (/ʃ/, /tʃ/, /dʒ/) on the “High factor”.

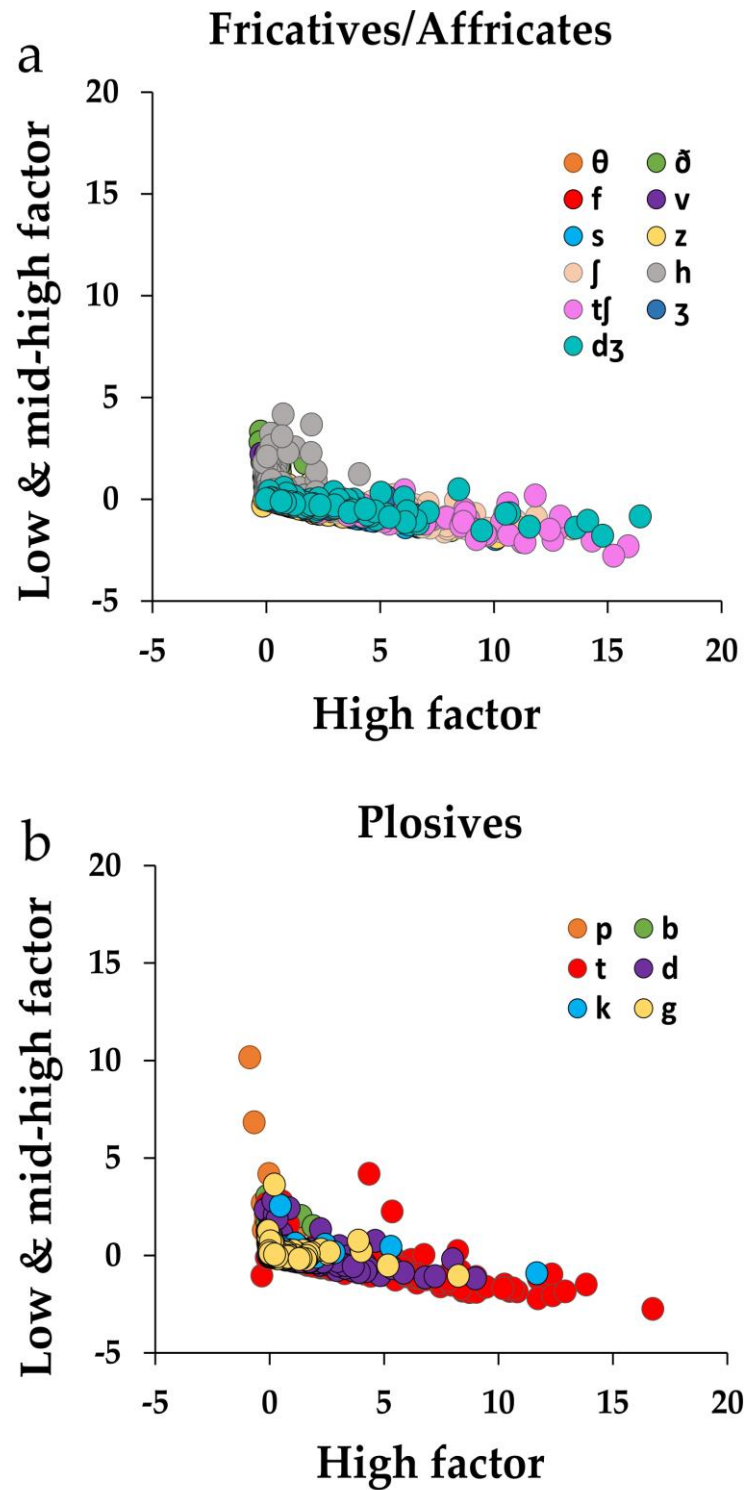


Figure 3.4. The distributions of the factor scores of obstruents divided into fricatives/affricates (a) and plosives (b) in the two-dimensional space of the “Low & mid-high factor” and the “High factor”.

The distributions of the average factor scores of the fricatives/affricates in Figure 3.5 (a) and of the plosives in Figure 3.5 (b), in the two-dimensional space of the “Low & mid-high factor” and the “High factor”, can be further observed. In Figure 3.5 (a), the five fricatives/affricates (/θ/, /ð/, /f/, /v/, /h/) were close to the origin and on both the “High factor” and the “Low & mid-high factor”. The distributions of the six fricatives/affricates (/s/, /z/, /ʃ/, /ʒ/, /tʃ/, /dʒ/) were almost parallel to and only located on the “High factor”. In Figure 3.5 (b), all six plosives (/p/, /t/, /k/, /b/, /d/, /g/) were close to the origin on both the “High factor” and the “Low & mid-high factor”, but they were mainly distributed on the “High factor”.

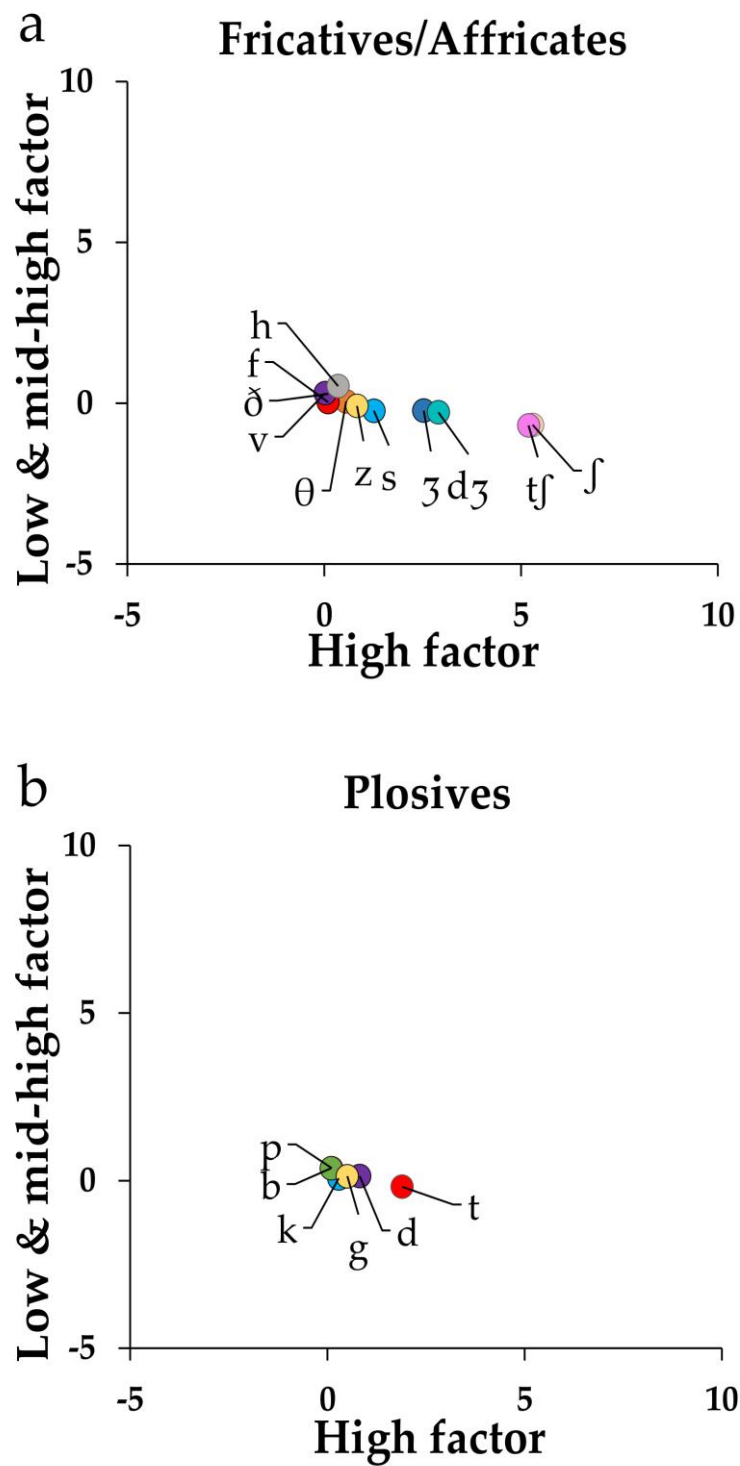


Figure 3.5. The distributions of the average of the factor scores of obstruents divided into fricatives/affricates (a) and plosives (b) in the two-dimensional space of the “Low & mid-high factor” and the “High factor”.

3.4 Discussion and conclusions

The purpose of Study 2 was to further investigate the acoustic features of English phonemes by using a newly-recorded English speech database. Origin-shifted factor analysis without cepstral liftering was applied to English speech samples taken from the new database that contained 100 sentences, each spoken by three native speakers of English. The database is called the “Fukuoka-Galway English Speech Database”. In total, 11816 labeled English phonemes were used as analysis samples. The three-factor analysis results again showed that the distributions of the obtained factor scores in the factor space well reflected the acoustic features of English phonemes and could be related to sonority features in phonology.

The results of factor analysis showed three spectral factors, i.e., the “Low & mid-high factor”, the “Mid-low factor”, the “High factor”, which appeared in four frequency bands (50–600, 600–1700, 1700–3000, and above 3000 Hz), consistent with the results by using the “ATR British English Database” in Study 1. The analysis proved again that the “Mid-low factor”, with high factor loadings on a frequency range around 1200 Hz, was only related to vowels and sonorant consonants. Vowels, sonorant consonants, and obstruents were separated very clearly on this factor. Vowels had the highest position on the “Mid-low factor”. Based on the results of Studies 1 and 2, it was confirmed that the “Mid-low” factor is related to sonority, as suggested by Nakajima et al. (2017). Because vowels are always playing the roles of syllable nuclei in English, they are high in the sonority hierarchy (Spencer, 1996). Sonorant consonants occupied the middle position on the “Mid-low factor”. They are also in the middle position in the sonority hierarchy, indicating that sonorant consonants play

roles close to those of vowels, in that some of them also can be syllable nuclei. For example, the nucleus of the first syllable of the word “little” is the vowel /ɪ/, but the nucleus of the second syllable is the sonorant consonant /l/ (Gordon, 2016). Obstruents had the lowest position on this “Mid-low factor”. They also have the lowest position in the sonority hierarchy, indicating that obstruents can hardly be treated as syllable nuclei.

Moreover, the acoustic natures of English obstruents in the other two spectral factors, i.e., the “Low & mid-high factor” with a peak frequency range around 300 Hz and around 2300 Hz, and the “High factor” with a peak frequency range around 4100 Hz, were further determined. As shown in Figures 3.4 and 3.5, the distributions on the “Low & mid-high factor” were not as prominent as the distribution indicated in Figure 2.4 (c) of Study 1. The distribution of the obstruents showed an L-shape in the two-dimensional factor space. When the factor scores of the “Low & mid-high factor” were higher, then the factor scores of the “High factor” were lower. The distribution of the factor scores of the obstruents in the present Study 2 was similar to those in Study 1 and in Nakajima et al. (2017).

In an English syllable, obstruents typically occupy a low position on the sonority hierarchy. This was reflected in our analysis: obstruents never went in the positive direction of the “Mid-low factor” – which can be identified as the “sonority factor”. By contrast, obstruents were only distributed on the “High factor”, or on both the “Low & mid-high factor” and the “High factor”. Fricatives/affricates occupied a wider distribution than plosives on the “High factor”. These pieces of evidence corresponded to the fact that obstruents do not constitute syllable nuclei.

With the present multivariate analysis, it is thus possible to connect acoustics and phonology. From the three extracted spectral factors, the “Mid-low factor” seemed to represent the sonority features of English speech sounds, confirming the conclusion of Nakajima et al. (2017). In the following Study 3, a four-factor analysis will be performed. When the number of the extracted spectral factors increases, one of the factors will be separated, as was shown in previous research (Ueda et al., 2008; Kishida et al., 2016; Ueda & Nakajima, 2017; Ueda et al., 2023). In Kishida et al. (2016), when the number of the extracted factors increased from three to four, the double-peaked “Low & mid-high factor” was separated into two factors. In order to confirm this and to further identify the acoustic features of English phonemes, and in particular those of obstruents, the “Fukuoka Galway English database” was analyzed by using a four-factor analysis in Study 3.

Chapter 4. Study 3: A four-factor analysis of English speech

4.1 Purpose

In Study 2, a high-quality English speech database containing 300 English-spoken sentences was analyzed by using origin-shifted factor analysis without cepstral liftering. The results showed that three spectral factors (the “Low & mid-high factor”, the “Mid-low factor”, and the “High factor”) appeared in four frequency bands (50–600, 600–1700, 1700–3000, and above 3000 Hz), consistent with the results of Study 1. On the “Mid-low factor”, the distributions of factor scores for the three categories of English phonemes from high to low were located in the order of vowels, sonorant consonants, and obstruents. However, obstruents got the highest position of factor scores on the “High factor”.

The purpose of Study 3 was to analyze the same English speech sentences as used in Study 2 by means of four-factor analysis to further observe the acoustic features of the extracted spectral factors. As Kishida et al. (2016), Ueda and Nakajima, (2017), Ueda et al. (2018), and Ueda et al. (2023) claimed, when the number of the extracted spectral factors increases, one of the factors will be separated. The first purpose of Study 3 was to investigate the separated factor. The second purpose was to determine how the acoustic features of English phonemes (vowels, sonorant consonants, and obstruents) appeared in the distribution of factor scores by using origin-shifted factor analysis with or without cepstral liftering. The third purpose was to specifically focus on how English obstruents appear in the obtained factor score spaces, since Study 2 showed that by using origin-shifted factor analysis with the

“Fukuoka-Galway English Speech Database”, even obstruents could be analyzed in detail.

4.2 Method

4.2.1 Speech samples

Following the method as described in Study 2, 100 spoken sentences uttered by three native speakers of English from the “Fukuoka-Galway English Speech Database” were used for the present analysis. All phoneme labeling was associated with the waveform and spectrogram of speech signals, as described in Study 2, section 3.2.1. Thus, a total of 300 sentences containing 11816 individual English phonemes (4526 vowels, 2467 sonorant consonants, and 4823 obstruents) were used. The speech sentences had been recorded with a sampling frequency of 44100 Hz and a linear quantization of 16 bits.

4.2.2 Procedure

The results of Study 1 and Study 2 showed that cepstral liftering did not have a great effect on factor analysis, especially on the factor scores. To confirm this again in the four-factor analysis, here the speech samples were analyzed with origin-shifted factor analysis both with and without cepstral liftering. In the present analysis, the speech samples were analyzed by dividing them into 20 critical-band filters, and all critical-band filters were constructed to cover the frequency range of 50–6400 Hz. The center frequencies ranged from 75–5800 Hz. The 20 power fluctuations were subjected

to origin-shifted factor analysis followed by varimax rotation, in the same way as described in Study 2, section 3.2.2.

Four spectral factors were extracted and their factor loadings were obtained. Following the English phoneme categories as used in Studies 1 and 2, all analysis samples of English phonemes were divided into three categories here as well: vowels, sonorant consonants, and obstruents. The factor scores of the central midpoints of time for all the labeled phoneme samples were calculated. To observe the acoustic correlates in more detail, the distributions of factor scores of all English phonemes and the average of each English phoneme were analyzed.

4.3 Results

4.3.1 Factor loadings of the four spectral factors

Figure 4.1 shows the factor loadings obtained by origin-shifted factor analysis of 100 English speech samples, from three native speakers in the “Fukuoka-Galway English Speech Database”. The result of the factor loadings was similar to that in Study 2, with the three-factor analysis: Here too, four main frequency bands were obtained. These four frequency bands were separated by the crossings of the factor loadings of the four factor curves. The first band was a low-frequency band, from about 50 to 600 Hz. The second band was a mid-low-frequency band, from around 600 to 1700 Hz. The third band was a mid-high-frequency band, from about 1700 to 2700 Hz. The fourth constituted a high-frequency band, which was above 2700 Hz. The frequency ranges are indicated by the center frequencies on the horizontal axis in Figure 4.1. One

division of the horizontal axis corresponds to 0.5 critical bandwidth. The cumulative contributions of the four spectral factors were around 50%.

Four factors appeared in these frequency bands. The double-peaked “Low & mid-high factor” in the three-factor analysis, as described in Study 2, was separated into two factors: one was the “Low factor”, and the other was the “Mid-high factor”. The “Low factor” was related to the frequency range around 300 Hz (red line). The “Mid-high factor” showed a double-peaked trend, related to two frequency ranges around 700 Hz and 2200 Hz (green line). The other two factors were the same as those in the three-factor analysis. One was the “Mid-low factor”, which was related to the frequency range around 1200 Hz (black line). The other was the “High factor”, related to the frequency range around 4100 Hz (blue line). By comparing Figures 4.1(a) and (b), it can be confirmed that the factor loadings of the four extracted factors were hardly influenced by whether cepstral liftering was used or not. The factor loadings on the “Low factor” (around 300 Hz) and the first peak (around 700 Hz) of the factor loadings on the double-peaked “Mid-high factor” were a little bit higher with cepstral liftering than without cepstral liftering.

