The Introduction of Nihongo Speech Trainer:

A Tool for Learning Japanese Sounds

Tanporn Trakantalerngsak

Abstract

This paper investigates the computer-assisted pronunciation training "Nihongo Speech Trainer", aimed at helping Thai learners of Japanese to improve their perception and production of Japanese contrasts. The tool focuses on specific sounds which are said to be problematic for Thai learners. Perceptual training using a high-variability phonetic training method ("HVPT perceptual training"), to be specific, a wide range of variations in training stimuli (real words and nonsense words), a number of native speakers' input, multiple phonetic contexts and multiple word positions of the L2 contrasts were used in the tool. Moreover, a tutorial video was also integrated in the Nihongo Speech Trainer. In this article, an overview of the Nihongo Speech Trainer system, key features and the user interface will be introduced.

Introduction

High-variability phonetic training (HVPT) method perceptual training has shown itself to be the most effective tool for improving learners' ability to accurately perceive L2 consonants, vowels and suprasegmentals such as pitch and tone. Furthermore, the improvement gained from this type of training has also been shown to have generalized to new tokens and new talkers, and these improvements were retained in the long-term (Lively, Logan, & Pisoni 1993; Bradlow, Akahane-Yamada, Pisoni, & Tohkura 1999; Hirata, 2004; Iverson, Pinet, & Evans, 2012). Some studies also report that perceptual improvement successfully generalizes to production (Bradlow et al., 1999; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005). However, despite these promising results, very little additional research has directly investigated the application of HVPT in computer-assisted pronunciation training applications (Barriuso & Hayes-Harb, 2018). Moreover, as Thomson (2011) also stated, if a web-based applications were available, it "would allow endless research

possibilities, as teachers and researchers could collaborate remotely, monitoring the effect of perceptual training and its impact on pronunciation, in order to improve future iterations of the software" (p. 760). However, regarding the idea of developing computer-assisted pronunciation training applications, it has been stated that an efficient and effective pronunciation learning requires guidance from a teacher who can evaluate individual students' segmental needs, identify appropriate software, select appropriate study units, and monitor student progress over time, making adjustments as needed (Derwing & Munro, 2015). Tsai (2015) also states that teachers should not expect that technology can solve all the students' learning problems. Instead, they should pay attention to the different roles assigned to technology and other kinds of mediation. If teachers can introduce various mediating tools to their students to facilitate their learning at different learning stages, they will be able to assist them to move to the next advanced learning stage.

This project has created the web-based tool called "Nihongo Speech Trainer", aiming to create a freely available website to Thai learners for teaching Japanese sounds they wish to work on by adopting the theoretical HVPT perceptual training within computer-assisted pronunciation training applications. Moreover, to this end, it is hoped that the teacher role will contribute to the efficacy of the training, and since the ultimate goal of this study is not to replace the teacher with technology (Derwing & Munro, 2015), a combination of traditional instruction by an instructor focusing on the pedagogical task and computer-assisted pronunciation training using HVPT perceptual training were integrated. The tool focuses on the 11 contrasts which are ts, z, tɕ, ɕ, (d)ʑ, d, b, g, long-short vowel, geminate consonant and diphthong/. The reason that these particular contrasts were chosen is because these sounds have been proven to be difficult for Thai learners and are the area in which the most frequent errors are made by Thai learners both in perception and production (Kawano, 2014; Sukegawa, 1993). The development of the Nihongo Speech Trainer was funded by Mahidol University as a one-year project. The project was based at Faculty of Liberal Arts, Mahidol University. This paper describes an overview of "Nihongo Speech Trainer", a web-based online pronunciation training program designed for Japanese pronunciation training.

Description of the System

General overview of Nihongo Speech Trainer

The web-based online pronunciation training "Nihongo Speech Trainer" is a free, open source online website (*www.nihongospeechtrainer.com*) which is easy to learn and to use. "Nihongo Speech Trainer" was designed by Ecgates Solution

Co., Ltd in Bangkok. The tool comprises two components: 1) the video tutorial, and 2) the perceptual training. "Nihongo Speech Trainer" took the form of a web-based learning tool accessible via computers and hand-held electronic devices. The layout of the tool's user interface was kept Japanese-style to attract the young generation of learners with Japanese characters and vivid colors (Figure 1).

Figure 1. Interface of Nihongo Speech Trainer

Components of the tool – Technical and pedagogical features

Nihongo Speech Trainer aimed to offer the following functions.

