Analytical Visualization for Learning Foreign Language Sounds

TOMITA Kaoru
(Faculty of Literature and Social Sciences)

山形大学紀要（人文科学）第18巻第 4 号別刷
平成29年（2017） 2 月

Analytical Visualization for Learning Foreign Language Sounds

TOMITA Kaoru
(Faculty of Literature and Social Sciences)

Abstract

This study explores how visual information helps us to learn foreign language pronunciation. Visual assistance and its effect for learning foreign language have been discussed widely. For example, simplified illustrations in textbooks are used for telling learners which part of the articulation organs are used for pronouncing sounds. Vowels are put into a chart that depicts a vowel space. Consonants are put into a table that contains two axes of place and manner of articulation.

When comparing a still picture and a moving picture for visualizing learners' pronunciation, it becomes clear that the former works better than the latter. The visualization of vowels was applied to class activities in which native and non-native speakers' English was compared and their feedback was collected: the positions of six vowels by the latter did not scatter as much as they were expected to do. Specifically, two vowels were not discriminated and were arranged very close in the vowel space.

It was surprising for the author to find that learners liked analyzing their own pronunciations by drawing formant ones and twos on a sheet of paper with a pencil. Even a simple method works well if it leads the learners to think about their pronunciation analytically.

1. Introduction

Visual information helps us to learn things happening around us. It is represented in forms of a symbol, a picture or a language. The links between pictures and languages are strong. As is described in [1], mastering language is one unique capacity of our species; another is use of pictures. The pictures have existed for longer periods of time than written languages. The former representations are about 40,000 years old and the latter ones have appeared much later.

While observing history, we can find strong connections between spoken languages and

written ones. Reference [1] describes that, for a long time, pictures and speeches were completely separate domains; around 5,000 years ago a revolutionary invention came about, the most important in the history of mankind; people began producing pictures that were not symbols of creatures or things but symbols of letters that could be ordered in sequence to imitate the strict temporal succession of the sounds and words of speech.

Languages and pictures are both used for explanation of objects or phenomena in media, such as books and magazines. Languages are added to pictures for their explanation, or the latter are added to the former to give readers some additional cues for understanding of these objects or phenomena. This is the case when a very difficult concept is explained. For example, [2] points out that hard- to-understand doctrines are to be changed into simpler ones when additional stories or visualization are utilized.

Things happening around us are put into pictures with a time factor. Moving pictures are usually created with twenty-four frames per second and give us information that changes according to the time axis. Still pictures, on the other hand, visualize objects at a certain period of time span. There is no change of the objects based on the time axis. The time factor, however, is sometimes hidden in the still pictures. For example, a roll painting lists several historical events that have been happening over a long period. Even standard types of pictures contain chronological facts, and by looking at it, we may feel a dimension of a time factor.

As is described in [3], the physicist's symbol t is a deceivingly simple representation of "what we mean by time." As for conceptions of time, people in nations can be resources for finding different approaches on the basis of different cultures. Reference [4] states that the Indian conception of time is very different from what the Western mind regards as intuitively obvious, and in Indian thought, time, like other phenomena, is conceived statically rather than dynamically. He also refers to Japanese Buddhism that emphasizes the transience of the phenomenal world. He claims that the Japanese attitude toward this transience is very different from the Indian, and the Japanese disposition tends to lay a greater emphasis upon sensible, concrete events, intuitively apprehended, than upon universals.

Phenomena in nature or human activities are visualized and they always have something to do with the time factor. Some situations are described in still pictures of two or three dimensions. Changes in these situations are also described in moving pictures of two or three dimensions. In these cases, the change of objects or phenomena according to the time

axis is visualized by putting it into one of three or four dimensions. As is pointed out in [5], in dimensional analysis, we may mix units of distance and those of time together in various ways, obtaining new sorts of physical entities. We may also think of velocity as being the ratio of a distance to a duration, and acceleration as the ratio of velocity change to duration.

Music is played or sung on the time axis and it possesses temporal features, such as tempi, rhythm and duration. It is depicted in two dimensions with the time factor. People understand visualized music and can hear the sounds that come from it. This is because, as [6] refers to, music is itself knowledge, and in its way, very specific knowledge. Musicians can hear what they are writing.

Without the dimension of the time, in a way, things cannot be looked at and sounds cannot be heard. The importance of time is pointed out in [7]: When we talk about visual patterns we consider only spatial dimensions, disregarding the dimension of time. Under most circumstances, time does seem to be irrelevant. Yet, physiological processes do have a temporal dimension, and even in the process of seeing, which strikes us as taking place instantaneously, time plays a role. The identification of such simple figures as triangles and circles requires time and consequently, requires temporal integration in the central nervous system.