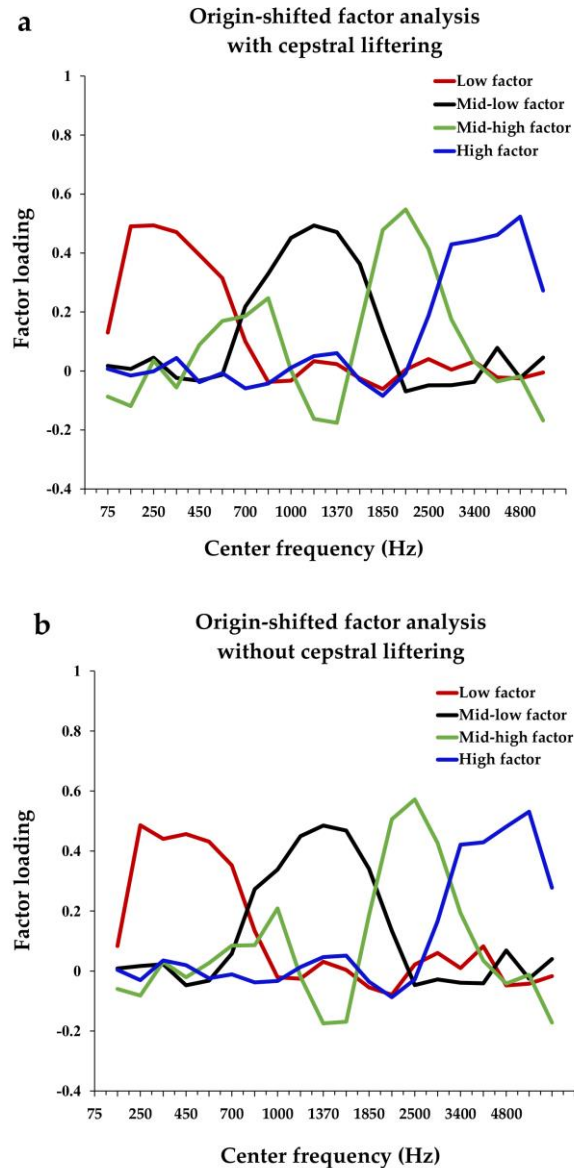


Figure 4.1. Factor loadings of the four extracted spectral factors of 100 English speech samples, from three native speakers in the “Fukuoka-Galway English Speech Database”, with origin-shifted factor analysis (a) with cepstral liftering, and (b) without cepstral liftering. Four factors appeared in four frequency bands. The first factor, the “Low factor” was related to the frequency range around 300 Hz (red line). The second factor, the “Mid-low factor” was related to the frequency range around 1200 Hz (black line). The third factor, the “Mid-high factor”, was double-peaked, related to two frequency ranges around 700 Hz and 2200 Hz (green line). The fourth factor, the “High factor” was related to the frequency range around 4100 Hz (blue line).

4.3.2 Factor scores of the four spectral factors

The factor scores were analyzed, following the analysis of factor loadings. The detailed results were as follows, as shown in Figure 4.2 below. Panels (a), (b), (c), (d), (e), and (f) show the distributions of six combinations of the four spectral factor scores. The combinations were made between the “Low factor” (with the frequency loadings around 300 Hz), the “Mid-low factor” (with the frequency loadings around 1200 Hz), the double-peaked “Mid-high factor” (with the frequency loadings around 700 Hz and 2200 Hz), and the “High factor” (with the frequency loadings around 4100 Hz). Following the same categories as in Study 1 and Study 2, the factor scores of all English phonemes were distributed into three categorical areas in the factor space, i.e., vowels (orange circles), sonorant consonants (blue circles), and obstruents (black crosses).

The four-factor analysis showed that the factor scores of the three categories of English phonemes on the “Mid-low factor” (with the frequency loadings around 1200 Hz) were consistent with the distribution results in Study 1 and Study 2, with the three-factor analysis. That was, vowels were the highest, followed by sonorant consonants, and finally obstruents. However, the distribution of factor scores also showed this tendency similarly on the other factor: the “Mid-high factor” (with the frequency loadings around 700 Hz and 2200 Hz). Secondly, the factor scores of obstruents were highly distributed between the “Low factor” and the “High factor”.

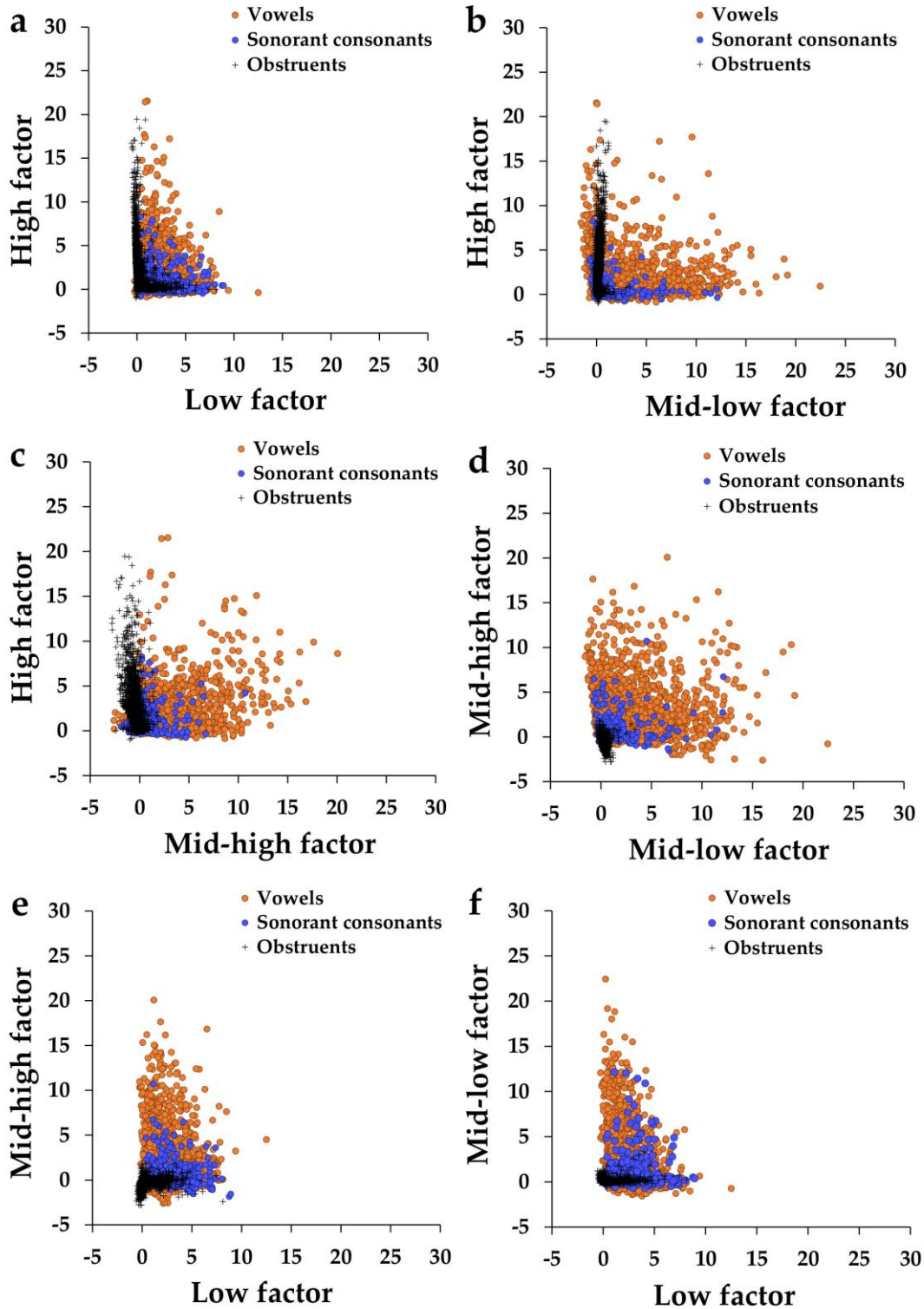


Figure 4.2. The distribution of the factor scores by origin-shifted factor analysis with cepstral liftering. (a–f) represent all possible combinations of two factors out of the four factors.

To further compare whether using cepstral liftering would affect the results of the distribution of factor scores in four spectral factors, a four-factor analysis was performed without cepstral liftering. The distribution of the factor scores is shown in Figure 4.3 below. Panels (a), (b), (c), (d), (e), and (f) show the distributions of six combinations of the four spectral factor scores. The distributions of the factor scores in Figure 4.2 and Figure 4.3 were very similar. The detailed distributions of Figure 4.3 were shown in Appendix D. Although the factor loadings on the “Low factor” (around 300 Hz) and the first peak (around 700 Hz) of the factor loadings on the double-peaked “Mid-high factor” were a little bit higher than without performing the cepstral liftering (Figure 4.1), the distribution of the factor scores of the “Low factor” was not greatly affected, by comparing the factor scores of the horizontal axis (the “Low factor”) between Figure 4.2 (a), (e), (f) and Figure 4.3 (a), (e), (f).

Moreover, the distributions of vowels and sonorant consonants on the “Mid-low factor” in Figure 4.2 (b), (d), (f) appeared to be a little wider than those in Figure 4.3 (b), (d), (f) by using cepstral liftering. The similar tendencies of vowels and sonorant consonants also appeared on the “Mid-high factor”. As the distributions shown in Figure 4.2 (b), (c), (e) were a little wider than those shown in Figure 4.3 (b), (c), (e), or on the “High factor” as the distributions shown in Figure 4.2 (a), (b), (c) were a little wider than those shown in Figure 4.3 (a), (b), (c) by using cepstral liftering. However, the distribution of obstruents hardly showed any changes whether cepstral liftering was used or not.

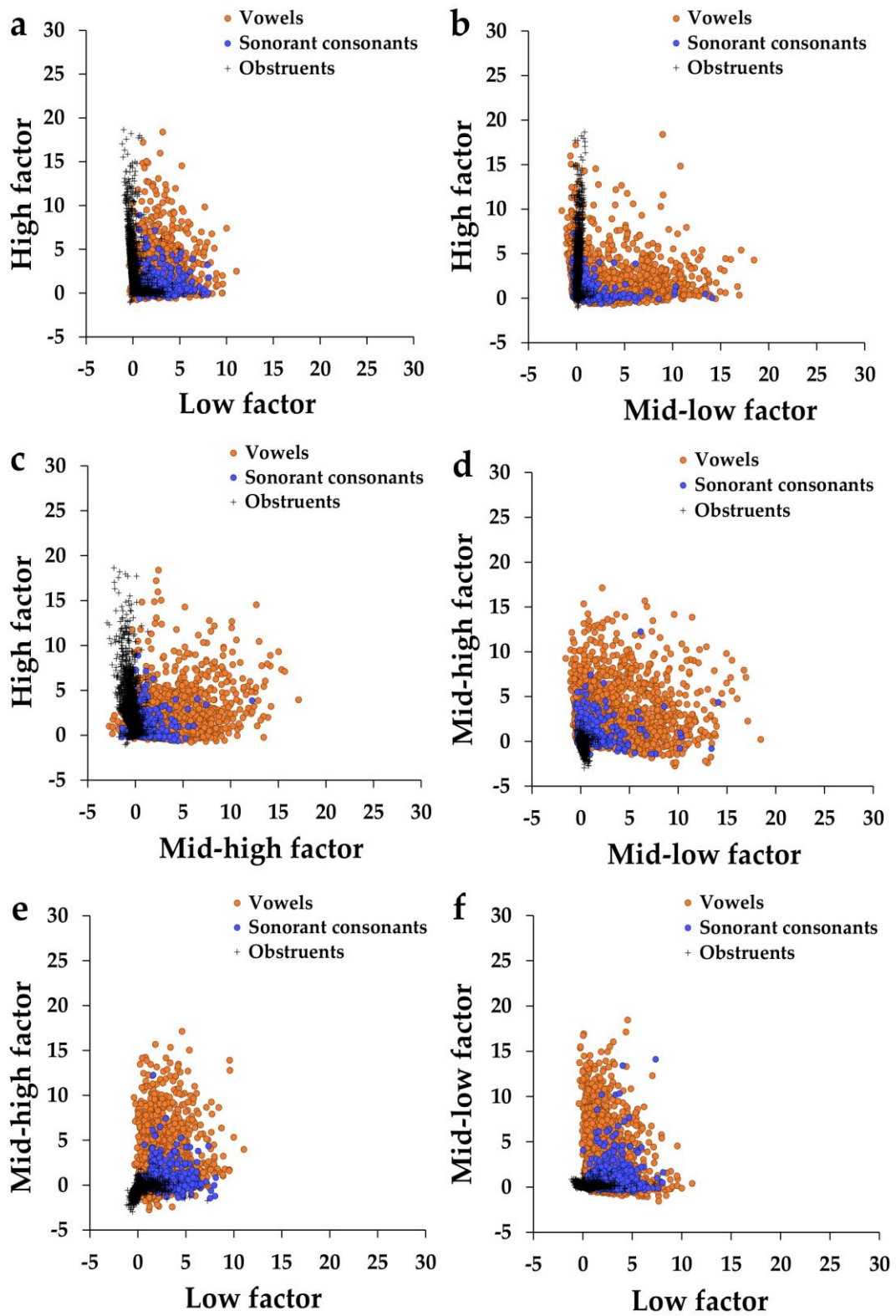


Figure 4.3. The distribution of the factor scores by origin-shifted factor analysis without cepstral liftering. (a–f) represent all possible combinations of two factors out of the four factors.

4.3.3 General distribution of English phonemes in the factor space

The results of the factor scores were highly similar with and without cepstral liftering, and therefore it was generally observed that on the “Mid-low factor”, with a center frequency range of around 1200 Hz, the highest factor scores were obtained by vowels, followed by sonorant consonants, and obstruents. A similar tendency also appeared on the “Mid-high factor”, and in a clearer tendency, by comparing the horizontal axis in Figure 4.2 (b) or 4.3 (b) and Figure 4.2 (c) or 4.3 (c). On the “Mid-high factor”, the factor scores of vowels were very high, while the distribution of sonorant consonants was almost half that of the vowels, and the obstruents were the lowest, close to the origin. By contrast, obstruents were associated with the two other spectral factors: the “Low factor” and the “High factor”. When the factor score of the “Low factor” was higher, the factor score of the “High factor” was lower, and vice versa, displaying an L-shape as in Figure 4.2 (a) or 4.3 (a). The L-shaped distribution of the factor scores of the obstruents across the “Low factor” and the “High factor” as analyzed here also appeared in the three-factor analysis across the “Low & mid-high factor” and the “High factor” in Study 1 and Study 2.

To investigate the features of vowels, sonorant consonants, and obstruents in more detail, the average of the factor scores of each phoneme was calculated, as shown in Figure 4.4 and Figure 4.5. The difference between Figure 4.4 and Figure 4.5 was whether or not to use cepstral liftering. Figure 4.4 shows the distribution of the average factor scores of each English phoneme by origin-shifted factor analysis with cepstral liftering, while Figure 4.5 shows the distribution without cepstral liftering. The horizontal axis and the vertical axis of panels (a), (b), (c), (d), (e), and (f) in Figure 4.4

and Figure 4.5 show the distributions of six combinations of the four spectral factor scores, the same as in Figures 4.2 (a-f) and 4.3 (a-f). The average of the factor scores of each phoneme is similarly distributed, by comparing Figure 4.4 and Figure 4.5.

The average of the factor scores of each phoneme was between the “Low factor” (with the frequency loadings around 300 Hz), the “Mid-low factor” (with the frequency loadings around 1200 Hz), the double-peaked “Mid-high factor” (with the frequency loadings around 700 Hz and 2200 Hz), and the “High factor” (with the frequency loadings around 4100 Hz). The average factor scores of all English phonemes were distributed in the factor space: Vowels are the red circles, sonorant consonants are the blue circles, and obstruents are the black crosses. The label of each phoneme was individually marked around the phoneme by using the International Phonetic Alphabet (Jones, 2011).

The first results were about the distribution of vowels, sonorant consonants, and obstruents on the two factors: the “Mid-low factor” and the “Mid-high factor”. In Figure 4.4 (e, f) and 4.5 (e, f), when the horizontal axis was the “Low factor”, the distribution of vowels, sonorant consonants, and obstruents had a similar trend on either the “Mid-low factor” (Figure 4.4f and Figure 4.5f) or the “Mid-high factor” (Figure 4.4f and Figure 4.5f). It was consistent with the results of the distribution of factor scores shown in Figure 4.4 (e, f) and 4.5 (e, f), that is, vowels were the highest, followed by sonorant consonants, and finally obstruents. But the difference was that on the “Mid-low factor”, shown in Figure 4.4 (f) or Figure 4.5 (f), the vowels /ɑ:/, /ɔ/, and /au/ occupied the highest positions of vowels, and the sonorant consonants /r/, /w/, /l/ occupied the highest positions of sonorant consonants. On the “Mid-high factor”

factor, shown in Figure 4.4 (e) or Figure 4.5 (e), the vowels /e/, /eə/, and /aɪ/ occupied the highest positions, while the sonorant consonants /j/, /r/, and /l/ occupied the highest positions of sonorant consonants.