- Use of variability: Research on perceptual training has shown that increasing the variability of the input – the so-called "HVPT method" – results in greater and more generalizable gains in L2 speech perception. This method provides a wide range of variations in training stimuli, in terms of the numbers of native speakers' input, multiple phonetic contexts and multiple word positions of the L2 contrasts (Lively et al., 1993). To follow the criteria of the HVPT method, the use of multiple talkers (seven Japanese native speakers), various phonetic contexts (/a/, /i/, /u/, /e/, /o/), multiple word positions (initial singleton, intervocalic and final positions) and word types (nonsense words and real words) were used in this tool. Seven Japanese native speakers recorded 11 sets of the target contrasts consisting of 75 minimal pairs (75 words x 2 minimal pairs x11 contrasts = 1,650 tokens). The contrast pairings were /ts/-/s/ or /ts/-/z/; /tɕ/-/ɕ/ or /tɕ/-/(d)ʑ/; /d/-/t/ or /d/-/r/; /g/-/k/; long-short vowel contrasts; geminate consonant & non-geminate consonant contrasts; diphthong & non-diphthong contrasts.

- Tutorial video: According to Tsai (2015), teachers should not expect that technology can solve all the students' learning problems. Instead, they should pay attention to the different roles assigned to technology and other kinds of mediation. If teachers can introduce various mediating tools to their students to facilitate their learning at different learning stages, they will be able to assist them to move to the next advanced learning stage. In other words, the role of the instructor is essential for learners to maximize the effect of the training (Gilakjani, 2017), hence, the production of the VDO tutorial focusing on perception and production techniques was employed in this project. First, vocal tract diagrams to visualize the articulation were presented in each contrast. Strategies such as encouraging learners not to merge the sound that does not exist in Thai such as /ts/ to the similar Thai sound category (/s/) by comparing them the differences of those sounds were used. Moreover, minimal-pairs used to demonstrate that the two phones are separate phonemes in Japanese were presented (Figure 3). Tutorial videos were limited to five minutes for each contrast so that the users did not feel too overwhelmed by the video content.

- Minimal-pairs exercises on perception that target common perception and production difficulties for Thai learners: the presentation of minimal pairs allows learners to focus and promote their awareness on specific phonemic contrasts (Celce-Murcia & Goodwin, 1996). Moreover, the perceptual training exercises use two alternative forced-choice identification tasks, since identification tasks are effective in developing new phonetic categories as they primarily expose learners to within-category variability or variability in the phonetic detail of the target phoneme which consequently facilitates generalization of learning (Logan & Pruitt, 1995; Carlet, 2017).

- Feedback: Feedback is an important feature to give opportunities for learners to detect their errors (Logan & Pruitt, 1995). Each training task included immediate trial by trial feedback and cumulative feedback which was provided at the end of each session. Immediate feedback was given by means of pictographic information (Figure 2) after each response. If the identification of the target segment was correct, participants could listen to the next trial, but if they identified the contrast incorrectly, a message was then displayed, and they could listen to the correct and the incorrect stimulus again and until they could choose the correct sound. Moreover, the participants were also asked whether they want to train on the tokens they misperceive again. Regarding the cumulative feedback, the overall grade (pre-test, training and post-test scores) was reported at the end of the training.
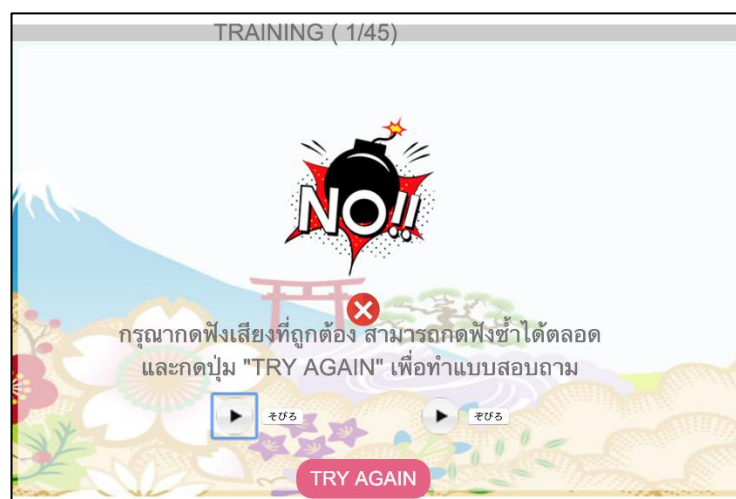
Figure 2. Feedback (a pictograph of a bomb shows up on the page when an incorrect choice is made)