Music nourishes human talents. As is pointed out in [8], listening to Mozart does temporarily enhance abstract spatial reasoning. There seems to be a critical period for perceiving features of music. For example, as is pointed out in [8], infants at six months can readily detect all rhythmic variations, but by twelve months their range has narrowed, albeit sharpened. Features of music are directly connected to brain regions. Reference [8] says that there are two basic categories of musical perception, one involving the recognition of melodies, the other the perception of rhythm or time intervals, and impairments of melody usually go with right-hemisphere lesions but representation of rhythm is much more widespread and robust and involves not only the left hemisphere, but many subcortical systems in the basal ganglia, the cerebellum, and other areas.

Perception of music and sound has a strong connection with dimensions. Reference [8] introduces a musician's comment; when we listen to music, we are actually perceiving multiple attributes or dimensions. Among these he includes tone, pitch, timbre, loudness, tempo, rhythm, and contour.

2. Pictures and figures for learning language sound

Visual assistance and its effect for learning language sounds have been discussed widely. Multimodal speech (e.g., auditory, visual, haptic) is the primary mode of speech perception [9]. For example, a very simplified illustration is used for telling learners which part of articulation organs are used for pronouncing sounds as is listed below:

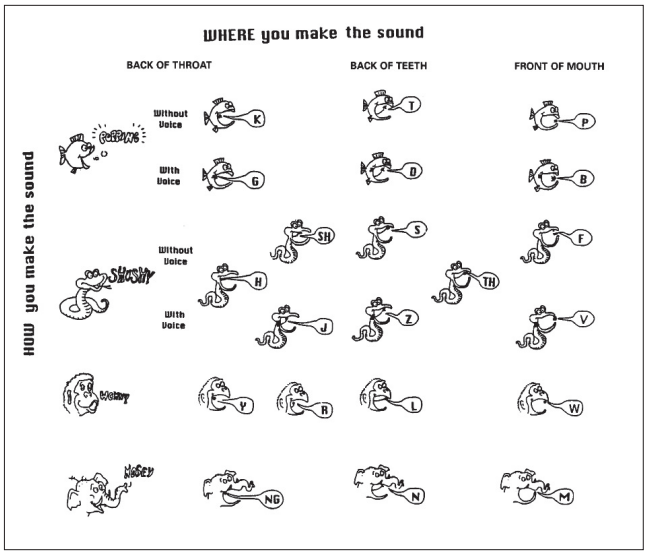

Figure 1　How to make consonants: quoted from [10]

Vowels and consonants are shown on charts that are made up with two axes; one is for opening/closing of a mouth and the other is for positions of a tongue. For vowels, the former is exactly the opening/closing of the mouth and the latter is the peak of the tongue shape. For consonants, the former corresponds to manners of sounds that are created with the opening/closing of two parts of organs, such as two lips, a tongue and palatal, velar, pharynx or glottal, and the latter matches the places of articulation that are decided by two parts of these organs.

It is said that musicians hear notes while they are writing them. It is claimed that phonology influences the recognition of printed words [11]. Similarly, if language learners hear what they look at, such as, phonetic alphabets or alphabetic letters, they would gain a lot of phonological and phonetic knowledge and become a good listener and speaker of that language.

2.1. Vowels

Vowels of world languages could be made into charts. They are used for documenting the learning characteristics of a certain language vowels or comparing them with other language ones. Number of vowels differs among languages. Their positions in a vowel space are different among languages. An example of Japanese ones drawn by the author on the basis of the International Phonetic Association is listed below:

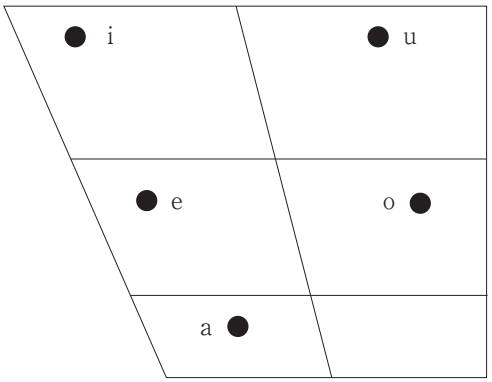

Figure 2 Japanese vowels

An example of American English ones drawn by the author on the basis of the International Phonetic Association is listed below:

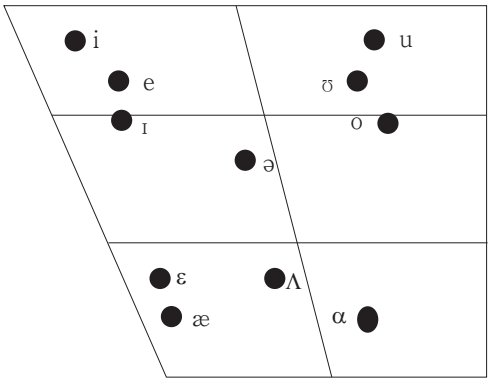

Figure 3 American English vowels

Putting too much information into one chart does not work very well in classes. Using presentation software is recommended in this case. By showing several charts that contain vowels of several languages, different positions in the vowel space for them can be clearly shown.

2.2. Consonants

Consonants of languages of the world could be made into tables. It is possible to learn that some languages contain, for example, two plosive sounds that are pronounced by using two lips while other languages contain more or less than that. An example of Japanese and American English ones drawn by the author on the basis of the International Phonetic Association is listed in Table 1.

Table 1 Consonants of Japanese and American English

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Uvular	Glotal
Plosive	p b		t d	*t* *d*			k g		
Affricate				t^{s}	***tʃ dʒ***				
Nasal	m		n	*n*			ŋ	N	
Flap					ɾ				
Fricative	ɸ	*f* *v*	θ ð	s z	ʃ ʒ				 h
Approximant				ɹ		j	w		
Lateral approximant				*l*					

Consonants which have the same manner and position of articulation for English and Japanese are presented in small letters and those whose manner or position differs in big letters, with non-italics for Japanese and italics for English.

To show different positions and manners of consonants for Japanese and American English, consonants of these languages have been put into one chart. Putting, however, too much information in one chart may not work very well. By presenting consonant spaces of several languages, it becomes clear that every language has different consonants that can be drawn in the consonant spaces.

3. Visualization for learning language sound

Visualization is made use of for academic textbooks. Science textbooks contain graphs, tables, and pictures. History books hold chronological lists and pictures. Language ones have illustrations and pictures. Media which contain moving pictures and sounds are used for various fields. For language learners, a film is an important medium for listening to dialogs and speech in natural speed.

For learning foreign language pronunciations, still or moving pictures can be used. Then, which one, still or moving, or two- or three-dimensions is better for learning pronunciation?

3.1 Still or motion pictures for learning language sounds

Time is an essential factor for motion pictures. They cannot be called so without a time axis. Audiences always perceive changes that are presented in front of them. Still pictures, on the other hand, do not reveal changes that should have happened or would be happening. A time factor is, in a way, hidden in them.

It takes a long time to learn a foreign language very well. It also takes time to be able to write, read, speak or listen to foreign languages. Writing may take more time than reading, and speaking may take a longer time than listening. In addition, listening needs a special skill to deal with time restrictions. It is a hard task for a listener to listen to a speaker when he/she is hasty in his/her speech.

For language learners, time sometimes could be a hardship that puts a restriction on fully understanding of what they are listening to. For learning meanings of languages, a time axis is treated indirectly. Saussure divides concepts of languages into two; form and substance. Form acts in a static situation and substance acts as a dynamic existence.

Static and dynamic experiences are sometimes perceived at the same time. They are, however, put into different modules in our mind. What brain scientists tell us is that a system dedicated to detecting movement is in reality distinct from ones detecting either shape or color [12].

3.2 Two dimensions or three dimensions for pronunciation training

Three dimensions are popular for phonetic studies. Reference [13] recommends three dimensions for generative topographic mapping studies. Since initial experiments suggested that a two-dimensional latent space produced too large an overlap between consonants, a three-dimensional generative topographic mapping was constructed.

Two dimensions are usually used to learn about language sounds. This may be the best way to learn foreign language pronunciations. Speaking takes time, and positions of sound production are always changing as time goes by. Speech organs change quickly and it is not easy to grasp moving organs or their positions. Then the factor of time may be better to be discarded.

Textbooks for pronunciation training have a long history. They started with explanations, examples, illustrations and pictures. Then they have come to include real sounds or even moving pictures. Last of all they come to have software with which learners can compare their own pronunciations with model pronunciations by native speakers. There are, however, the ones published with very simple illustrations even now [14].