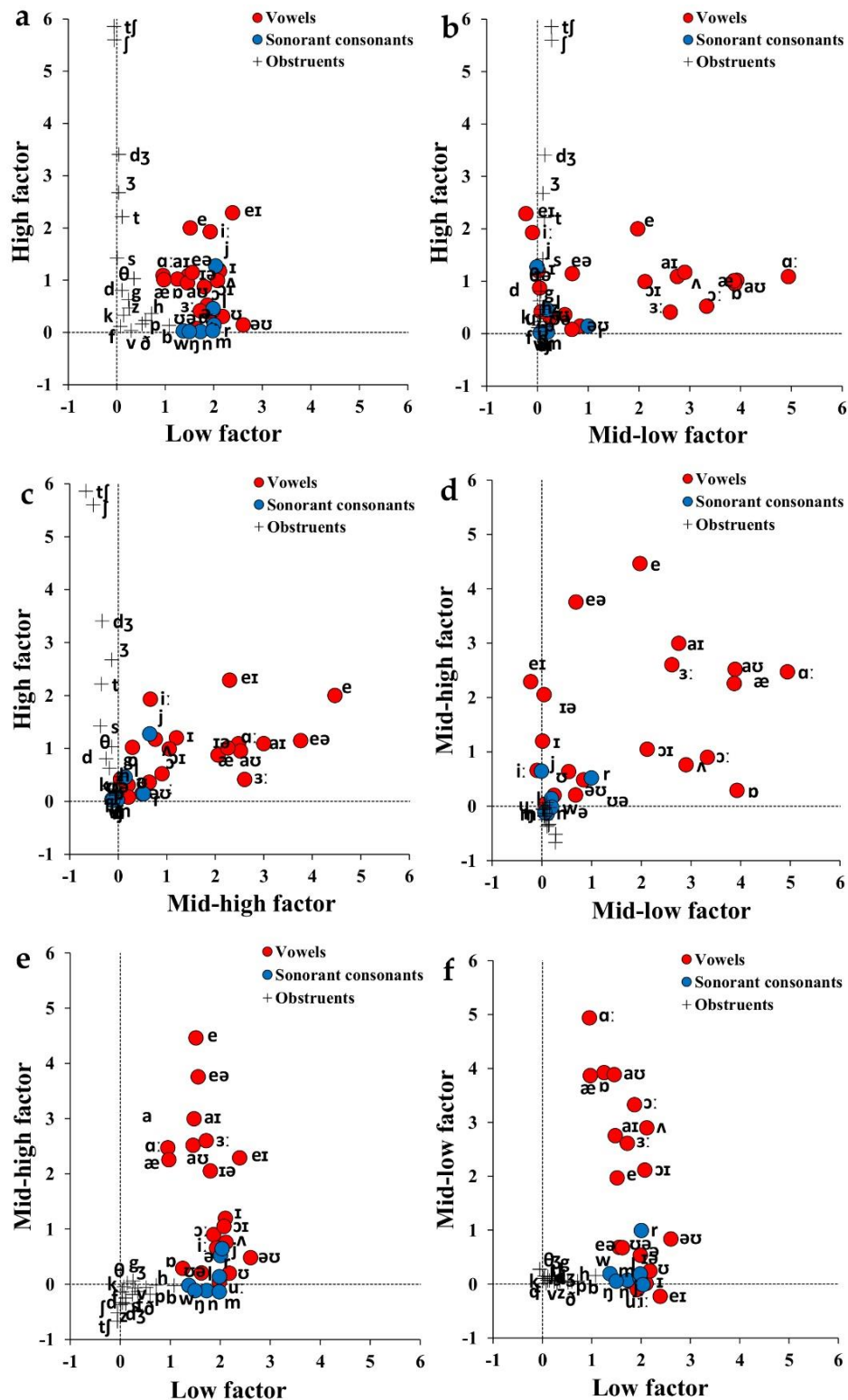


Figure 4.4. The distribution of the average factor scores of each English phoneme by origin-shifted factor analysis with cepstral liftering. (a–f) represent all possible combinations of two factors out of the four factors.

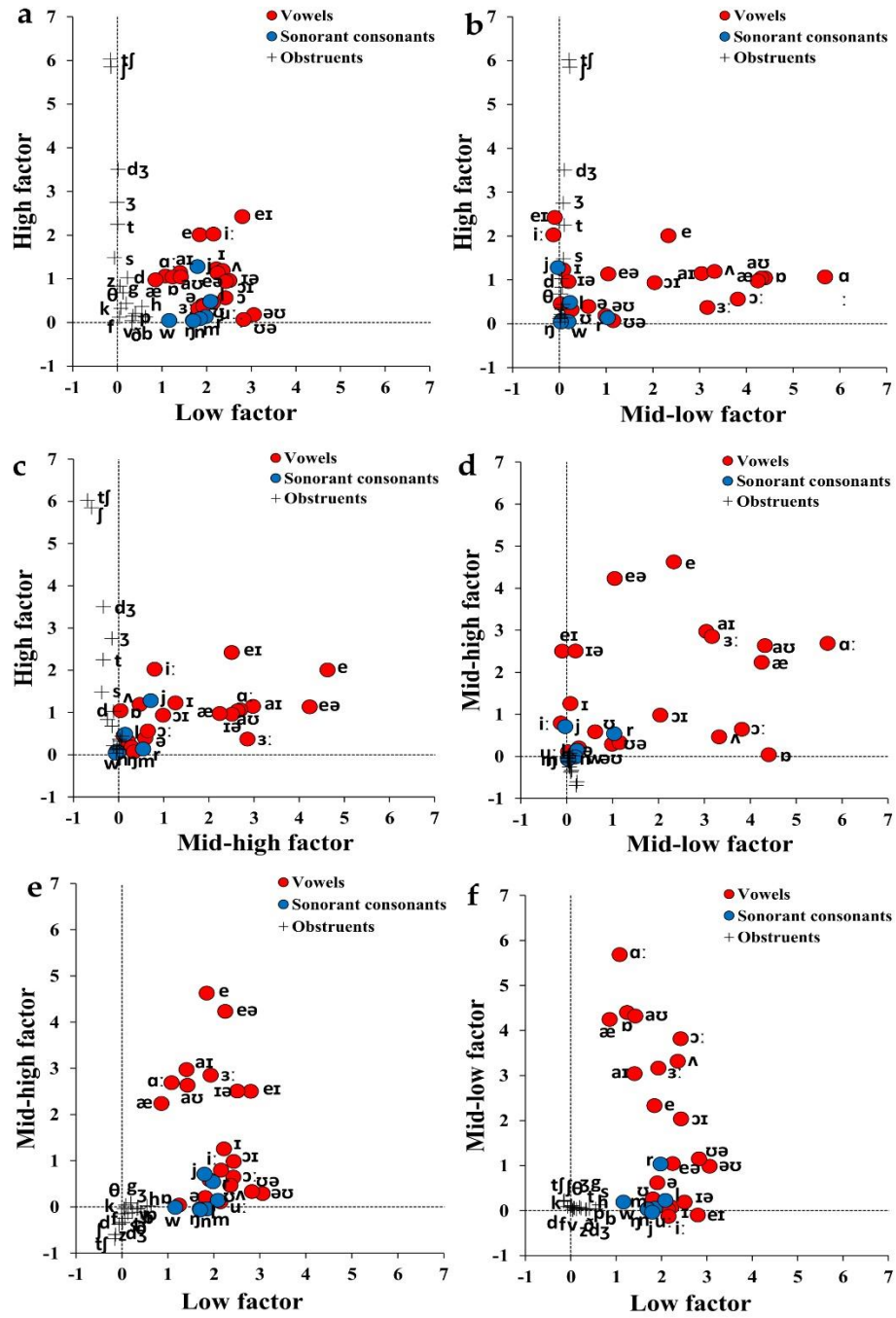


Figure 4.5. The distribution of the average factor scores of each English phoneme by origin-shifted factor analysis without cepstral lifiering. (a–f) represent all possible combinations of two factors out of the four factors.

The second result was that the distribution of the average factor scores of each English phoneme by origin-shifted factor analysis hardly showed any changes whether cepstral liftering was used or not (between Figure 4.4 a-f and Figure 4.5 a-f). In order to better observe the average factor scores of each English phoneme, Figures 4.6 to 4.11 show the enlarged distribution of Figure 4.5. Panel (a) in Figure 4.6 to Figure 4.11 represents the distribution of the average factor scores of each English phoneme in the two-dimensional space of four factors by origin-shifted factor analysis without cepstral liftering (cf. Figure 4.5a). Panel (b–d) represents the enlargement of panel (a) in three phoneme categories, i.e. vowels, sonorant consonants, and obstruents.

Figure 4.12 shows that the distribution of English phonemes obtained by origin-shifted factor analysis without cepstral liftering in the three-dimensional factor space. Panel (a) shows a two-dimensional mapping of the four factors, and the horizontal axis represents the combination of the “Mid-low factor” and the “Mid-high factor”. Panel (b) represents the enlargement of obstruents in panel (a). The results showed that for the combination of the “Mid-low factor” and the “Mid-high factor”, vowels, resonant consonants, and obstruents present a clearer ranking from high to low. Obstruents occupied the positions near the origin of these two factors. Moreover, on the “Low factor”, it can be clearly seen that vowels and sonorant consonants are distributed together and completely separated from obstruents. This also proves once again that the “Low factor” (with the frequency loadings around 300 Hz) could divide English phonemes into obstruents and non-obstruents (vowels and resonant consonants).

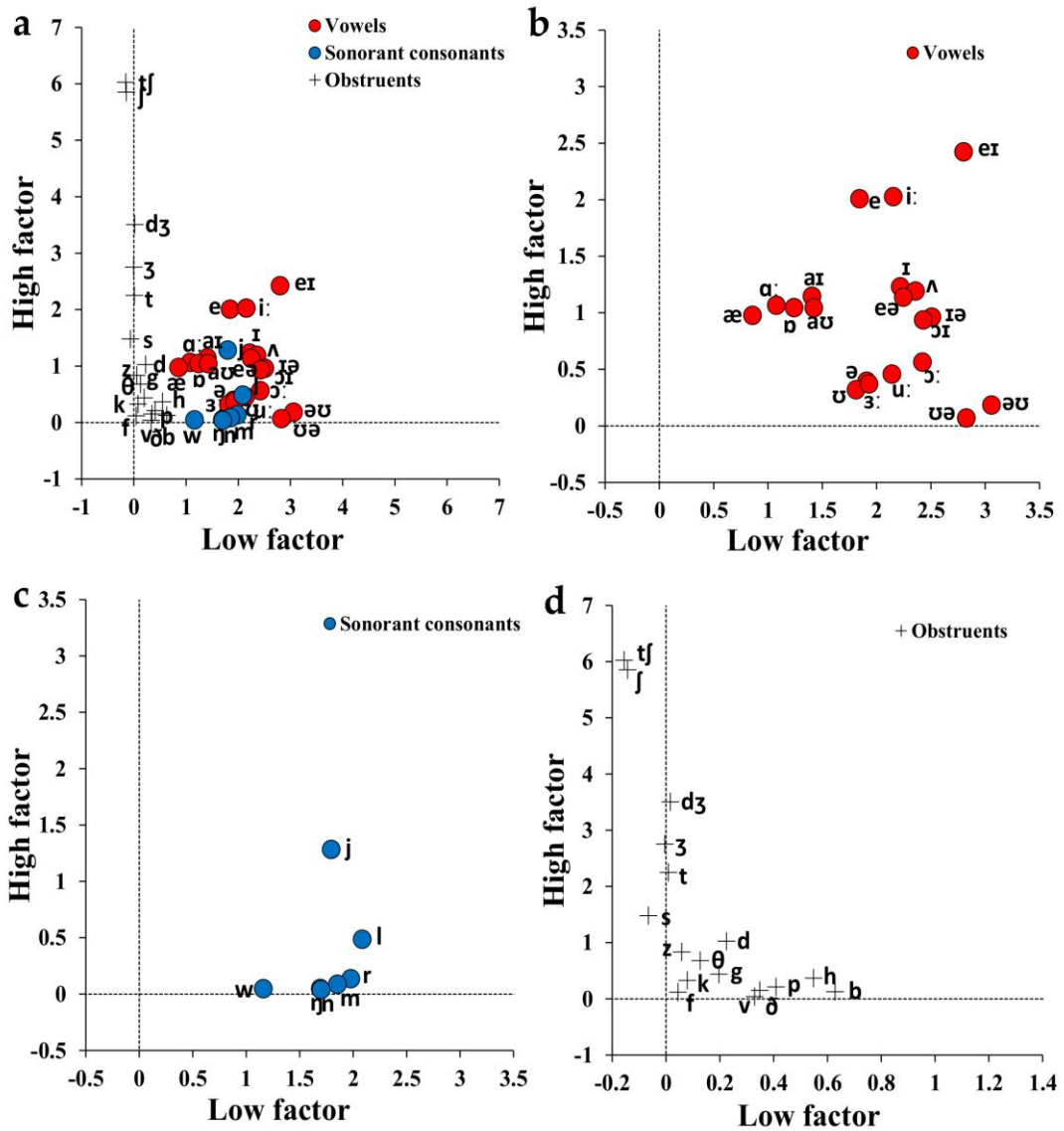


Figure 4.6. The enlargement of Figure 4.5 (a). Panel (a) represents the distribution of the average factor scores of each English phoneme in the two-dimensional space of the “Low factor” and the “High factor” by origin-shifted factor analysis without cepstral liftering (cf. Figure 4.5a). Panels (b–d) represent the enlargement of panel (a) in three phoneme categories, i.e. vowels, sonorant consonants, and obstruents.

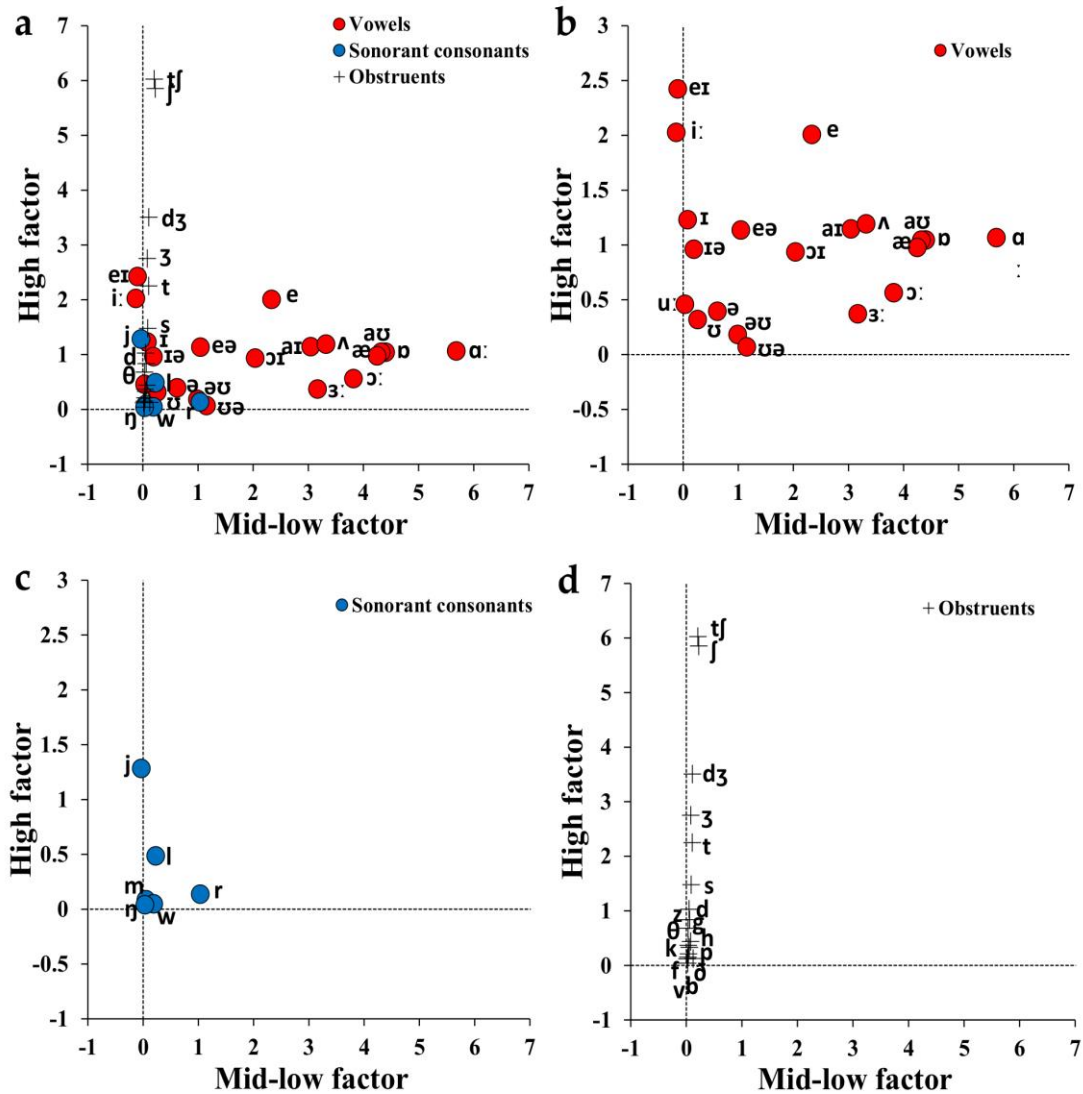


Figure 4.7. The enlargement of Figure 4.5 (b). Panel (a) represents the distribution of the average factor scores of each English phoneme in the two-dimensional space of the “Mid-low factor” and the “High factor” by origin-shifted factor analysis without cepstral liftering (cf. Figure 4.5b). Panels (b–d) represent the enlargement of panel (a) in three phoneme categories, i.e. vowels, sonorant consonants, and obstruents.

