Visualization for Learning Foreign Speech

Kaoru TOMITA
(Faculty of Literature and Social Sciences)

Abstract

This paper examines the articulatory properties of six vowels, /i/, /ɪ/, /æ/, /a/, /ʊ/ and /u/ in English produced by twenty-two Japanese-speaking English learners with an acoustic analysis of spoken words measured by Praat. Formant 1 and 2 of these vowels are compared with the ones produced by native speakers of English. Feedbacks from learners about a method of pinpointing each vowel in vowel space and comparing them with the ones by a native speaker are collected and estimated.

Keywords

feedback, language, pronunciation, visualization, vowel

1 Introduction

Language generally can be divided into two main forms. They are spoken and written ones. It is said that the former came to our world first and then the latter has developed from it. We say that languages look different, in a case where their structures, spellings or sounds are different. They, however, really have a strong connection with each other. Ngugi (1986, 14) claims that the written word imitates the spoken one.

As is stated in Schiff (2013, 409), children who acquire the ability to read must first learn the visual code used in their culture of representing speech as a series of symbols. As such, learning to read is ultimately a matching process in which unique visual symbols are matched to units of sound, with the relationship between symbols and sounds being systematic in many languages and are acquired with relative ease. Trial to visualize speech has been found in old history of linguistics. Alexander Melvil Bell invented Visible Speech in 1867 named by himself, which transcribes sounds into wave forms (Coulmas, 2014, 34).

To pursue an intimate relationship between symbols and sounds in a linguistic issue, this study introduces concepts, based on which an original research is conducted. From section two to five, connections between visual symbols and sound units found in sign language,

human senses, phonetic features and trainings of pronunciations are explained. Results of the research in section six lead to a conclusion in the last section, where a trial to visualize speech is promoted as it works for learning foreign languages.

2 Sign language

Main types of visualized methods for communication used in society are sign languages. Not everybody use them in our life but they are regarded as authorized ways to communicate with each other. As visualized tools for communication, sign languages might parallel written languages. Just like written languages, sign languages have letters. They, however, do not have systems for forming words and phrases. For sign language users, concepts of motions, such as *to eat, to go, to see,* are to be formed with combinations of hand and finger movements. As is described in Crystal (2007, 159), a few of the signs in any systems are indeed iconic and the vast majority of signs are arbitrary, just like the words of spoken languages.

For those who do not use sign languages in their life, it looks like they contain too complicated and delicate moving of hands, fingers and faces. Sign languages have nature and function in themselves (Sze, 2012). They are composed of wonderful systems which have features, such as synchronism, persistence of vision, perspective, comic storytelling, shape of mouth, image clarity, and montage (Sakata, et al., 2008, 110). These types of variations are not to be found in verbal languages.

As Berent (2013, 12) claims, signed and spoken phonologies share many structural characteristics. Hand shapes and arm movements that are obligatory aspects of sign languages themselves hold phonological characteristics. They may act just like tongue shapes and mouth's openings or closings.

Spoken languages have pauses in utterances, and also sign languages have holds for showing several functions and meanings. As is stated in Groeber (2012, 133), holds have been shown to be a powerful resource in social interaction that participants draw on not only to project a next action to take place, but also to display on-line their understanding of its relevant accomplishment.

3 Human senses

We process information around us with making full use of human senses. When we speak with others, we hear what they say with looking at expressions on their faces, moving lips,

and gestures. This happens when we watch moving pictures on screens. We usually listen to some explanations recorded on tracks of the same media materials. Each human sense becomes supportive for the other senses to capture information. Visual information can support audio perception in speech and sound information would be supportive for visual perception in films. Shigeno (2014, 161) claims two ways to this visual or perceptual information; color hearing or tone seeing. In general, the more noise alters auditory speech perception, the more visual information is used (So, C. K. et, al., 2014, 614, quoted from MacLeod, A., et al. 1987).