Systems for creating objects in three dimensions have been invented, and they have now become rather trendy tools. Human beings are living not in two but three dimensions and living things and natural environment around us all have bodies of three dimensions. They, however, can be depicted on a sheet of paper with two dimensions by making use of visual perspectives. Even when there is a necessity to make a copy of these things, they are always printed in two dimensions. Of course there are cases where things are made to look like having bodies of three dimensions by making full use of artificial techniques in perspectives.

Objects that look like being in three dimensions are drawn on a sheet of paper. In addition, pictures, illustrations or some types of shapes are easily depicted in perspective. We know, for example, how to write a shape of box in two dimensions which looks like a real box standing in three dimensions.

Graphs are also written in ways in which they look like existing in a world of three dimensions. The axes of x, y, z are used and factors are put into each of them. Sometimes, a time factor is put into one of them, in which changes that happen based on the time are presented clearly.

Traditionally, two dimensions were used to depict things or phenomena happening around people. Depicting sounds started a long time ago with two dimensions. As is introduced in [15], the Arab tradition produced the first known vocal tract diagram with a visual representation of speech organs by appropriate arranging of Arabic letters to articulation places for Arabic speech sounds in the late 12th or early 13th century A.D.

4. Discussion and conclusion

This is a trial to appeal to all enthusiasts of seeing the invisible [16], in the fields of phonetics

and their application to language education. Through our eyes, we see things and through our ears we hear sounds. These two types of channels are important for acquiring skills. Sounds are invisible but there should be ways to make them be visible. Pictures are silent but there should be ways to make them be heard.

People connect symbols to sounds and these connections have an effect on their daily lives. It is possible for people to infer product attributes through the phonetic symbolism of these sounds in the same way they do that from the sound of a brand name [17], [18].

Visualization would help people to listen to sound. Visualization itself, however, may not be an easy task. As is described in [19], it is almost impossible to visualize exactly what is happening when an audible vibration occurs, because everything moves so fast.

Supposing visualization helps learners to listen to foreign language sounds better, then how this happens would be an important issue to be explored. It might be better to start with studying mechanisms that are happening in the native speakers' mind. Or it might be more useful to directly study those in foreign language learning.

Learning about language sounds and their visualization is similar to asking essential questions, such as "What is sound?" After reading the following long passage of quotation from [19], it may be helpful to think about a similar question, "What is language sound?":

> Pythagoras believed we should try to understand music in terms of arithmetic. Galileo and Mersenne argued that the arithmetic wasn't important of itself, but was merely a reflection of the physical motions of sound sources. Then Sauveur and Rameau shifted attention from these motions to the properties of the sounds they gave rise to, and argued that therein was to be found an understanding of the structure of music. Lastly, Helmholtz showed, again, that it wasn't the sound itself which was important but the way our ears respond to that sound. All the time the centre of attention is getting closer to the human brain.

The visualization of vowels has been applied to class activities and learners' feedback is quoted from [20].

> Feedback 1: Positions of six vowels did not scatter as much as they were expected to do. Two vowels were not discriminated and they were arranged very close in the vowel space. Pronunciations produced by myself were not as clear as expected.

Feedback 2: I tried to close my mouth for uttering the word “heed” and open it a little for the word “hid”. The result was, however, vice versa. My mouth was closed to pronounce the word “hid”. This kind of task usually did not come to me so this was a very interesting activity.

Feedback 3: It was thought that my mouth was always tightly closed or widely opened when English was spoken. After looking at the figures, however, it was found that it did not move as expected. Those who spoke good English would open or close their mouths and moved their tongues more accurately than was believed.

Feedback 4: This was the first time for me to analyze my own vowel pronunciation. That caused me to be a little nervous. The result showed that vowels in “heed” and “hid” were discriminated very well. That made me happy. Vowels in “hood” and “hoodoo”, however, did not show much difference. That brought me into a sorrow.

Feedback 5: Formant one and formant two varied and among six vowels, two were pronounced well, but the rests were so terrible. It must be necessary to do practice for clear pronunciation.

Samples of this feedback resulted from activities in which native and non-native speakers' English were measured by sound analyzing software. Formant one and formant two of vowels produced by Japanese-speaking English learners were compared with the native speakers' counterparts. Feedback from learners about methods of pinpointing each vowel in vowel spaces and comparing them with the native speakers' ones were collected and estimated.

Formant one, which is a reflection of opening of a mouth, and formant two, that of a tongue shape, are measured. Results show that even university students whose English skills are expected to be high do not discriminate some vowels very well.

All feedbacks show a positive attitude, in which she/he will pay a special attention to opening or closing of mouth and shapes of tongues from then on. Learners seem to like to visualize their own vowels.