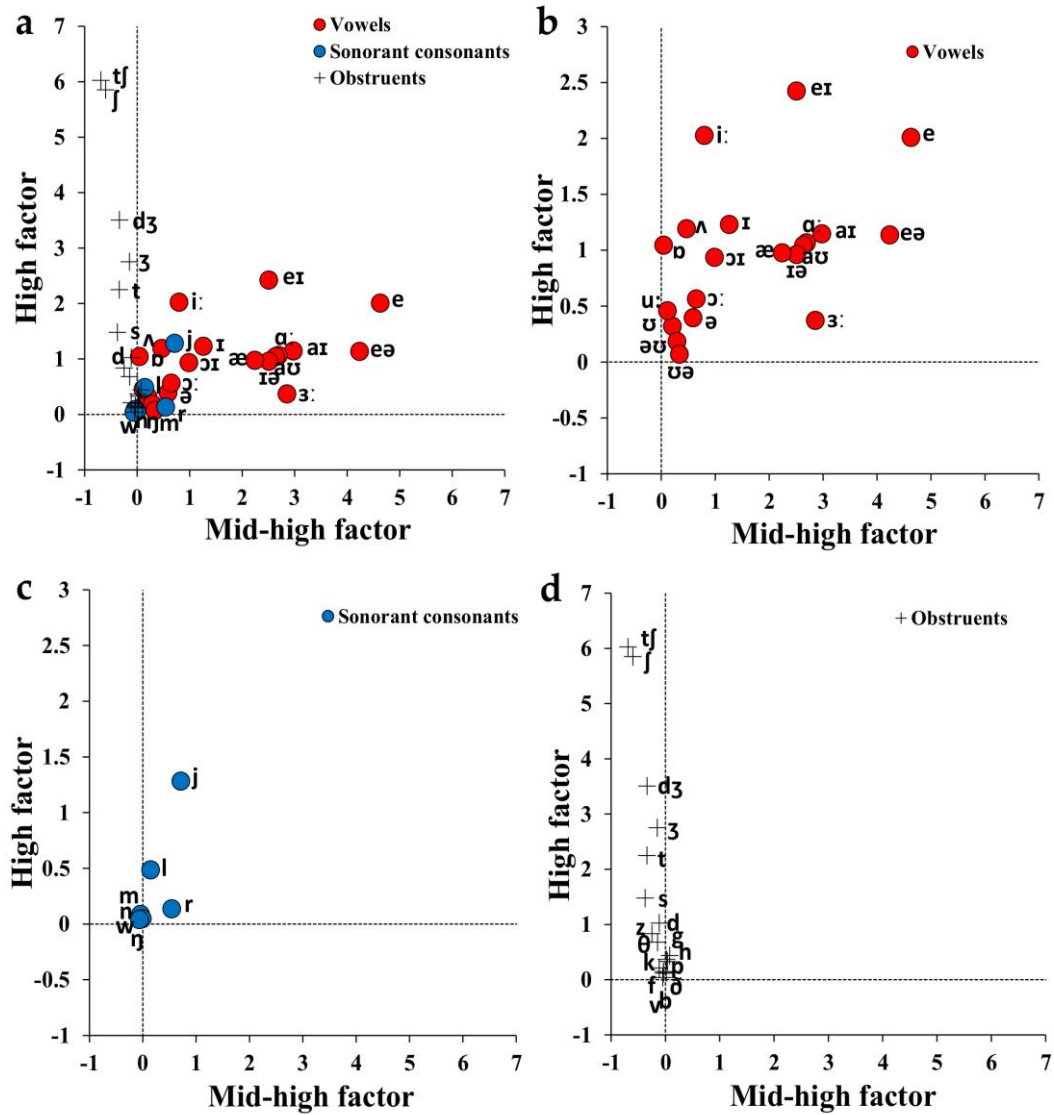


Figure 4.8. The enlargement of Figure 4.5 (c). Panel (a) represents the distribution of the average factor scores of each English phoneme in the two-dimensional space of the “Mid-high factor” and the “High factor” by origin-shifted factor analysis without cepstral liftering (cf. Figure 4.5c). Panels (b–d) represent the enlargement of panel (a) in three phoneme categories, i.e. vowels, sonorant consonants, and obstruents.

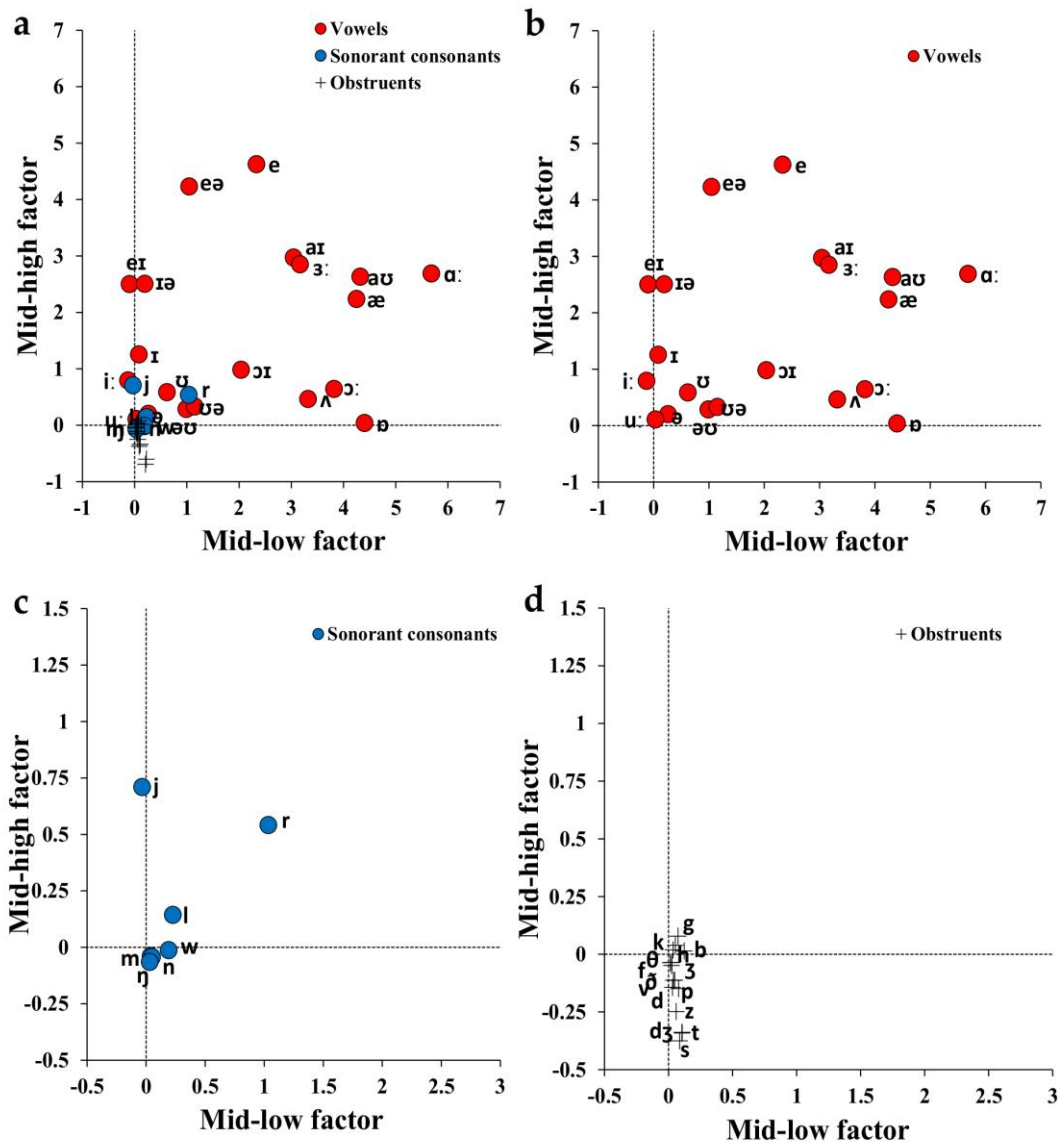


Figure 4.9. The enlargement of Figure 4.5 (d). Panel (a) represents the distribution of the average factor scores of each English phoneme in the two-dimensional space of the “Mid-low factor” and the “Mid-high factor” by origin-shifted factor analysis without cepstral liftering (cf. Figure 4.5d). Panels (b–d) represent the enlargement of panel (a) in three phoneme categories, i.e. vowels, sonorant consonants, and obstruents.

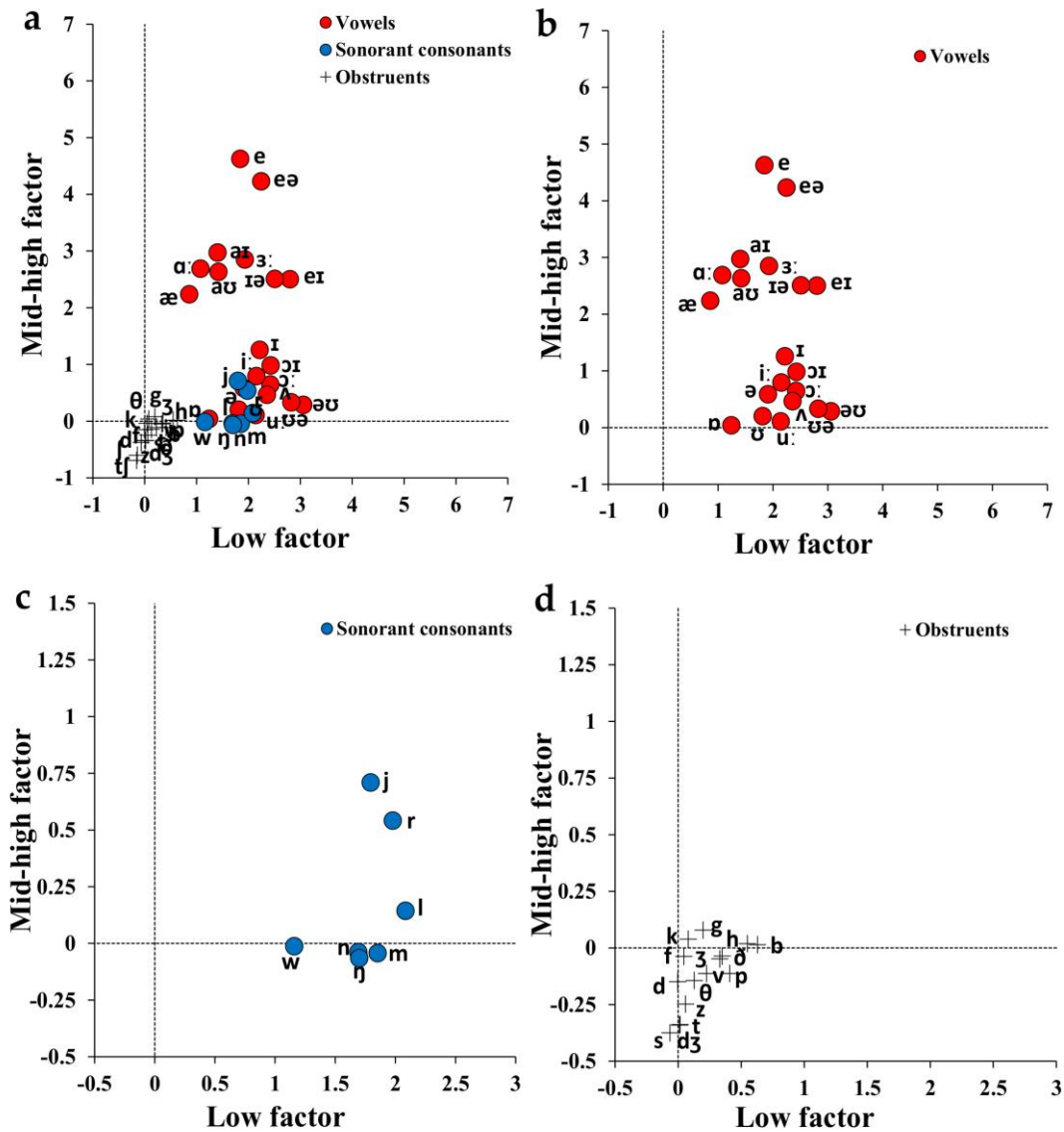


Figure 4.10. The enlargement of Figure 4.5 (e). Panel (a) represents the distribution of the average factor scores of each English phoneme in the two-dimensional space of the “Low factor” and the “Mid-high factor” by origin-shifted factor analysis without cepstral liftering (cf. Figure 4.5e). Panels (b–d) represent the enlargement of panel (a) in three phoneme categories, i.e. vowels, sonorant consonants, and obstruents.

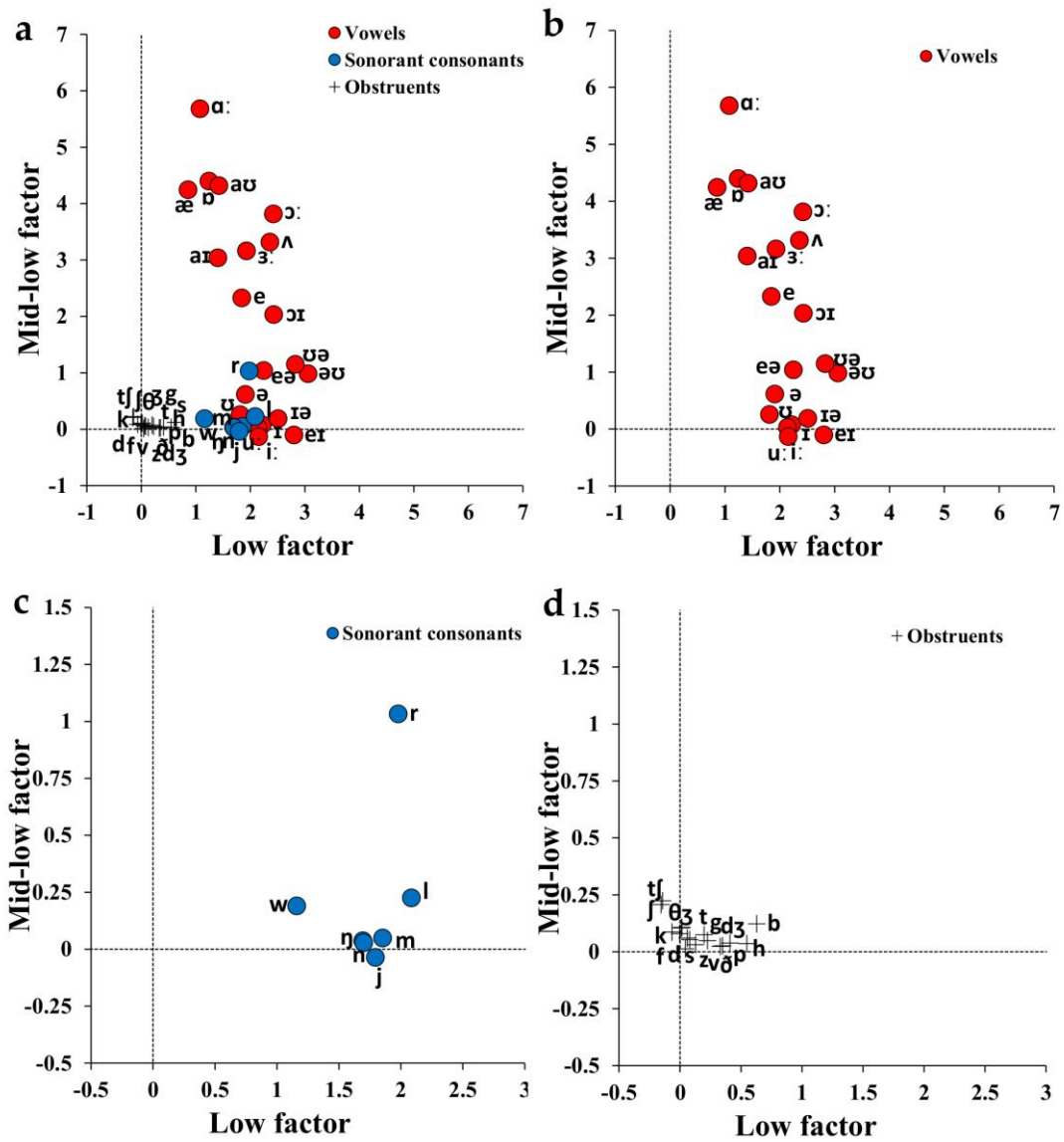


Figure 4.11. The enlargement of Figure 4.5 (f). Panel (a) represents the distribution of the average factor scores of each English phoneme in the two-dimensional space of the “Low factor” and the “Mid-low factor” by origin-shifted factor analysis without cepstral ltering (cf. Figure 4.5f). Panels (b–d) represent the enlargement of panel (a) in three phoneme categories, i.e. vowels, sonorant consonants, and obstruents.

