How the Acoustic Correlates of English Obstruents Appear in Multivariate Analysis

Yixin Zhang, Kyushu University, Japan
Yoshitaka Nakajima, Sound Corporation, Japan
Kazuo Ueda, Kyushu University, Japan
Gerard B. Remijn, Kyushu University, Japan

The Asian Conference on Language 2021
Official Conference Proceedings

Abstract
Multivariate analysis of the acoustic characteristics of speech can provide insight into English phonology. To identify the acoustic correlates of obstruents, we here performed an origin-shifted factor analysis of English speech samples. Our multivariate analysis of spectral changes of speech demonstrated that one extracted factor, the so-called "mid-low factor" with high factor loadings around 1100 Hz, was strongly associated with vowels and sonorant consonants. Obstruents never went into the positive direction on this spectral factor. By contrast, two other spectral factors showed some association with obstruents, i.e., a "high factor" with high factor loadings above ~3300 Hz, and a "low & mid-high factor", with high factor loadings around 300 Hz and 2300 Hz. We focused on two general categories of English obstruents: fricatives/affricates and plosives in more detail. The results showed that six fricatives/affricates (/s/, /z/, /ʃ/, /ʒ/, /tʃ/, /dʒ/) occupied the positive direction only on the "high factor", while five other fricatives/affricates (/θ/, /ð/, /f/, /v/, /h/) occupied the positive direction on both the "high factor" and the "low & mid-high factor". All plosives (/p/, /t/, /k/, /b/, /d/, /g/) also occupied the positive direction on both these two spectral factors, but seemed mainly distributed on the "high factor". Overall, if we assume that the "mid-low factor" is correlated with what is called sonority, our factor analysis provides corroborating evidence for a widely-accepted idea in phonology: In the sonority hierarchy, obstruents are given the lowest rank, with plosives located at the bottom, and they do not constitute syllable nuclei.

Keywords: Speech Perception, English Phonemes, Factor Analysis, Sonority

Introduction

When we pronounce or perceive English speech, we prefer to combine several phonemes into a combination of sounds called a syllable. Syllables play a prominent role in theories that have proved useful in describing many prosodic phonological features of different languages [1]. In its common typology, a syllable is divided into three parts: the onset (or beginning), the nucleus (or peak), and the coda (or end). The nucleus is the most obligatory part in a syllable, and a syllable may even consist only of a nucleus, e.g., the first syllable in the disyllable word about. The nucleus is typically a vowel, while the onset of the coda is often made up of consonants [2]. This reflects a principle that is used to classify English phonemes: the sonority hierarchy. A high position in the sonority hierarchy represents a close proximity to the syllable nucleus. Vowels occupy the highest position in the sonority hierarchy, followed by sonorant consonants (glides, liquids, nasals). Obstruents (fricatives/affricates, plosives), formed by obstructing airflow, occupy the lowest position in the sonority hierarchy [3]. Thus obstruents are always considered as the syllable boundaries, while vowels always have a position on the syllable nucleus.

It has been argued that multivariate analysis of acoustic characteristics of English speech sounds can be related to phonological features. In one of the most recent studies using the classic factor analysis, Ueda and Nakajima (2017) found that multivariate analysis of speech samples from eight languages/dialects universally showed three spectral factors that appeared in four frequency bands [4]. Nakajima et al. (2017) further showed that the distributions of factor scores of English vowels, sonorant consonants, and obstruents, respectively, went from high to low on a spectral factor with high factor loadings on a frequency range around 1100 Hz. This tendency was highly related to the sonority hierarchy [5]. Furthermore, in our recent study, we have shown that the distribution of obstruents in English speech is related to two spectral factors: one is a bimodal factor with high factor loadings around 300 Hz and around 2300 Hz, and the other is a factor with high loadings on a frequency range around 4100 Hz [6]. In these studies, however, an English speech database was used in which the labeling of some phonemes was incorrect, and for this reason samples were omitted. To further investigate how acoustic characteristics of English and phonology are related, a new English speech database was created. In the present study, we performed a newly-developed type of factor analysis over speech samples from this database to confirm the previous findings, and to identify the acoustic correlates of English obstruents in more detail.