In natural environments, these human senses integrate very well for perception of information. In human made environments, such as virtual worlds and social nets, visuals work very well and several types of visual information help users to get to know that even something unusual is happening in the virtual worlds. As is described in Jones (1996, 105), graphics, texts, tables and animations exploit only the human vision system. Beside the vision, as he claims, human perception relies on four other senses for processing complex information, and at least some of these senses should be exploited all the time. Changizi (2011) claims visual ambiguity can be reduced by auditory information, and vice versa. Furthermore, he points out that there are regions of cortex responsible for making vision and audition fit one another.

Multi-sense is explained in Shams (2011, 264) with its processes: First, multi sensory experiences quickly recalibrate unisensory maps in the brain. Second, a new connection between unisensory cortical areas in the brain is created. Third, unisensory representation of stimuli is integrated with those stimuli in a multi sensory manner. As is well known, visual, perceptual and body senses are associated in left side of brain and that makes language processing, such as reading and writing (Nakagome, K., 2010, 73).

When we read English passages, we automatically assimilate sound features elicited from letters. On the basis of an experimental study, Lee (2013, 191) suggests that phoneme-to-phoneme transformations involved in uttering a word may also be involved in identifying the word visually.

4 Phonetic features

Languages are classified into several types from viewpoints of phonological features, such as types of closed or open syllables, quality of vowels, that of consonants and their alignments. English is placed on closed syllables and Japanese is placed on open syllables. However, as is

claimed in Granlund, et al. (2012, 510), there is conflicting evidence as to whether both global and segmental features are language-universal or language-specific.

Japanese has a smaller vowel inventory of 5 monophthongs instead of 11 in English and the former has short or long vowel contrasts which differ almost exclusively in duration, whereas the latter has tense-lax contrasts which differ primarily in vowel spectrum and, less importantly, in duration. For example, the high front vowel minimal pair /i-ɪ/ in English and the long-short distinction /i-i:/ in Japanese are typical differences.

Links between sound and meaning have been main interests to researchers of linguistics, psychology and sociology. Various studies have established a robust existence of sound symbolism, the phenomenon in which speakers link phonetic features with meanings in a non-arbitrary fashion (D'Onofrio, 2014, 367). As is claimed in Feist (2013, 116), sound "symbolism" in the wide sense sometimes serves the expressive function, either alone or in combination with the communicative function and in other uses, and it characteristically serves the dramatising function, as well as the communicative one, as do such other elements of English as climactic syntactic structure, exclamatory phonology and emotive wording. Language sounds play an important role to convey information in many types of speech style. Brown (2014, 45) states that politeness does not merely reside in verbal markers but is co-signaled by phonetic cues.

5 Learning pronunciation

Great importance is put on visual information for identifying words. As is stated in Lidestam (2014), only audiovisual training improves speech-in-noise identification, demonstrating superiority over auditory-only training. This type of attention to audiovisual training is to be found for improving listening skills but not for speaking skills, nor for improving pronunciation.

As is pointed out in Yamada (2014, 448), easy-to-use methods for presentation of learners' articulation are demanded. Ian (2014, 563) states that it is not easy to lead learners to change shape of their tongues without looking at them. He concludes that with showing them supersonic wave of their tongue shapes displayed on screen, they come to articulate speech sound according to teachers' direction. Adekunle (2014, 726) states that it is observable in data analysis that some foreign segments which are absent in native phonology are substituted with their closest alternative phonemes.

Technologies have produced ways to visualize human utterances. One of them is an x-ray

of vocal tracts (Wilson, et al., 2014, 106). Pictures of vocal tracts listed in pronunciation textbooks would help learners to find a relation between shapes of their tongues and sounds produced. Textbooks of foreign language pronunciations list, however, not real pictures but illustrations. Simplified illustration might be better for learners to grasp features of sound. In a way, pictures of their speaking organs are too lurid for them to learn their way of pronunciation (words by one of presenters at *General Meeting of the Phonetic Society of Japan*).

It needs to be considered very well before putting technique of visualization to learning pronunciations. Takei (2014, 21) states that there is an importance of knowing how to control his own body form and motion to be a good athlete. Appropriate motion comes from a good form and the good form is created through appropriate motion. To know form of his/her part of body by putting great thought to his/her own form is also important for learning pronunciations.