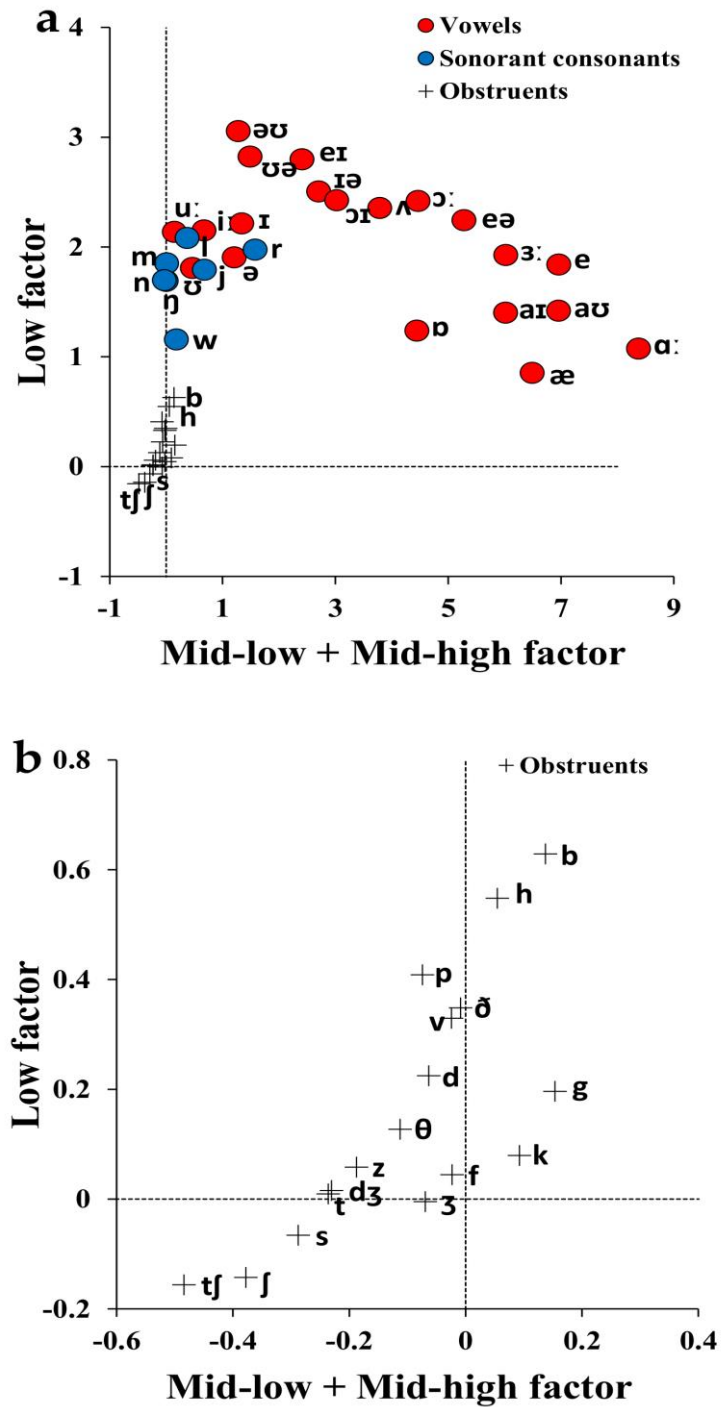


Figure 4.12. The distribution of English phonemes obtained by origin-shifted factor analysis without cepstral lustering in the three-dimensional factor space. Panel (a) shows a two-dimensional mapping of the four factors, and the horizontal axis represents the combination of the “Mid-low factor” and the “Mid-high factor”. Panel (b) represents the enlargement of obstruents in panel (a).

4.3.4 Distribution of English obstruents in the factor space

The third result was a new finding about obstruents. In Figure 4.4 (a, e, f), Figure 4.5 (a, e, f), or Figure 4.12 (a), on the “Low factor”, it was seen that the distribution of obstruents can be distinguished from vowels and sonorant consonants. This distribution never occurred for the other three factors. Regarding the obstruents, compared with Figure 4.4 (e) and (f) or Figure 4.5 (e) and (f), in Figure 4.4 (a) or Figure 4.5 (a), obstruents were very widely distributed on the “High factor” and occupied the highest position.

Moreover, in Figure 4.6 (d), Figure 4.7 (d), and Figure 4.8 (d), fricatives /tʃ/, /ʃ/, and /dʒ/ occupied the highest positions, while plosives always gathered around the origin. Plosives /t/, /d/ and /g/ were positioned slightly higher than other plosives. As shown in Figure 4.4 (d, e, f) or Figure 4.5 (d, e, f), most of the obstruents were concentrated around the origin of the “Mid-low factor” and the “Mid-high factor”. However, compared with the “Mid-low factor”, the distribution of the obstruents on the “Mid-high factor” tended to be in a lower direction.

In Figure 4.6 (d), on the “High factor”, the average factor scores of obstruents from high to low position were /tʃ/, /ʃ/, /dʒ/, /ʒ/, /t/, /s/, /d/, /z/, /θ/, /g/, /h/, /k/, /p/, /f/, /b/, /ð/, /v/. However, In Figure 4.12 (b), on the “Low factor”, the average factor scores of obstruents from high to low position were close to the opposite: The order was /b/, /h/, /p/, /ð/, /v/, /d/, /g/, /θ/, /k/, /z/, /f/, /dʒ/, /t/, /ʒ/, /s/, /ʃ/, /tʃ/. Moreover, on the combination of the “Mid-low factor” and the “Mid-high factor”, the average factor scores of some plosives (such as /b/, /g/, /k/) were higher than all fricatives or affricatives, especially fricative /s/.

4.4 Discussion and conclusions

The purpose of Study 3 was to further analyze the acoustic features of the extracted four spectral factors, by analyzing the same English speech sentences as used in Study 2. When the number of the extracted spectral factors increases, one of the factors will be separated (Ueda et al., 2008, Kishida et al., 2016, Ueda and Nakajima., 2017, and Ueda et al., 2023). This indeed also happened in the four-factor analysis, and the first purpose of Study 3 was to examine the separated factor. The results showed that the double-peaked “Low & mid-high factor” in the three-factor analysis, as described in Study 2, was separated into two factors: one was the “Low factor”, and the other was the “Mid-high factor”. As shown in Figure 4.1, the “Low factor” was related to the frequency range around 300 Hz. The “Mid-high factor” showed a double-peaked trend, related to two frequency ranges around 700 Hz and 2200 Hz, which never clearly appeared in Study 1 and Study 2. The other two factors were the same as in the three-factor analysis. One was the “Mid-low factor”, which was related to the frequency range of around 1200 Hz. The other was the “High factor”, related to the frequency range of around 4100 Hz.

The second purpose was to determine how the acoustic features of English phonemes (vowels, sonorant consonants, and obstruents) appeared in the distribution of factor scores by using origin-shifted factor analysis with or without cepstral liftering. Here it was confirmed again that cepstral liftering would hardly affect the results of the factor loadings and the factor scores. The factor loadings on the “Low factor” (around 300 Hz) and the first peak (around 700 Hz) of the factor loadings on the double-peaked “Mid-high factor” were a little bit higher when the cepstral liftering

was used. The distributions of factor scores of vowels, sonorant consonants, and obstruents together showed a similar tendency to the sonority hierarchy on both the “Mid-low factor” and the “Mid-high factor”. In Figure 4.2 (d) and Figure 4.3 (d), on either factor, the position of the distribution of the phonemes from high to low was vowels, followed by sonorant consonants, and finally obstruents. In the distribution of the average factor scores of the phonemes (Figure 4.3), however, phonemes showed characteristic configurations. Some vowels (i.e. /ɑ:/, /ɒ/, and /aʊ/) occupied the highest positions of vowels on the “Mid-low factor” (Figure 4.3f), while some other vowels (/e/, /eə/, and /aɪ/) occupied the highest positions of vowels on the “Mid-high factor” (Figure 4.4e and Figure 4.5e). Furthermore, sonorant consonants occupied similar places on the “Mid-low factor” and on the “Mid-high factor”, because the factor scores were around half of the factor scores of vowels (Figure 4.2d and Figure 4.3d). They were distributed in the positive direction. Obstruents, however, never went in the positive direction on the “Mid-low factor” and the “Mid-high factor”. As the conclusion, the present results corroborate the Sonority Sequencing Principle (Spencer, 1996): vowels and sonorant consonants have higher sonority and are more likely to be found closer to the nucleus or peak of a syllable. Obstruents have lower sonority: They cannot occur as the syllable nucleus and are likely to occur towards the boundaries of a syllable.

Finally, the third purpose of Study 3 was to specifically focus on how English obstruents appear in the obtained factor score spaces. A new finding of the present research is that the distribution of obstruents can be distinguished from non-obstruents (vowels and sonorant consonants) on the “Low factor”. This distribution never

occurred on the other three factors (Figure 4.4a, e, f and Figure 4.5a, e, f, and Figure 4.12a). The “Low factor” was located in a frequency range around 50-600 Hz, and had the peak frequency around 300 Hz. A conclusion in the three-factor analysis (Study 2) was that obstruents were prominent on the “High factor”, related to high frequency components around 4100 Hz. Thus, obstruents serve as syllable boundaries. The current results revealed that the frequency range below 600 Hz is also likely to serve as an important clue for distinguishing syllable boundaries.

Chapter 5. General discussion

The general purpose of the present dissertation is to elucidate how acoustic features of English speech can be expressed using multivariate analyses of speech spectra. To achieve this purpose, a series of factor analyses on English speech was performed, described here as Studies 1, 2, and 3.

There were three main research purposes. The first purpose was to investigate the suitability of multivariate analysis methods as applied to speech phonemes (Study 1 in Chapter 2). The second purpose was to determine the acoustic features of English phonemes in general by analyzing the results of the aforementioned factor analyses in relation to sonority, focusing on how English phonemes appear in the obtained factor score spaces (Study 1 in Chapter 2; Study 2 in Chapter 3; and Study 3 in Chapter 4). The third purpose was to obtain clear insight into whether and how obstruents would appear in the factor spaces (Study 1 in Chapter 2; Study 2 in Chapter 3; and Study 3 in Chapter 4).

The first purpose (related to Study 1) was to determine whether the use of normal factor analysis and a modified “origin-shifted” factor analysis caused differences in the extracted spectral factors when applied to English speech. To achieve this purpose, comparative analyses were performed between four different factor analysis methods applied to English speech. The differences between them were two processing aspects: (1) whether cepstral liftering was applied or not, and (2) whether the origin shift was used or not. The results showed that the factor loadings of one factor in a three-factor analysis, the “Low & mid-high factor”, were more pronounced when the analysis was without cepstral liftering, specifically on the

frequency of relatively high loadings around 2300 Hz. This suggested that cepstral analysis might have an effect on components with higher frequency bands in the signal spectrum. Cepstral liftering is a smoothing process of the speech signal (Rabiner & Schafer, 1978). However, it only affected factor loadings and had almost no effect on factor scores. There is little difference between the analyses with and without cepstral liftering, and for the present purposes it is recommendable not to use cepstral liftering because it is better not to change the original signal information. Furthermore, the distributions of the factor scores were similar whether cepstral liftering or the origin shift was used or not. Although the origin shift did not have a large effect on the factor scores, origin-shifted factor analysis is highly recommendable for speech resynthesis research and subsequent listening experiments. The reason was that it can remove a continuous background noise that almost always appears otherwise. Obstruents are always difficult to perceive because of their low-intensity natures. The obstruent consonants are more susceptible to noise than vowels (Li & Loizou, 2008). If the quality of the resynthesized speech is better, it will be related more closely to the real auditory signal, which makes the speech signal more useful for listening experiments. As a first conclusion of this dissertation, origin-shifted factor analysis without cepstral liftering is more recommendable for speech analyses. It was therefore used in the further analyses in Studies 2 and 3.

The second purpose (related to Studies 1, 2, and 3) was to determine the acoustic correlates of English phonemes in general in relation to sonority, by using origin-shifted factor analyses. An overview of the results is given in Table 5.1. Study 1, using a three-factor analysis, showed that a factor with a frequency range around

1100 - 1200 Hz (the “Mid-low factor”) was related to sonority. The order of factor scores from high to low on this “Mid-low factor” was similar to the sonority hierarchy for English phonemes, ranking vowels as highest, followed by sonorant consonants, and obstruents as lowest. Importantly, the same factor in the same frequency range was found with the newly-recorded database and the three-factor analysis (Study 2 in Chapter 3), and the four-factor analysis (Study 3 in Chapter 4). Together, this confirmed the existence of the “sonority” factor, which is only related to vowels and sonorant consonants.

Table 5.1. An overview of the results of the acoustic correlates of English phonemes concerning sonority by using origin-shifted factor analyses of Studies 1, 2, and 3 in Chapters 2, 3, and 4, respectively.

		Study 1 (Chapter 2)	Study 2 (Chapter 3)	Study 3 (Chapter 4)
Analyzed speech database	Name	“The ATR British English Speech Database” (Campbell, 1993)	“Fukuoka-Galway English Database”	
	Sampling frequency	12000 Hz	44100 Hz	
	Linear quantisation	16-bit	16-bit	16-bit
	Spoken sentences	200 sentences were spoken by 1 male and 2 females native speakers of English	100 sentences were spoken by 1 male and 2 females native speakers of English	
Extracted number of factors		3	3	4
Factor 1	Factor name	“Low & mid-high factor”	“Low & mid-high factor”	“Low factor”
	Frequency range	around 300 and 2300 Hz	around 300 and 2300 Hz	around 300 Hz
Factor 2	Factor name	“Mid-low factor”	“Mid-low factor”	“Mid-low factor”
	Frequency range	around 1100 Hz	around 1200 Hz	around 1200 Hz
Factor 3/4	Factor name	-	-	“Mid-high factor”
	Frequency range			around 700 Hz and 2200 Hz
	Factor name	“High factor”	“High factor”	“High factor”
	Frequency range	around 4100 Hz	around 4100 Hz	around 4100 Hz

In Study 3, the double-peaked “Low & mid-high factor” in the three-factor analysis (Study 1, 2) was separated into two factors: one was the “Low factor” (with a frequency range around 300 Hz), and the other was the “Mid-high factor” (with a double-peaked frequency range around 700 Hz and 2200 Hz). It is important that the “Mid-high factor” was also related to sonority. The distributions of factor scores on this factor were easily correlated to the sonority hierarchy. The frequency zone below 3000 Hz, in which the first three formants of vowels are located, is called the sonorant frequency zone (Harris, 1994; see also Nakajima et al., 2017). The “Mid-low factor” and the “Mid-high factor” were located in the frequency range from 600 to 2700 Hz in the present four-factor analysis. This strongly suggests that the acoustic aspects of English vowels are connected to sonority through these factors. Furthermore, the factor scores of sonorant consonants on the “Mid-high factor”, occupied similar distributions and on the “Mid-low factor”, because the factor scores were around half of the factor scores of vowels. As the second conclusion of this dissertation, the distributions of factor scores of the “Mid-low factor” and the “Mid-high factor” in the three- or four-factor analysis well reflected the rule of the Sonority Sequencing Principle (Spencer, 1996) in phonology. The phoneme with a higher position in the sonority hierarchy, is close to the syllable nucleus (Harris, 1994). Vowels are considered to have the highest sonority and always serve as syllable nuclei, while sonorant consonants play roles close to those of vowels, in that they can be next to the syllable nucleus, and even some of them can also be a syllable nucleus.

The third purpose of this dissertation (related to Studies 1, 2, and 3) was to obtain new insight about obstruents in the factor spaces. In phonology, English

phonemes with higher sonority are closer to the nucleus of the syllable, and conversely, phonemes with lower sonority are closer to a boundary of the syllable (Harris, 1994; Spencer, 1996). Therefore, obstruents are designated as the syllable boundaries. To determine this point of view through the perspective of acoustic features, this dissertation proposed three pieces of evidence by using factor analysis.

1. Regarding the distribution of factor scores, on the “Mid-low factor” of the three- and four-factor analysis and the “Mid-high factor” of the four-factor analysis that were highly correlated with sonority, obstruents were often near the origin and were almost not distributed in the positive direction. They had the lowest position among all categories of English phonemes. It proved that obstruents do not constitute the syllable nucleus, but rather delimit the syllable, corresponding to obstruents having the lowest position in the sonority hierarchy.
2. It was also found that obstruents were related to two factors: the “Low & mid-high factor” and the “High factor” in the three-factor analysis, and the “Low factor” and the “High factor” in the four-factor analysis. Whether obstruents were divided into voiced obstruents and voiceless obstruents or were divided into fricatives, affricates, and plosives, they were in a higher position on both factors, especially on the “High factor”. The lower part of the “Low & mid-high factor” in the three-factor analysis and the “Low factor” in the four-factor analysis were located at the lowest frequency range (50-600 Hz). The “High factor” in the three- and four-factor analysis was located at the highest frequency range (above 2700 Hz). The center frequencies of these two ranges

were 300 Hz and 4100 Hz, respectively. Obstruents seem highly related to these two frequency ranges.

3. The present research shows for the first time that the distribution of obstruents can be distinguished from non-obstruents (vowels and sonorant consonants) on the “Low factor”, where obstruents occupied the lowest position. Such configurations did not appear for the other three factors (the “Mid-low factor”, the “Mid-high factor”, and the “High factor”). This suggests that the fact that obstruents have a low position on the “Low factor” and can be a cue of syllabic boundary formation. The frequency range below 600 Hz is likely to serve as an important region for distinguishing syllable boundaries.