Methods

Speech Samples

First, a new English database was created, which contained 100 sentences uttered by one male and two female native-English speakers. The speech samples in the database were recorded with a sampling frequency of 44100 Hz and 16-bit linear quantization. The speech signals of all the spoken sentences ..... segmented into individual phonemes and were labeled utilizing the International Phonetic Alphabet, with the Cambridge Advanced Learner's Dictionary as pronunciation reference [7]. A total of 11935 English phonemes was used as analysis samples.

The Origin-shifted Factor Analysis

The speech samples were analyzed with the origin-shifted factor analysis, that has been recently developed [8]. This method is potentially suitable to resynthesize speech, because every starting point of the data is moved from the gravity center to the origin, so that all silent parts in speech signals would remain silent. In our recent study [6], we compared a classic factor analysis and the origin-shifted factor analysis and focused on the two main differences between these multivariate analysis methods: the origin shift and cepstral liftering. Cepstral liftering is a smooth processing of the signal and applied to the analysis of speech [9]. We argued that multivariate analysis with cepstral liftering would not relate to the real auditory signal quality, since it will smooth the speech signals and may weaken the features of speech. It was concluded that the origin-shifted factor analysis without cepstral liftering is more recommendable for speech analyses, and we therefore used it in the present analysis.

The two-dimensional axes derived from the origin-shifted factor analysis were rotated by varimax rotation [10]. As a result, similar to our previous study [6], three main spectral factors were extracted and their factor loadings were obtained. All analysis samples of English phonemes were divided into three categories: vowels, sonorant consonants, and obstruents. In total, 4528 vowels, 2467 sonorant consonants, and 4940 obstruents were used. The factor scores of the central midpoints of time for all the labeled phoneme samples were calculated. To observe the acoustic correlates of obstruents in more detail, the distributions of all English phonemes were analyzed. We focused on the distributions of obstruents in particular (4940 in total), as represented in the two-dimensional factor space.

Results

Factor Loadings of the Three Spectral Factors

Figure 1 shows the factor loadings obtained by the origin-shifted factor analysis for all the English speech samples spoken by three native speakers. 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 1. The first band was a low-frequency band, from about 50 to 600 Hz. The second band was a mid-low frequency band, around 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” (Figure 1, red line) was bimodal, 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 our previous study [6]. The second factor, the “mid-low factor”, was related to the frequency range around 1100 Hz (Figure 1, black line). The third factor, the “high factor”, was related to the frequency range around 4100 Hz (Figure 1, blue line).

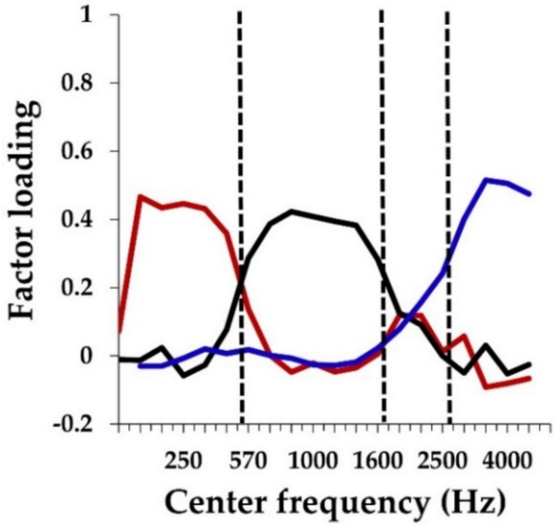

Figure 1: Factor Loadings of the Three Extracted Spectral Factors of 100 English Speech Samples from Three Native Speakers with the Origin-Shifted Factor Analysis.

Factor Scores of the Three Spectral Factors

The factor scores of all English phonemes were obtained and distributed into three categorical areas in the factor space, i.e., vowels, sonorant consonants, and obstruents. On the "mid-low factor" with a center frequency range around 1100 Hz, the highest factor scores were obtained by vowels, followed by sonorant consonants, and obstruents. 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", similar to the result in our earlier research [6].