Putting technology of visualization to pronunciation training is, in a way, a mixture of an advanced technology and a traditional feat. Vowel space measured with sound analysis software in this study is drawn on a sheet of paper and checked by learners themselves. This way is reliable because it is measured with most advanced technologies. Validity of this way is high as what is measured and checked by learners is their own pronunciations. More than anything, this is practical from an economic view point. What are necessary for learners in a class are a sheet of paper and a piece of pencil. PC and software for measuring learners' pronunciation are not necessary for all of them. A teacher can measure the words produced by learners with some amount of time in the class.

The way to move muscles to utter some types of sounds with the aid of visualization should work very well. The same is stated in Osawa, et al. (1985, 198) for practicing writing letters, in which what part of muscles are moved should be always thought about because that would improve learning effects. Imitation might be a part of learning foreign language sounds. As is presented in Babel (2011, 177), participants accommodated toward vowels selectively; the low vowels /æ a/ showed the strongest effects of imitation compared to the vowels /i o u/.

In the current study, an experimental research is conducted to show that vowels, such as /i/ and /ɪ/ or /ʊ/ and /u/ discriminated by native speakers of English are not done so by all of Japanese learners of English. Feedbacks from learners about acoustical analyses of vowels and results dotted on vowel space are collected to propose that visualization of

sounds on two dimensions promote learners to know how to control their tongues for English pronunciations.

6 Research

6.1 Aims

There are two aims for the experiment. One is to present an arrangement of vowels in vowel space produced by Japanese learners of English. The other is to collect learners' feedbacks of a learning method in which visualization of vowel space is used for training pronunciation of vowels.

6.2 Method

Subjects are asked to read listed words that include six different vowels. These words are recorded and whose formant 1 and 2 are measured with Praat. Results are given to each subject. They put dots for each vowel in a sheet of paper on which vowel space is drawn. After that they put their comments for their own pronunciations.

6.2.1 Participants

Twelve female Japanese learners of English (mean age 19 years) and ten male Japanese learners of English (mean age 19 years) take part in the research. All are university students who are majoring in linguistics. They come from several different regions in Japan. They are brought up as monolinguals and have learned English as a second language at school for over six years. They are intermediate level English speakers, which are reflected in their self-reported English skills.

6.2.2 Materials

Six words, “heed”, “hid”, “had”, “hod”, “hood” and “hoodoo”, each of which include different vowels are selected.

6.2.3 Apparatus

PC (MacBook Air) and sound analysis software (Praat) are used for recording and analyzing vowels.

6.2.4 Procedure

Recording and analyzing of vowels are conducted for each subject respectively. Recording and analyzing are done by the author.

6.2.5 Measurements

Middle part of vowels in each word is selected and formant 1 and 2 are measured.

6.3 Results

Table 1 presents mean formant 1 and 2 of six vowels produced by 22 subjects.

Table 1 Mean formant 1 and 2 values of six vowels

Word	Formant 1	Formant 2
heed	399.64	2370.05
hid	491.32	2430.50
had	786.05	1513.55
hod	619.00	1094.82
hood	462.45	1399.86
hoodoo	445.09	1379.50
Mean	533.92	2133.13
F-value	16.95	56.91
p-value	< 0.01	< 0.01
Comparison	heed, hoodoo, hood, hid, hod< had	hod< hoodoo, hood, had < heed, hid

The degrees of freedom are all 5 and 126.

Formant one and two values represent spreading of six vowels in vowel space. Statistical analysis, however, shows that some of them are not discriminated very well. There are not significant differences between “heed” and “hid”, or “hoodoo” and “hood”.

Subjects’ feedbacks about their own pronunciations of six vowels that are dotted in vowel space are collected. They all present positive attitudes to analyzing and learning pronunciations. They are listed in Appendix.

7 Conclusion

Raw data of learners’ speech or oral reading of listed words and their analyses must be an important issue for study of language learning. Database of learners’ speech have been constructed at a large scale and usage of these data must be useful. For some areas of study, such as archeology and history, accumulation of data is increasing too much and field work and collection of data are not considered to be important and useful anymore (Kobayashi, 2014, 1). For study of language learning, however, not many collections of data have been conducted at a large scale and an individual researcher is, in a way, free to collect any types of speech for his/her aim of studies.