As for the future research on this topic, it is expected that the analysis results of this dissertation will be a useful starting point and will function as a resource for connecting sonority to phonology. The concept of sonority in phonology plays an essential role in the present acoustic analyses. However, sonority itself has been connected to acoustics only conceptually. It should also be empirically investigated from a psychophysical viewpoint in perceptual experiments, by focusing on the assumed role of sonority in the production and perception of syllables. As mentioned, obstruents are typically considered as syllable boundaries, while vowels and some sonorant consonants are designated as the syllable nucleus (Harris, 1994; Spencer, 1996). Harris (1994) argued that frequency components below 3000 Hz are closely related to vowels, and called this frequency range the *sonorant frequency zone*. Referring to the acoustic analysis results of the present study, obstruents occupied the

highest position of factor scores on the “High factor” (with the frequency range above 2700 Hz) of the three- or four-factor analysis, indicating that obstruents could become the boundaries of syllables. For example, English phonemes containing frequency components above 2700 Hz but lacking components below 600 Hz should serve as syllable boundaries, and English phonemes containing frequency components below 600 Hz and between 600 and 2700 Hz should serve as syllable nuclei. A potential procedure of a perceptual experiment is to select some English phonemes that had higher factor scores only on the “High factor” and another set of English phonemes that had higher factor scores on either the “Mid-low factor” or the “Mid-high factor”, and make a speech chain by resynthesizing and reconnecting these phonemes (referring to the sequence in Figure 4.12). We can then ask the participants to judge the number of syllables they perceived. This way, one can obtain some perceptual results about the syllable perception of chimeric English words, and it is expected that this kind of experiment can help to clarify the syllable structure related to sonority.

The present results show for the first time that the distributions of factor scores of vowels, sonorant consonants, and obstruents following a similar tendency as that in the sonority hierarchy on two factors in the factor space: the “Mid-low factor” and the “Mid-high factor”. More research is necessary to investigate how the acoustic features of English phonemes differ between these two spectral factors. Supported by future acoustic analyses and the psychological listening experiments, it is expected to provide more possibilities for exploring the speech perception in acoustics and linguistics.

References

- Awan, S. N., Giovinco, A., & Owens, J. (2012). Effects of vocal intensity and vowel type on cepstral analysis of voice. *Journal of Voice*, 26(5), 670-e15. <https://doi.org/10.1016/j.jvoice.2011.12.001>
- Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer* (Version 5.3.53) [Computer program]. <https://www.praat.org/>
- Campbell, N. (1993). The ATR British English speech database. *ATR Interpreting Telephony Research Labs*. ATR: Kyoto, Japan. TR-I-0363.
- de Saussure, F. (1959). *Course in general linguistics* (pp. 38-49). New York: Philosophical Library.
- Frigo, M., & Johnson, S. G. (2020). *The fastest fourier transform in the west* (Version 3.3.10). <https://www.fftw.org/fftw3.pdf>
- Féry, C., & van de Vijver, R. (2003). *The syllable in optimality theory*. Cambridge University Press.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Hoequist, Jr, C. (1983). Syllable duration in stress-, syllable-and mora-timed languages [J]. *Phonetica*, 40(3), 203-237. <https://doi.org/10.1159/000261692>
- Harris, J (1994). *English Sound Structure* (pp. 53-81). Blackwell Press.
- Harris, J. (2006). The phonology of being understood: Further arguments against sonority. *Lingua*, 116(10), 1483–1494. <https://doi.org/10.1016/j.lingua.2005.07.009>
- Hayes, B. (2008). *Introductory phonology* (1st edition). Wiley-Blackwell Press.
- Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, 40(2), 269–279. <https://doi.org/10.1016/j.wocn.2011.11.001>

- Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Prentice Hall Press.
- Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer Press.
- Jones, D. (2011). *Cambridge English pronouncing dictionary* (18th ed.). Cambridge University Press.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, 374(2065).
<https://doi.org/10.1098/rsta.2015.0202>
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187-200.
- Kishida, T., Nakajima, Y., Ueda, K., & Remijn, G. B. (2016). Three factors are critical in order to synthesize intelligible noise-vocoded Japanese speech. *Frontiers in Psychology*, 7, 517. <https://doi.org/10.3389/fpsyg.2016.00517>
- Lennig, M., & Brassard, J. P. (1984). Machine-readable phonetic alphabet for English and French. *Speech Communication*, 3(2), 165-166.
- Li, N., & Loizou, P. C. (2008). The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise. *The Journal of the Acoustical Society of America*, 124(6), 3947-3958. <https://doi.org/10.1121/1.2997435>
- Møller, H., & Pedersen, C. S. (2004). Hearing at low and infrasonic frequencies. *Noise & Health*, 6(23), 37-57.
- Nyquist, H. (1928). Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2), 617-644.
- NTT-AT. (2002). *Multilingual Speech Database 2002*. Tokyo: NTT Advanced Technology Corporation; Available online at: <http://www.ntt-at.com/product/speech2002/> (Accessed December 1, 2015)
- Nakajima, Y., Ueda, K., Fujimaru, S., Motomura, H., & Ohsaka, Y. (2017). English phonology and an acoustic language universal. *Scientific Reports*, 7(1), 46049.

<https://doi.org/10.1038/srep46049>

- Nakajima, Y., Ueda, K., Remijn, G. B., Yamashita, Y., & Kishida, T. (2018). How sonority appears in speech analyses. *Acoustical Science and Technology*, 39(3), 179–181. <https://doi.org/10.1250/ast.39.179>
- Ohala, D. K. (1999). The influence of sonority on children's cluster reductions. *Journal of Communication Disorders*, 32(6), 397–422. [https://doi.org/10.1016/S0021-9924\(99\)00018-0](https://doi.org/10.1016/S0021-9924(99)00018-0)
- Plomp, R., Pols, L. C. W., & van de Geer, J. P. (1967). Dimensional Analysis of Vowel Spectra. *The Journal of the Acoustical Society of America*, 41(3), 707–712. <https://doi.org/10.1121/1.1910398>
- Pols, L. C., Tromp, H. R., & Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *The Journal of the Acoustical Society of America*, 53(4), 1093–1101. <https://doi.org/10.1121/1.1913429>
- Plomp, R. (1976). *Aspects of tone sensation: A psychophysical study*. Academic Press.
- Price, P. J. (1980). Sonority and syllabicity: Acoustic correlates of perception. *Phonetica*, 37(5-6), 327–343. <https://doi.org/10.1159/000260001>
- Plomp, R. (2001). *The intelligent ear: on the nature of sound perception*. Psychology Press.
- Parker, S. (2008). Sound level protrusions as physical correlates of sonority. *Journal of Phonetics*, 36(1), 55–90. <https://doi.org/10.1016/j.wocn.2007.09.003>
- Rabiner L. R., & Schafer R. W. (1978). *Digital processing of speech signals*. Prentice Hall Press.
- Selkirk, E. (1984). On the major class features and syllable theory. *Language sound structure* (pp. 107-136). MIT Press.
- Spencer, A. (1996). *Phonology: theory and description*. Blackwell Press.
- Smith, J. O., & Abel, J. S. (1999). Bark and ERB bilinear transforms. *IEEE Transactions on Speech and Audio Processing*, 7(6), 697–708. <https://doi.org/10.1109/89.799695>

- Ueda, K., & Nakajima, Y. (2008). A consistent clustering of power fluctuations in British English, French, German, and Japanese. *Technical Committee of Psychological and Physiological Acoustics*, 38, 771–776.
- Ueda, K., Nakajima, Y., & Satsukawa, Y. I. (2010). Effects of frequency-band elimination on syllable identification of Japanese noise-vocoded speech: Analyses of confusion matrices. *Proceedings of Fechner Day*, 26(1), 33-38.
- Ueda, K., & Nakajima, Y. (2017). An acoustic key to eight languages/dialects: Factor analyses of critical-band-filtered speech. *Scientific Reports*, 7, 42468. <https://doi.org/10.1038/srep42468>
- Ueda, K., Araki, T., & Nakajima, Y. (2018). Frequency specificity of amplitude envelope patterns in noise-vocoded speech. *Hearing Research*, 367, 169–181. <https://doi.org/10.1016/j.heares.2018.06.005>
- Ueda, K., Doan, L. L. D., & Takeichi, H. (2023). Checkerboard and interrupted speech: Intelligibility contrasts related to factor-analysis-based frequency bands. *The Journal of the Acoustical Society of America*, 154(4), 2010-2020. <https://doi.org/10.1121/10.0021165>
- Wedin, S., Leanderson, L., & Wedin, L. (1978). Evaluation of voice training spectral analysis compared with listeners' judgements. *Folia Phoniatrica et Logopaedica*, 30(2), 103-112. <https://doi.org/10.1159/000264116>
- Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *The Journal of the Acoustical Society of America*, 33(2), 248. <https://doi.org/10.1121/1.1908630>
- Zahorian, S. A., & Rothenberg, M. (1981). Principal-components analysis for low-redundancy encoding of speech spectra. *The Journal of the Acoustical Society of America*, 69(3), 832–845. <https://doi.org/10.1121/1.385539>
- Zsiga, E. C. (2013). *The sounds of language: an introduction to phonetics and phonology* (Vol. 7). Wiley-Blackwell Press.

Appendix A. The 100 speech sentences used in the Fukuoka-Galway English Speech Database (Campbell, 1993)

1. The price range is smaller than any of us expected.
2. They asked if I wanted to come along on the barge trip.
3. Amongst her friends she was considered beautiful.
4. The smell of the freshly ground coffee never fails to entice me into the shop.
5. I'm often perplexed by rapid advances in state of the art technology.
6. John could lend him the latest draft of his work.
7. From forty love the score was now deuce and the crowd grew tense.
8. The Presbyterian minister managed to curb the drinking habits of the loitering youths.
9. The bulb blew when he switched on the light.
10. It is futile to offer any further resistance.
11. They launched into battle with all the forces they could muster.
12. The chill wind caused them to shiver violently.
13. The government triumphed four years ago and we have every reason to believe that it will triumph again.
14. He jerked round in an instant to face his assailant.
15. He emphasized his strengths while concealing his weaknesses.
16. The table is made so sloppily that it tilts.
17. It was important to be perfect since there were no prompts.
18. I ran for cover whilst he hurled several stones.
19. We have proof that the regime wields sufficient power in the North to exploit the entire population.
20. She flicks through a magazine when she gets a chance.
21. Thank goodness it's Friday and time to go home.
22. Itches are always so tempting to scratch.
23. I'll hedge my bets and take no risks.
24. The length of her skirt caused the passers-by to stare.
25. I always seem to follow my instincts rather than reason.

26. It's strange that I slept for so long since I wasn't feeling tired.
27. I'll draft those new proposals before the next meeting.
28. The mud squelched loudly and he realized that his suede boots were doomed.
29. He grasped the rope with his free hand and swung his legs round to the side.
30. Opportunities like this don't grow on trees as you should know by now.
31. It's difficult to choose between two such equally good alternatives.
32. Henry normally viewed Bob's zealous enthusiasm with disdain.
33. She'll think of an excuse if given enough space and time.
34. When forced to make a choice, Sarah chose ping-pong as her favorite game.
35. It's a shame that architects design for themselves and not for the general public.
36. Coe beat him to the line by five thousandths of a second.
37. The bath plug is missing so you'll have to take a shower.
38. You ought to brush your teeth before you go to bed.
39. The earth used to be flat but now it's a sphere.
40. I wish he'd either grow a beard or shave his moustache.
41. Judith found the manuscripts waiting for her on the piano.
42. The huge castle was encircled by a deep moat.
43. Jane adored maths and French but hated the rest of school.
44. Mashed potatoes are more fattening than either boiled or baked ones.
45. The world is becoming increasingly dangerous but hardly anyone cares.
46. Patrick will need speech therapy because of his cleft palate.
47. These practical jokes have been taken much too far.
48. I shall paint this room mauve with a few beige dots.
49. My car was stolen while I was shopping in Kensington High Street.
50. Changing gear half way up a steep hill can be quite risky.
51. I always enjoy a pint of lager when I come off the squash court.
52. Jean might prepare more salmon and cucumber sandwiches if we're lucky.
53. There'll be big trouble if you dare to touch that surface.
54. Tom says that ancient Saabs are far more stylish than British Leyland Triumphs.
55. Jill tumbled after Jack because she was shoved by little Miss Muffet.
56. Doctor Philips raised a number of moot points about the professor's article in the

recent journal.

57. This new display attracts more customers than ever.
58. Mrs. Stewart pleased her entire family when she produced malt loaf for tea.
59. Pete couldn't bear to show us his scar so soon after the accident.
60. Joy fibbed about the money missing from the secret cashbox.
61. The hearse will arrive at the morgue between a quarter and half past twelve.
62. Water was cascading down the mountain at a rate of knots.
63. Clara went through a phase when she always served Hungarian goulash followed by rhubarb crumble.
64. We'd be hard pushed to catch the bus to Newcastle tonight.
65. Our butcher makes his own pork and beef sausages.
66. Martin and Craig grow dwarf tulips and exhibit them all over the county.
67. Our two kids adored the little iced buns with the cherries on top.
68. He remembered he needed a passport to get a visa stamp.
69. His attempts to use his credit card failed because his account was in arrears.
70. They were still frisked by Customs though they had nothing to declare.
71. The nurse tended the fierce-looking wound on Paul's scalp.
72. Everyone talks of the birds and the bees but they never mention wasps.
73. Most of the scenes were filmed on location in the Alps.
74. Mr. Vernon helped himself to dessert and a large cup of coffee.
75. It was a sheer fluke that I bumped into the same chap yesterday.
76. During his last year at university they shared -many hilarious moments.
77. I loved to pedal my tricycle in the back-yard when I was a child.
78. Some people find the Welsh language very difficult to learn.
79. The food varies from place to place but the price remains fairly constant.
80. The ceremony overwhelmed me and I was moved to tears.
81. The area was so thoroughly blitzed that now virtually nothing remains.
82. Bob milked the cows after he'd gathered the chickens' eggs.
83. We really will need to defrost the fridge now that the strawberries have oozed out all over the place.
84. The coach swerved to the left in an attempt to avoid the blind pedestrian.

85. After the road veers to the south you'll see a junction ahead.
86. Her evening gown was a trifle too garish for the occasion.
87. He glimpsed the traffic warden out of the corner of his eye.
88. She had scarcely divulged the scandal before it was splattered over the front pages of the tabloids.
89. The questionnaire was short and to the point.
90. We were plunged into darkness as the clouds engulfed the moon.
91. Sharon vowed never again to jeer at the concierge.
92. Compared with the previous guests she appeared coy and demure.
93. She returned from holiday bronzed by the mediterranean sun.
94. They suggested that the ice would thaw as soon as spring arrived.
95. Gordon's words were lost amidst the hub-bub of the crowded hall.
96. Mr. Forsyth has cheered the team on now for five whole years.
97. It seems as if Susan does all the chores for this household.
98. I want each person here to give George their report by Thursday afternoon.
99. Beth chaired the first meeting in order to oblige her overworked boss.
100. I've just seen Charles tearing off in a hurry.

Appendix B. Informed consent and instructions for creating the Fukuoka-Galway English Speech Database

No. _____

Informed Consent Form for Participation in Observation/Experiment Comprehensive Investigation on Perception and Cognition: Speech Perception

Director/Supervisor	Kazuo Ueda, Ph.D.	Human Science, Faculty of Design, Kyushu University
Supervisor	Gerard B. Remijn, Ph.D.	Human Science, Faculty of Design, Kyushu University

Place of Experiment _____

1. We kindly ask you to collaborate with our research project(s) by participating in behavioral observation(s)/experiment(s) and/or non-invasive physiological experiment(s) on human perception and cognition. The obtained data are utilized to understand psychological/neurophysiological mechanisms of human perception and cognition. The experimenter will explain the procedures to be followed, how long they will take, their frequency, and the honorarium.
2. If you decide to participate in the observation(s)/experiment(s), risks to your mental/physical health are not expected.
3. Even after signing this informed consent form, you can always stop participating in the observation(s)/experiment(s) when you are not able or willing to. Discontinuation will never disadvantage or prejudice your future relations with Kyushu University and the supervisor(s)/experimenter(s) involved with the observation(s)/experiment(s).
4. The results of the observation(s)/experiment(s) may be presented in the form of a thesis, a conference presentation or a scientific paper, all in academic fields. Supervisors/experimenters will also have the liberty to make the results of the observations/experiments public via media. Any information obtained from your participation, including this informed consent form and face sheets, will be kept strictly confidential. In the case data are presented, every effort will be made to make name identification impossible by utilizing participant numbers. The supervisors/experimenters may utilize the same obtained data in different institutes/universities for scientific purposes related to the research.
5. If you understand the contents of the above explanation, and if you agree with us to participate in the observation(s)/experiment(s), please sign below:

Participant's name:

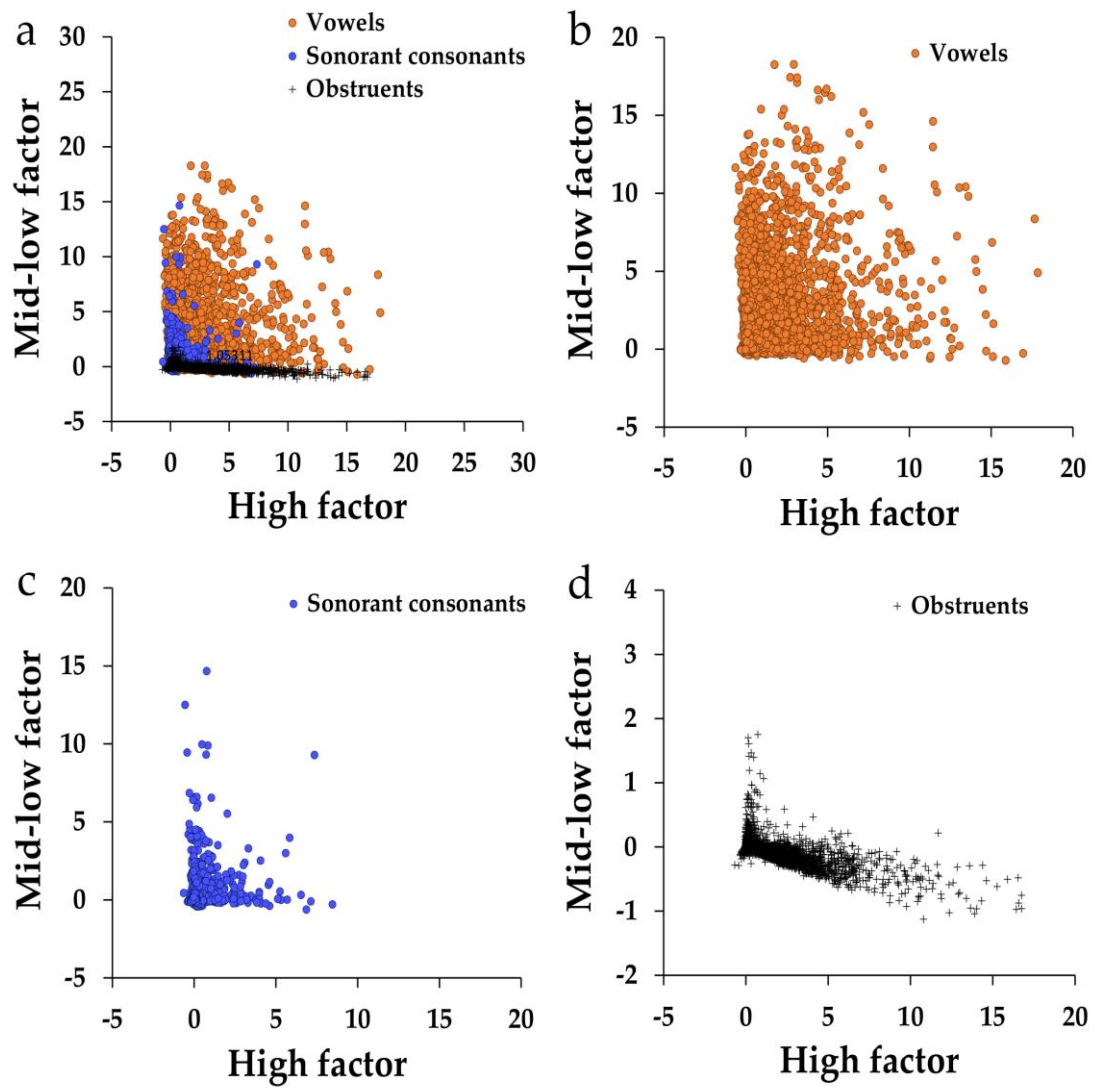
Signature by the participant or guardian:

Date (day/month/year):

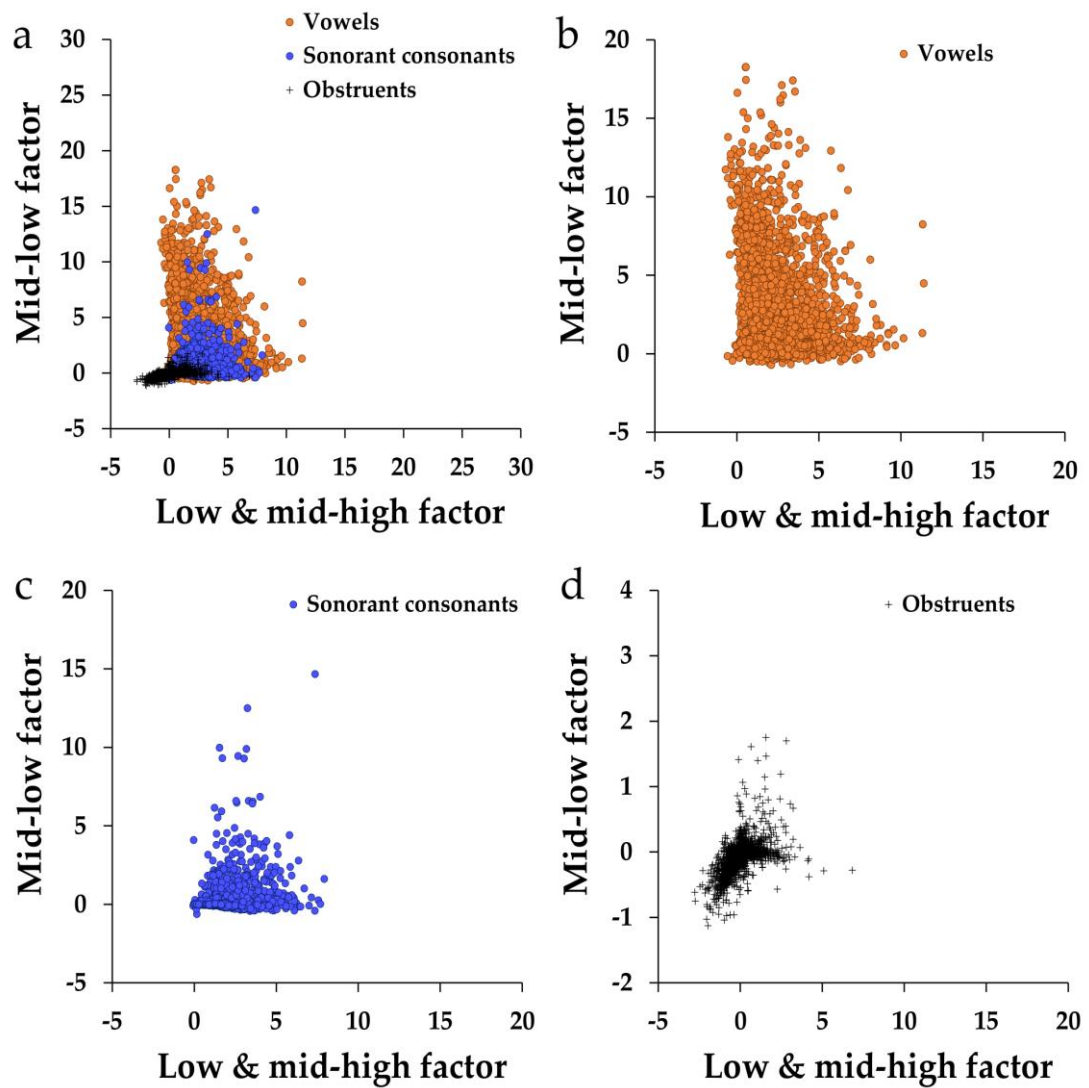
Contact address, phone number, e-mail address, or affiliation:

Appendix C. The expanded version of Figure 3.3 (a)-(d)

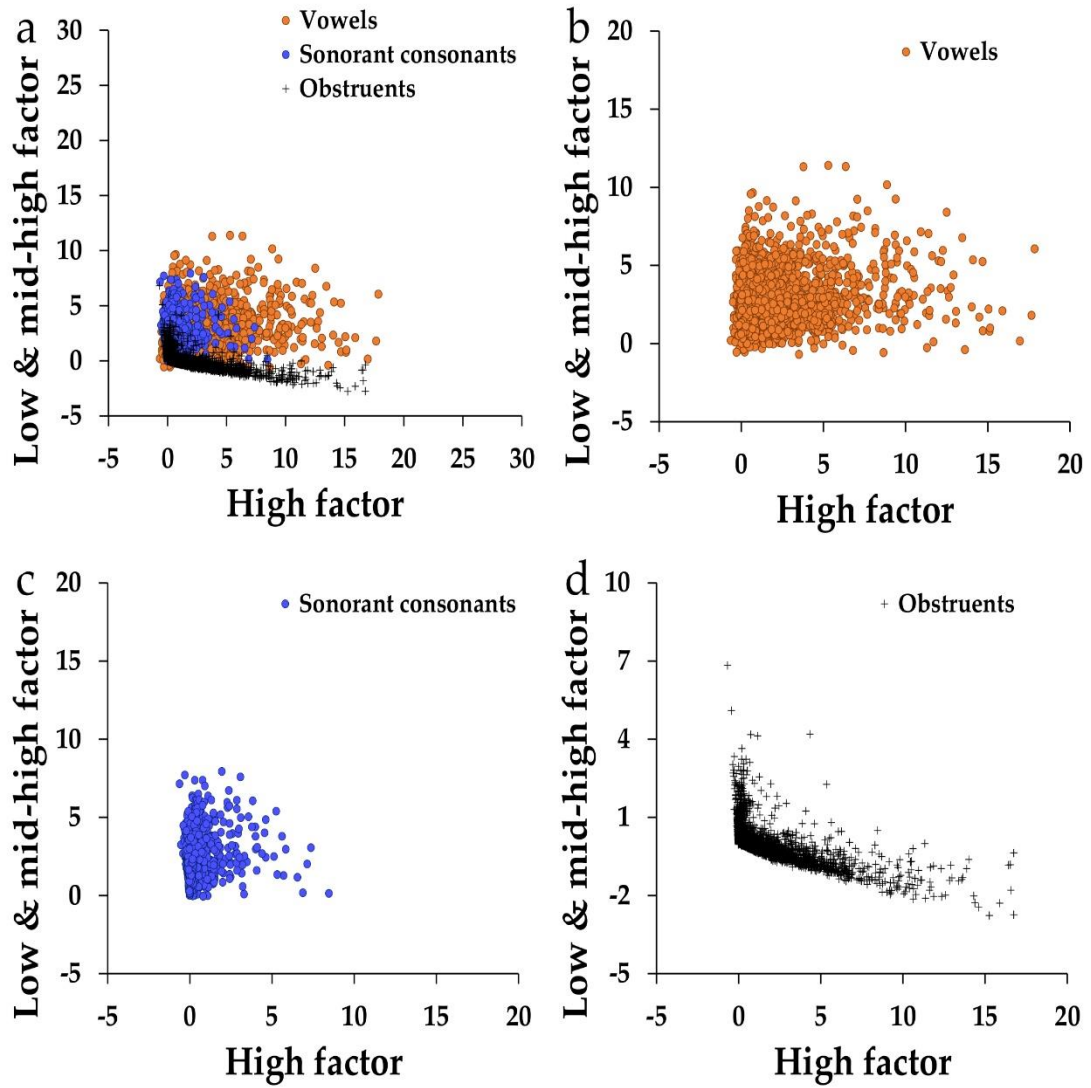
The distributions of the factor scores of vowels, sonorant consonants and obstruents in the two-dimensional space of the “Mid-low factor” and the “High factor”.



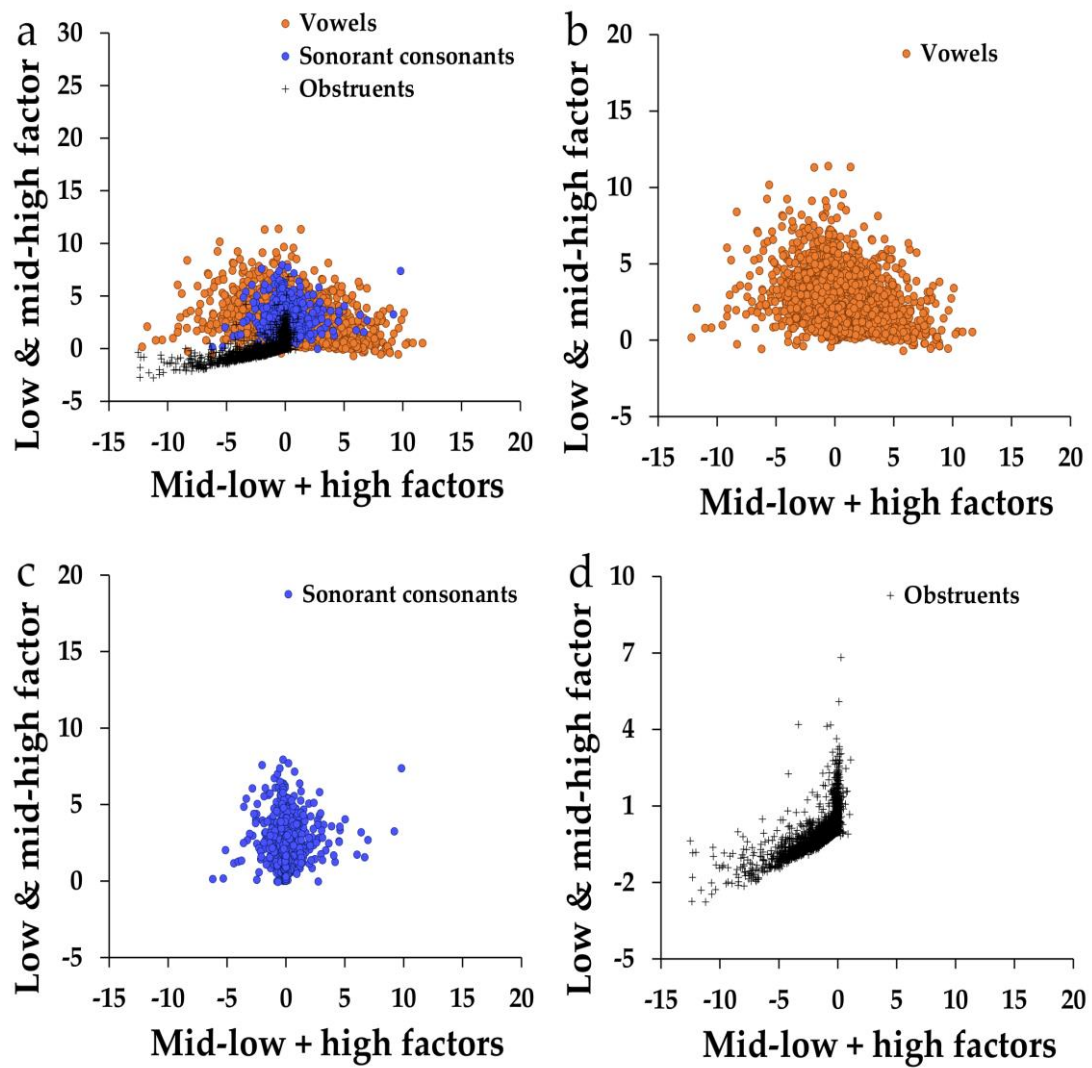
The distributions of the factor scores of vowels, sonorant consonants and obstruents in the two-dimensional space of the “Mid-low factor” and the “Low & mid-high factor”.



The distributions of the factor scores of vowels, sonorant consonants and obstruents in the two-dimensional space of the “Low & mid-high factor” and the “High factor”.

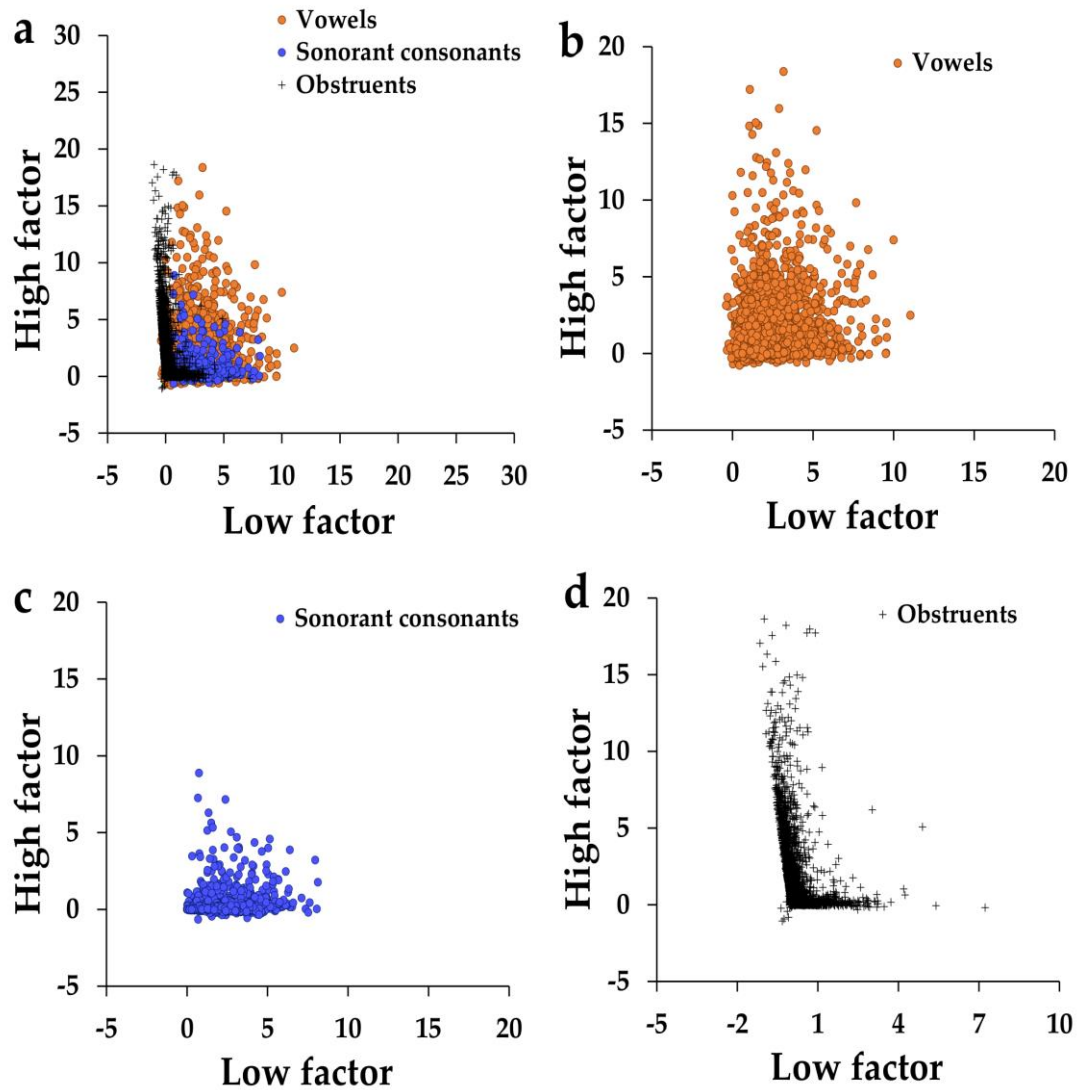


The distributions of the factor scores of vowels, sonorant consonants and obstruents in the two-dimensional space of the “Low & mid-high factor” and the “Mid-low + high factors”.