It is surprising for the author to find that learners like analyzing their own pronunciation

with connecting formant ones and twos on a sheet of paper with a pencil. Even a simple method works if it leads learners to think about their pronunciation analytically.

Acknowledgements

Much appreciation goes to colleagues, students, loved ones for their support of my language experiments. This research is partially 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

[1] Janson, T. (2012). *The History of Languages: An Introduction.* Oxford: Oxford University Press.

[2] Komatsu, K. (2010). *Youkai Emaki: Nihon no Ikai wo Nozoku (Roll Paintings of Specters: Gazing into Japanese Alien World).* Tokyo: Heibonsha Co., Ltd.

[3] Fraser, J.T. (1966). *The Voices of Time.* Amherst: The University of Massachusetts Press.

[4] Nakamura, H. (1966). Time in Indian and Japanese thought. J.T. Fraser (Ed.) *The Voices of Time.* Amherst: The University of Massachusetts Press.

[5] Clemence, G.M. (1966). Time measurement for scientific use. J.T. Fraser (Ed.) *The Voices of Time.* Amherst: The University of Massachusetts Press.

[6] Leppert, R. (2007). *Sound Judgment.* Hampshire: Ashgate Publishing Limited.

[7] Lenneberg, E.H. (1967). *Biological Foundations of Language.* New York: Wiley & Sons.

[8] Sacks, O. (2007). *Mussicophilia.* New York: Vintage Books.

[9] Rosenblum, L.D. (2004). Primacy of multimodal speech perception. In D.B. Pisoni and R.E. Remez (Eds.), *The Handbook of Speech Perception.* M.A.: Blackwell.

[10] Nicholson, T. (2005). *Phonics Handbook.* London: Whurr Publishers.

[11] Gaskell, G. (2007). *Oxford Handbook of Psycholinguistics.* Oxford: Oxford University Press.

[12] Matthews, P. H. (2003). *Linguistics: A Very Short Introduction.* Oxford: Oxford University Press.

[13] Lecumberri, M. L. G., Gong, J. and Cooke, M. (1995). Exploring the representation of second language sounds in first language space. *Journal of the Phonetic Society of*

Japan. 19: 1, 32-43.

[14] Cravotta, J. S., Takagi, M. and Takeda, S. (1999). *Eigo no Hatsuon (English Pronunciation).* Tokyo: Eikosha.

[15] Heselwood, B., and Hassan, Z.M. *Instrumental Studies in Arabic Phonetics.* Philadelphia: John Benjamins Publishing Company.

[16] Settles, G.S. (2001). *Schilieren and Shadowgraph Techniques: Visualizing Phenomena in Transparent Media.* New York: Springer.

[17] Yorkston, E., and Menon, G. (2004). A sound idea: Phonetic effects of brand names on consumer judgments. *Journal of Consumer Research,* 31:1, 43-51.

[18] Elder, R.S., Aydinoglu, N.Z., Barger, V., Caldara, C., Chun, H., Lee, C.J., Mohr, G.S., and Stamatogiannakis, A. (2010). A sense of things to come: Future research directions in sensory marketing. In A. Krishna (Ed.) *Sensory Marketing: Research on the Sensuality of Products.* New York：Routledge.

[19] Johnston, I. (2002). *Measured Tones.* New York: Taylor & Francis.

[20] Tomita, K. (2016). Visualization for Learning Foreign Speech. *Bulletin of Yamagata University (Humanities),* 18:3, 1-14.

Analytical Visualization for Learning Foreign Language Sounds

TOMITA Kaoru
(Faculty of Literature and Social Sciences)

Abstract

This study explores how visual information helps us to learn foreign language pronunciation. Visual assistance and its effect for learning foreign language have been discussed widely. For example, simplified illustrations in textbooks are used for telling learners which part of the articulation organs are used for pronouncing sounds. Vowels are put into a chart that depicts a vowel space. Consonants are put into a table that contains two axes of place and manner of articulation.

When comparing a still picture and a motion picture for visualizing learners' pronunciation, it becomes clear that the former works better than the latter. The visualization of vowels was applied to class activities in which native and non-native speakers' English was compared and their feedback was collected: the positions of six vowels by the latter did not scatter as much as they were expected to do. Specifically, two vowels were not discriminated and were arranged very close in their vowel space.

It was surprising for the author to find that learners liked analyzing their own pronunciation by drawing formant ones and twos on a sheet of paper with a pencil. Even a simple method works well if it leads the learners to think about their pronunciations analytically.