Combination of visual and perceptual information has been promoted by development of technologies. Digital museums, for example, are now taking roles of backing up this type of artifacts. They make resources of multi-media in a wide area including perceptions and touching things that are open to public inspection, and that means digital museums are

surpassing old style museums that have been displaying only visual information (Nishino, 1996, 288).

Studies of perceptual learning focused on training with making use of one sensory modality fails to tap into natural learning mechanisms that have evolved to optimize behavior in a multisensory environment (Shams, et al. ibid., 7). Visualization of vowel space for language learning is sure to be made use of to let learners have a special interest in their pronunciations.

A concept of effectiveness of still picture is made use of for this study. The author thinks that still pictures work better than moving pictures for learning foreign language pronunciation. Besides, two-dimension pictures are easy to understand than three-dimension pictures in some cases. This issue should be proved well with objective measurements in future studies. Current research on visual communication suggests that still images can convey complex conceptual structures like categorization, analogy, causality and even temporal intervals (Oversteegen, et al. 2014, 93).

Most feedbacks suggest positive attitudes to visualization of foreign language pronunciations. This might be because learners prefer to a primitive way of sensing information for learning a foreign language. Small children prefer to use visual information for sensing things, such as size of the things. As is presented in Tribushinina (2013, 205), for an experiment conducted for 2 to 7-year-old children, the results demonstrate that there is a gradual increase in the ability to inhibit visual cues and to use world knowledge for interpreting size terms.

There are pros and cons for the argument that some phonetic features of native language are assimilated into foreign language ones. One of the proponents is So, et al. (ibid., 611), who point out that Cantonese might have assimilated their vowels to their closest native vowels. One of those who is against that is Darcy (2012, 568), in which discrimination task provides evidence that children who are native speakers of Turkish and begin learning German as an L2 in kindergarten categorize difficult German contrasts differently from age matched native speakers.

Questionnaires are used in this research and the author thinks their result reflect learners' thinking and feeling to the method for promoting natural pronunciations. Objective ways of measurements, such as amount of time that learners engaged in learning pronunciation with the method treated in this study and change or improvements of their pronunciation measured in acoustic features, such as formants and durations, are better to

be employed in future studies.

Acknowledgements

Many thanks are due to colleagues, friends, loved ones. My students are among my best teachers. I wish to thank all of them.

Funding

This research is supported by a Project Grant-In Aid for Scientific Research by the Ministry of Education, Culture, Sports, Science and Technology (Basis C-26370655, "Applied study on ability of analyzing English sound with visualized vowel spaces").

References

Adekunle, O.G. (2014). Devient realization of foreign vowels in the speech form of Yoruba-English Nigerian bilinguals. *Open Journal of Modern Linguistics*, 4, 720-727.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40, 177-189.

Berent, I. (2013). *The Phonological Mind*. Cambridge: Cambridge University Press.

Brown, L., Winter, B., Idemaru, K. and Grawunder, S. (2014). Phonetics and politeness: Perceiving Korean honorific and non-honorific speech through phonetic cues. *Journal of Pragmatics*, 66, 45-60.

Changizi, M. (2011). *Harnessed: How Language and Music Mimicked Nature and Transformed Ape to Man*. In audiobook. Dallas: BenBella Books, Inc.

Coulmas, F. (2014). *Moji no Gengogaku (Writing Systems – An introduction to their linguistic analysis)*. (Translated by Saito, S.). Tokyo: Taishukan.

Crystal, D. (2007). *How Language Works*. London: Penguin Books.

Darcy, I. and Kruger, F. (2012). Vowel perception and product in Turkish children acquiring L2 German. *Journal of Phonetics*, 40, 568-581.

D'Onofrio, A. (2014). Phonetic detail and dimensionality in sound-shape correspondences: Refining the *Bouba-Kiki* paradigm. *Language and Speech*, 57:3, 367-393.

Feist, J. (2013). "Sound symbolism" in English. *Journal of Pragmatics*, 45, 104-118.