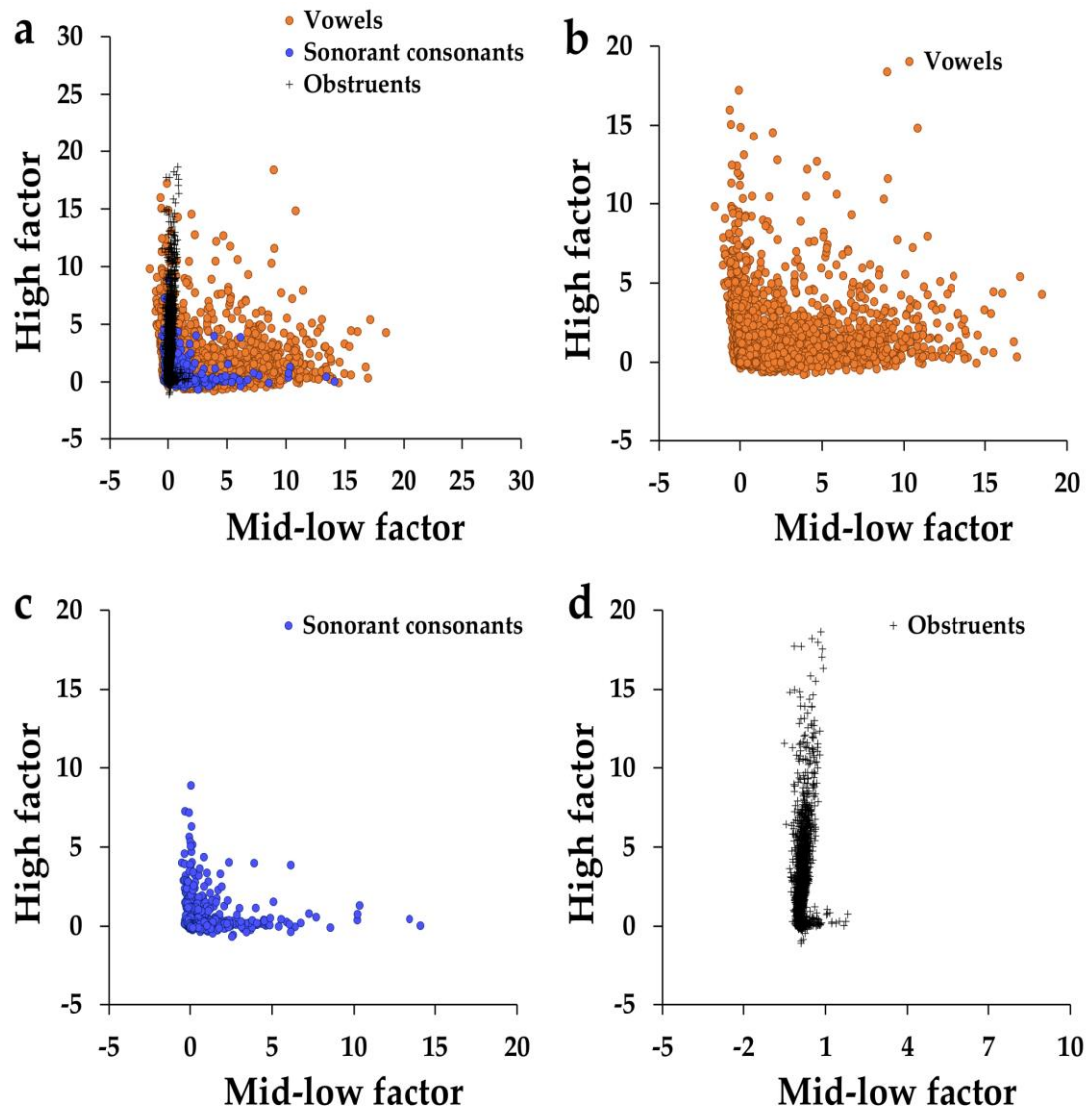


Appendix D. The expanded version of Figure 4.3 (a)-(e)

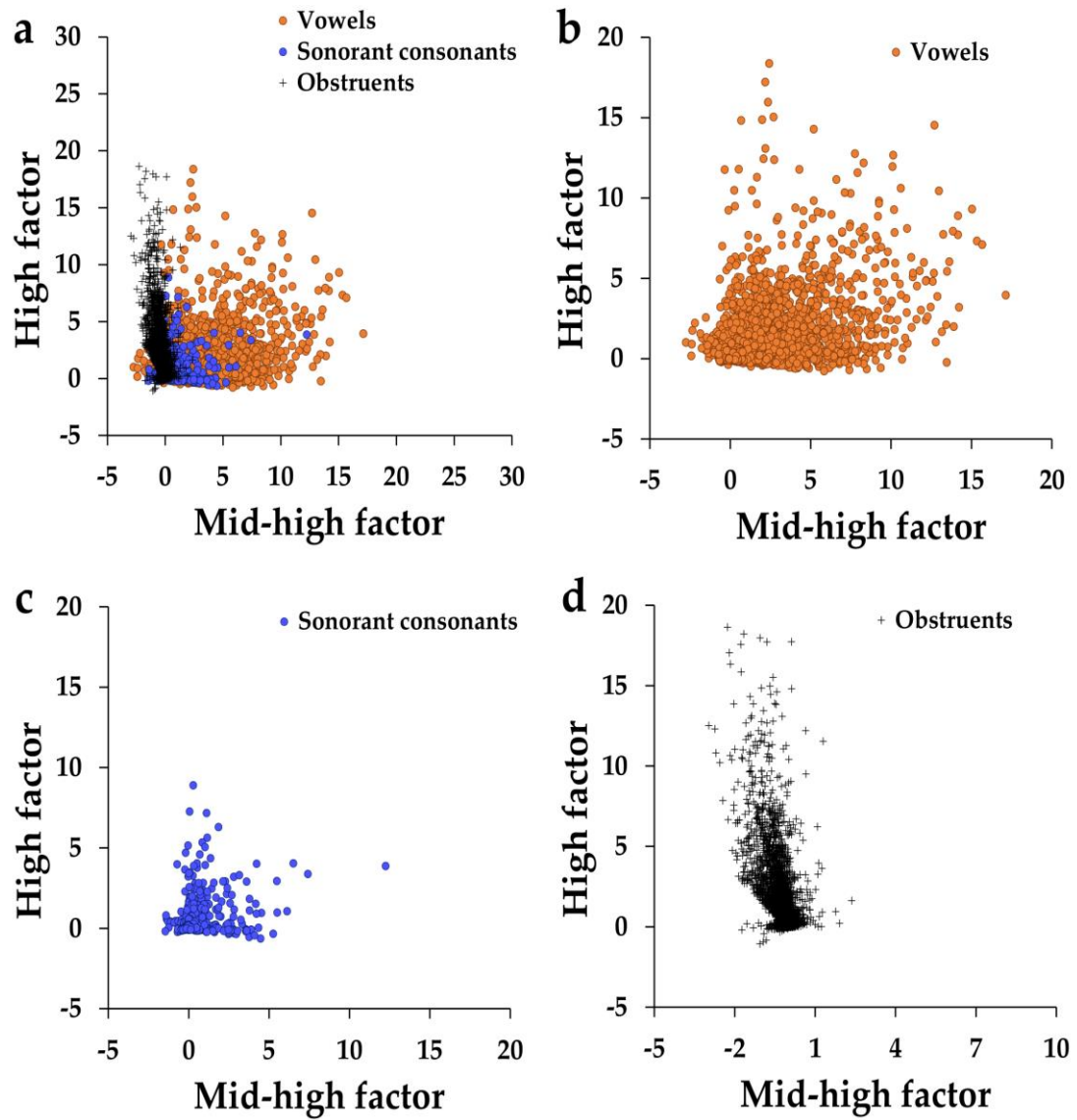
The distributions of the factor scores of vowels, sonorant consonants and obstruents in the two-dimensional space of the “High factor” and the “Low factor”.



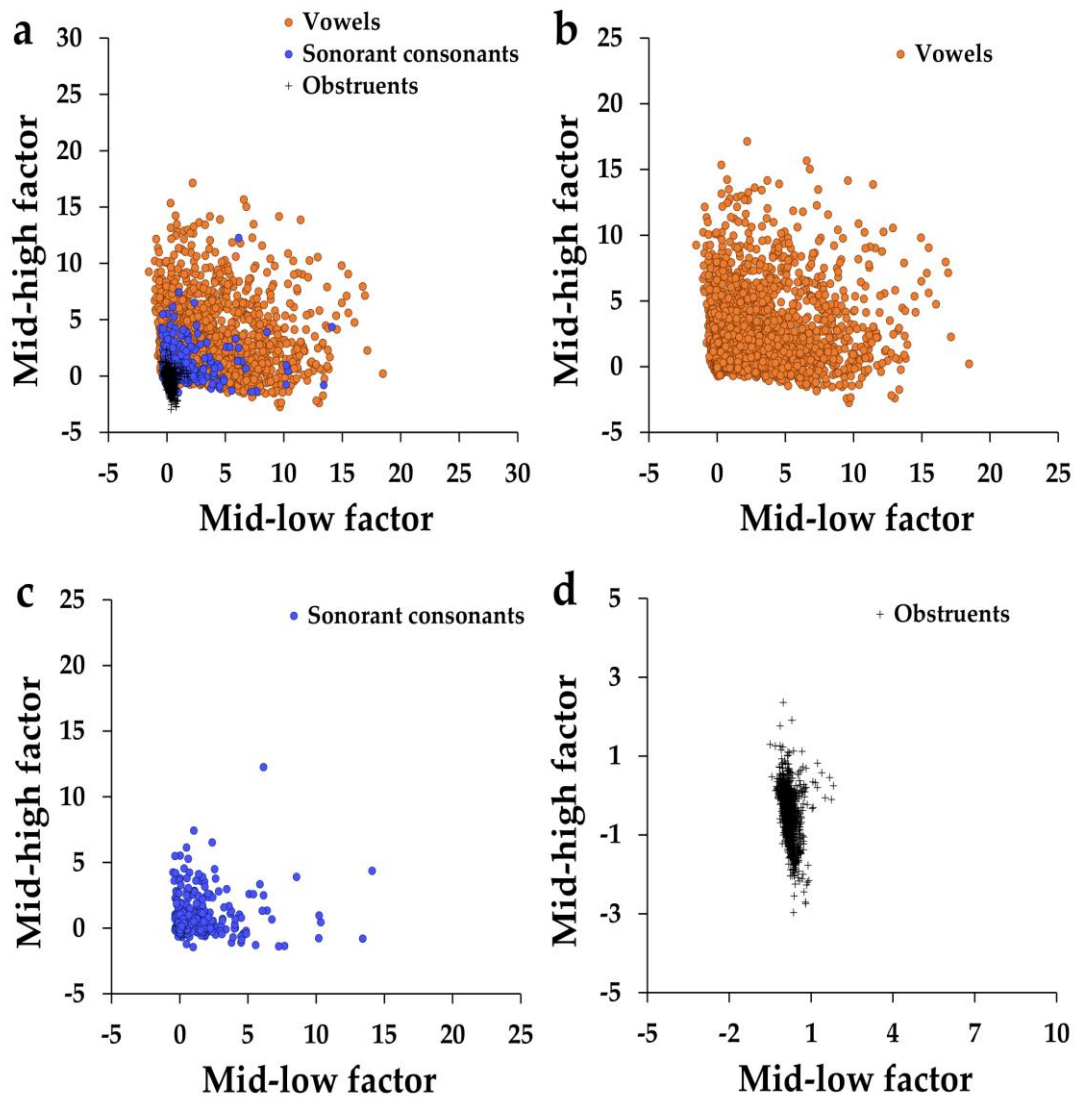
The distributions of the factor scores of vowels, sonorant consonants and obstruents in the two-dimensional space of the “High factor” and the “Mid-low factor”.



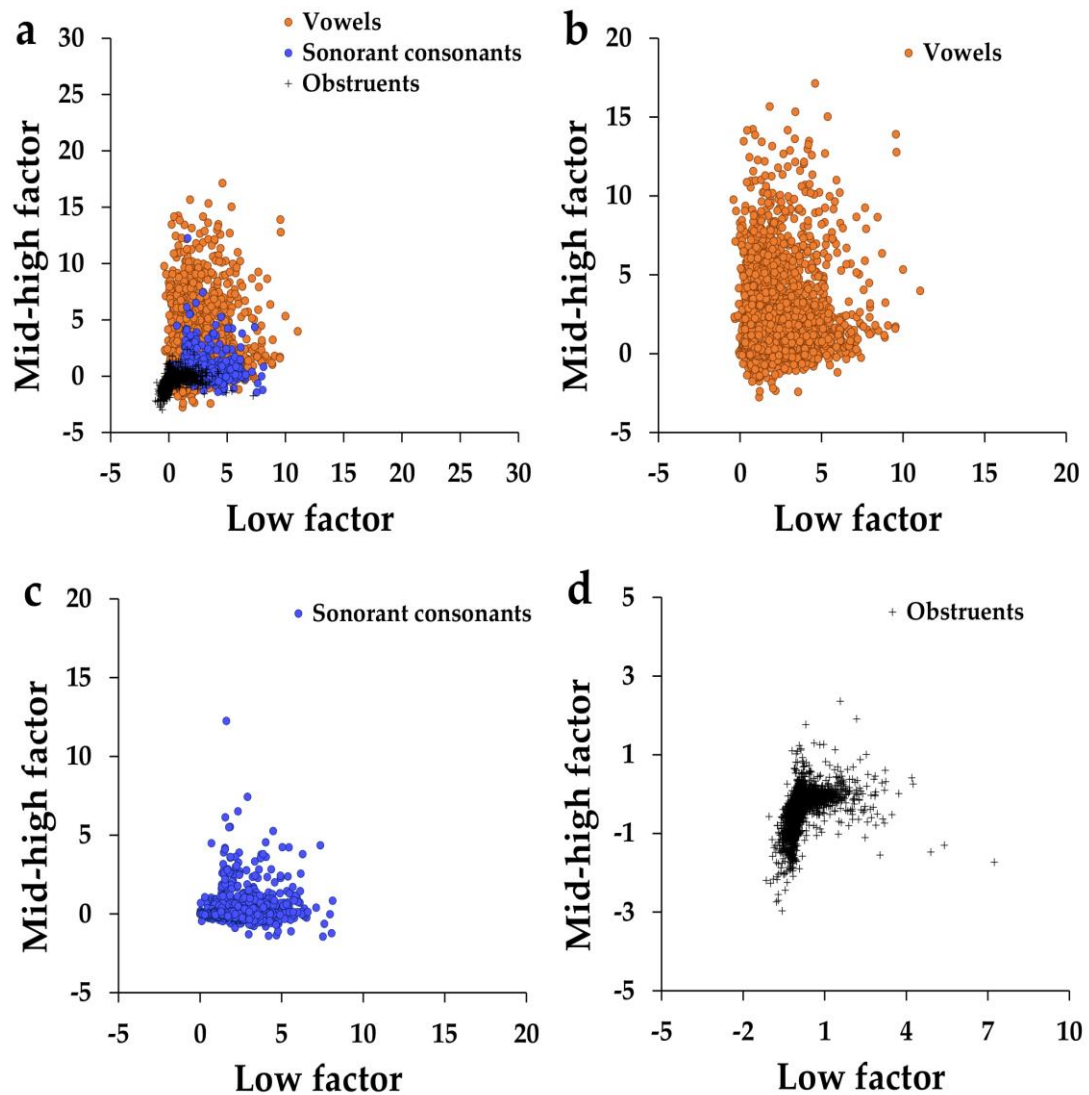
The distributions of the factor scores of vowels, sonorant consonants and obstruents in the two-dimensional space of the “High factor” and the “Mid-high factor”.



The distributions of the factor scores of vowels, sonorant consonants and obstruents in the two-dimensional space of the “Mid-high factor” and the “Mid-low factor”.



The distributions of the factor scores of vowels, sonorant consonants and obstruents in the two-dimensional space of the “Mid-high factor” and the “Low factor”.



The distributions of the factor scores of vowels, sonorant consonants and obstruents in the two-dimensional space of the “Mid-low factor” and the “Low factor”.