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 2 shows the distributions of the factor scores of individual obstruents divided into eleven fricatives/affricates in Figure 2(a), and six plosives in Figure 2(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".

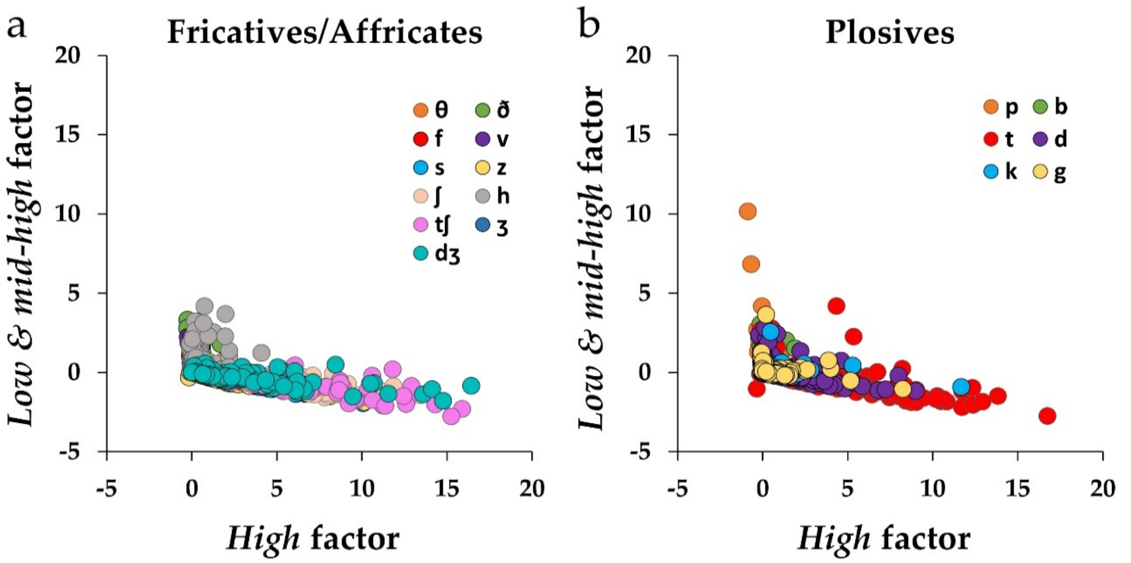

Figure 2: The Distributions of the Factor Scores of Obstruents Divided into Fricatives/Affricates (**A**) and Plosives (**B**) in the Two-Dimensional Space of the "Low & Mid-High Factor" and the "High Factor".

We further observed the distributions of the average factor scores of the fricatives/affricates in Figure 3(a) and of the plosives in Figure 3(b), in the two-dimensional space of the "low & mid-high factor" and the "high factor". In Figure 3(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(b), all six plosives (/p/, /t/, /k/, /b/, /d/, /g/) were close to the origin and on both the "high factor" and the "low & mid-high factor", but they were mainly distributed on the "high factor".