Granlund, S., Hazan, V. and Baker, R. (2012). An acoustic-phonetic comparison of the clear speaking of Finnish-English late bilinguals. *Journal of Phonetics*, 40, 509-520.

Groeber, S. and Pochon-Berger, E. (2012). Turns and turn-taking in sign language interaction:

A study of turn-final holds. *Journal of Pragmatics*, 65, 121-136.

Ian, W. (2014). Choonpa wo mochiita choon no shido to kenkyu (Study of articulation with supersonic wave). *The Journal of the Acoustical Society of Japan*, 70:10, 560-564.

Jones, C. V. (1996). *Visualization and Optimization*. London: Kluwer Academic Publishers.

Kobayashi, K. (2014). Kokogakukyouiku he no Jyoumonsyuraku detabesu no riyou (Usage of database of Jomon colony for archeology education). *Resource Sharing Newsletter for the Humanities*, 8, 1-12.

Lee, Y., Moreno, M. A., Carello, C. and Turvey, M. T. (2013). Do phonological constraints on the spoken word affect visual lexical decision? *Journal of Psycholinguistic Research*, 42, 191-204.

Lidestam, B., Moradi, S., Pettersson, R. and Ricklefts, T. (2014). Audiovisual training is better than audio-only training for auditory-only speech-in-noise identification. *Journal of Acoustical Society of America*, 136:2, published online.

MacLeod, A. and Summerfield, A. W. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, 21, 131-141.

Nakagome, K. (2010). *Gengo to Igaku (Language and Physiology)*. Tokyo: Asakura Books.

Nishino, Y. (1996). *Rekishi no Moji (Letters in History)*. Tokyo: Tokyo University Press.

Ngugi, W. T. (1986). *Decolonising the Mind: The Politics of Language in African Literature*. Oxford: James Curry.

Osawa, K., Kuwayama, S., Kurauchi, H., Yajima, K., Yoshimura, M. and Yamada, H. (1985). *Moji no Kagaku (Science of Letters)*. Tokyo: Hosei University Press.

Oversteegen, E. Schilperoord, J. (2014). Can pictures say no or not? Negation and denial in the visual mode. *Journal of Pragmatics*, 67, 89-106.

Sakata, K., Yano, K., and Yoneuchiyama, A. (2008). *Odorokino Shuwa "Pa" "Po" Honyaku (Amazing Sign Language Translation with "Pa" and "Po")*. Osaka: Seikosha.

Schiff, R. (2013). Shallow and deep orthographies in Hebrew: The role of vowelization in reading development for unvowelized scripts. *Journal of Psycholinguistic Research*, 41, 409-424.

Shams, L., Wozny, D.R., Kim, R., and Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Front. Psychol*, 2, 264.

Shigeno, S. (2014). *Otono Sekai no Shinligaku (The Psychology of the World of Sound)*. Kyoto: Nakanishi Publishing.

So, C. K. and Attina, V. (2014). Cross-language perception of Cantonese vowels spoken by

native and non-native speakers. *Journal of Psycholinguistic Research*, 43:611-630.

Sze, F. Y. B. (2012). Right dislocated pronominals in Hong Kong sign language. *Journal of Pragmatics*, 44, 1949-1965.

Takei, S. (2014). Attaka taidan 47 (Hot car interview no. 47). *JAF Mate*, 52:10, 20-22.

Tribushinina, E. (2013). Adjective semantics, world knowledge and visual context: Comprehension of size terms by 2- to 7-year-old Dutch-speaking children. *Journal of Psycholinguistic Research*, 42, 205-225.

Wilson, I. and Kanada, S. (2014). Pre-speech postures of second-language versus first-language speakers. *Journal of the Phonetic Society of Japan*, 18:2, 106-109.

Yamada, R. (2004). Gaikokugogakusyu no tameno ICT kyozaikaihatsu no choryu (Trends for development of information and communication technology for foreign language learning). *The Journal of the Acoustical Society of Japan*, 70:8, 446-451.

Appendix

Feedback from learners.

1. I could not clarify differences of vowels in “hod”, “hood” and “hoodoo”. I thought that was because I was nervous and could not stabilize my own pronunciations. I thought that it was difficult for the Japanese to pronounce very good English.
2. I could not clarify differences of vowels in “heed” and “hid”, or “hood” and “hoodoo”. My “hood” was outside of vowel space drawn on the paper. It was interesting to quantify my own pronunciations.
3. An arrangement of my vowels spread very tight. Formant 2 values were around 1000s and that made me find that I did not put my tongue to the front of the mouth very well.
4. Positions of six vowels did not scatter as much as I expected. I thought I did not discriminate two vowels that were arranged close in vowel space. I thought my pronunciations were not clear as I thought by myself.
5. I tried to close my mouth for uttering the word “heed” and open a little for the word “hid”. The result was, however, vice versa. I closed my mouth to pronounce the word “hid”. This kind of things usually did not come to me so this was a very interesting activity.
6. I thought I always tightly closed or widely opened my mouth when I was speaking in English. With looking at the figures, however, I found I could not do that as I expected. I thought those who spoke good English would open or close mouth and moved tongue

more accurately than I thought they were doing.

7. I could not put as much differences as I expected. I did not speak with opening my mouth and that became my own style. So-called good English was not mine. Now I have found the reason why I could not pronounce very clearly.
8. This was the first time for me to analyze my own vowel pronunciation. I felt a little nervous. The result showed that vowels in “heed” and “hid” were discriminated very well. I felt good. Vowels in “hood” and “hoodoo”, however, did not show much difference. I felt sorry for that. I thought vowel pronunciations revealed individuals’ characteristics.
9. It was very difficult to clarify differences of vowels that did not really differ a lot. I thought I could only discriminate different sounds halfway. I made up my mind to pay my attention to vowel sounds and pronounce them clearly from then on.
10. Formant 1 and 2 varied and to speak good English, I needed to do practice for pronunciation. Among six words, two were pronounced so so, but the rests were so terrible.
11. Six vowels almost gathered together. This was why my speaking could not be heard very well. I thought I would pay my attention to the shape of my mouth from then on.
12. Among six vowels, the three that were produced at the front of my mouth were pronounced very well. The other three that were produced at the back of my mouth were not pronounced very well.
13. Results made me feel that I should speak more clearly. This was a very good experience for me.
14. I could not estimate whether my pronunciation was good or not just with looking at the results. Anyway I thought it would be better to pay my attention to the shape of my mouth.
15. I paid my attention to the shape of my tongue but I could not move it into a proper position. Results of measurement surprised me a lot. An arrangement of six vowels was just messy. I have now found the reason why I could not communicate with one of my German friends. I always communicated not by speaking but by showing pictures.
16. Six vowels gathered together and they were not discriminated very well. I paid my best attention to the shape of my tongue but I could not pronounce them very clearly. I came to think how I could change my own pronunciations.
17. Vowels were not discriminated as much as I expected. Especially formant 1 values showed a big gap between the one by the model speaker and mine.

18. It was much more difficult than I expected to clarify differences in each vowel. I have found I could not pronounce vowels in front position of the mouth that was opened widely. I thought I would think about that from then on.
19. I was nervous and I could not discriminate vowels very well. Most of the vowels were produced at the back of my mouth and also I could not close my mouth in a good way.
20. All the vowels gathered together and that seemed to visualize my tendency for not opening my mouth when I spoke in Japanese.
21. I thought I could not clarify my pronunciations.
22. Pronunciations of these vowels were too difficult for me who did not understand English at all.

Visualization for Learning Foreign Speech

Kaoru TOMITA
(Faculty of Literature and Social Sciences)

This paper examines the articulatory properties of six vowels, /i/, /ɪ/, /æ/, /a/, /ʊ/ and /u/ in English produced by twenty-one Japanese-speaking English learners with an acoustic analysis of spoken words measured by Praat. Formant 1 and 2 of these vowels are compared with the ones produced by native speakers of English. Feedbacks from learners about methods of pinpointing each vowel in vowel space and comparing them with the ones by a native speaker are collected and estimated.