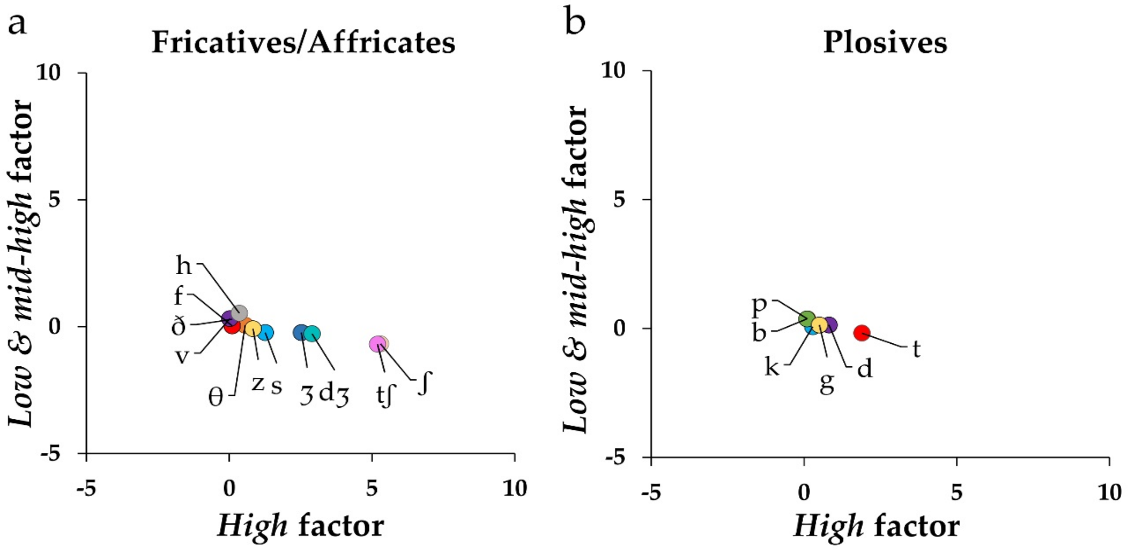

Figure 3: The Distributions of the Average of the Factor Scores of Obstruents Divided into Fricatives/Affricates (**A**) and Plosives (**B**) in the Two-Dimensional Space of the "Low & Mid-High Factor" and the "High Factor".

Conclusions

In the present study, the origin-shifted factor analysis was applied to English speech samples. The samples were taken from a newly-recorded database that contained 100 sentences each spoken by three native speakers of English. In total, 11935 individually-labeled English phonemes were used as analysis samples. The analysis showed that the distributions of the obtained factor scores in the factor space well reflected the phonological roles that English phonemes are considered to have.

Confirming earlier research [5, 6], our analysis proved again that the "mid-low factor", with high factor loadings on a frequency range around 1100 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". Always playing the roles of syllable nuclei in English, vowels are high in the sonority hierarchy. Sonorant consonants occupied the middle position on the "mid-low factor" 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 in the word *little* is the vowel /i/, but the nucleus of the second syllable is the sonorant consonant /l/ [2]. 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. Given these findings, we can therefore call the "mid-low factor" the "sonority factor", confirming previous results [5, 6].

Secondly, we investigated the acoustic natures of English obstruents in the other two spectral factors as also extracted in our early studies [6], i.e., the "low & mid-high factor" with bimodal frequency ranges around 300 Hz and around 2300 Hz, and the "high factor" with a frequency range around 4100 Hz (Figure 2, Figure 3). In an English syllable, obstruents typically occupy a low position on the sonority hierarchy. This was clearly reflected in our analysis: obstruents never went into the positive direction of the "mid-low factor" - 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 confirmed that obstruents do not constitute the syllable nucleus.

With the present multivariate analysis, we thus connected acoustics and phonology, by extracting spectral factors that seem to represent sonority features of English speech sounds. To further identify the acoustic features of obstruents, the newly-recorded English database will be used for more analyses about the categorical perception of English phonemes.

Acknowledgments

This study was supported by JSPS KAKENHI Grant Number 19H00630. Y.Z. is supported by a MEXT Super Global University Project scholarship from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

References

Féry, C., Vijver, R. (2003). *The syllable in optimality theory*. Cambridge University Press: Cambridge, UK.

Gordon, K. M. (2016). *Phonological typology*. Oxford University Press, UK.

Spencer, A. (1996). *Phonology: Theory and Description*. Blackwell Press: Oxford, UK.

Ueda, K., Nakajima, Y. (2017). An acoustic key to eight languages/dialects: Factor analyses of critical-band-filtered speech. *Scientific Reports*, 7, 42468.

Nakajima, Y., Ueda, K., Fujimaru, S., Motomura, H., Ohsaka, Y. (2017). English phonology and an acoustic language universal. *Scientific Reports*, 7, 46049.

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.

McIntosh, C. (2013). Cambridge Advanced Learner's Dictionary, 4th Edition.

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.

Awan, N.S., Giovinco, A., Owens, J. (2012). Effects of vocal intensity and vowel type on cepstral analysis of voice. *Journal of Voice*, 26, 5, 670.e15–670.e20.

Kaiser, H.F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187–200.

Contact email: 3DS20005M@s.kyushu-u.ac.jp
remijn@design.kyushu-u.ac.jp