- Design: According to Yan Lin, and Liu (2018), a user-friendly interface can motivate students to participate in more pronunciation practice than traditional paper-based and classroom exercises. Also, to mitigate user's feeling that the computer is "mechanical", the design incorporated an interactive environment based on the modern Japanese style. Moreover, with this specific design it is hoped that learners feel encouraged from the beginning of the training (Figure 1 and 3). In the aspect of new directions in teaching pronunciation proposed by Celce-Murcia and Goodwin (1996, pp. 311-315), this training meet the criteria as illustrated in Table 1. Some features such as speech spectrographic devices is helpful in visualizing the pitch movement (Matsuzaki, 2002). However, this study does not focus on suprasegmentals so the speech spectrographic was not used in this tool. This tool presented here make only a first step into a thorough application of HVPT perceptual training on a computer assisted pronunciation training. The features such as Automatic Speech Recognition and the functions which provide interaction with other classmates are to be considered in the future research since these functions require high performance computing.

Table 1. General features of CAPT

1	Using multimedia in teaching pronunciation	✓
2	Audio feedback	✓
3	Video	✓
4	Computer-assisted instruction	✓
5	Speech Spectrographic devices	
6	System incorporating Automatic Speech Recognition modules	
7	Stress free environment	✓

8	Learners centered: focus on individual problem, allow self-pace and self-directed learning	✓
9	Provides immediate corrective feedback	✓
10	Provides multiple samples of native speakers	✓
11	Interaction with the speakers in the software and classmates (incorporating Automatic Speech Recognition modules)	
12	Focus on those segmental and suprasegmental aspects	✓

System Structure

The training was structured as follows:

1. Introduction

Nihongo Speech Trainer was self-paced and completed outside of class time. Users were given the username and password to log into their own account. The simple instructions of how to use the website in Thai and English were given on the "Home" page. Participants were presented with a list of 11 phonemic contrasts on the page (/ts, z, tɕ, ɕ, (d)ʑ, d, b, g, long-short vowel, geminate consonant and diphthong/). They could choose the contrast that best suited their needs and interest hence the training content differed for each participant, since their problems varied in content and number, which added an extra independent variable that was not controlled for (Figure 3).

Figure 3. Interface of Nihongo Speech Trainer (11 target contrast)

2. Pre/post-test

After participants chose the problematic contrast to train on the main page, they were given a pre-test, the training itself and then the post-test. The pre/post-test were conducted to measure and compare possible improvements on perception

and production ability. The post-test and the pre-test were similar. Participants were asked to perform an identification task before they continue to the training phase. There were 20 words produced by a female native speaker. At least two-alternative forced choices available formed choices relating to the target word (e.g., after hearing "つきに" the participants were asked to identify it as "すきに" or "つきに" or "じゅきに"). The words used in the pre/post-test were not used in the training. There was no feedback provided in this section and they could not replay the stimulus. After they finished the pre-test phase they then were subsequently directed to the training phase. The pre/post-test lasted approximately three minutes.

3. Pronunciation video tutorial

As suggested by Derwing and Munro (2005), it is necessary to have the help of teachers who have a foundation in pronunciation research and are able to draw comparisons between L1 and L2. This training part has a pedagogical basis in providing a tutorial video consisting of a description of Japanese, presented with animated vocal tract diagrams, phonetic and articulatory descriptions and tips for learning Japanese pronunciation. The audio and video recordings were made by the author and a native Japanese teaching assistant. This tutorial aims to help strategy development to guide learners in developing rules to perceive the contrast. To be specific, it focuses on targeted contrasts which are difficult for Thai learners. It is hoped that the tutorial video could help Thai learners understand how Japanese contrasts differ from the Thai contrasts. Ideally with such explanation, it is hoped that learners could narrow down the difference between the Thai and the Japanese contrasts. The tutorial videos were limited to five minutes because it was hoped that it would not be too long for the learners (Figure 4).

4. Training

In each training, participants completed a two forced-choice identification task (e.g., Is the word you hear "あすま /asuma/" or "あずま /azuma/"?). The sounds used in the training are produced by seven different speakers. The order was randomly chosen by the system. Target sounds are provided in a wide variety of phonetic environments (e.g., [a], [i], [o]) situated in various word locations (e.g., initial, medial and final) and word types (nonsense and real words). Moreover, there were three selections of stimuli size in the training of this study to see the effect of stimuli volume and to see which input size participants tend to choose. The users can choose to train with the volume of 45, 60 or 75 minimal pairs on the home screen. The quantity and length of the training varied according to a participants' training performance, varying from 15 minutes to 20 minutes each contrast. In the training phase, participants identified the sounds and were given

immediate feedback regarding the correct answer after each trial. If the identification of the target segment was correct, participants could listen to the next trial, but if they identified the contrast incorrectly, a message was then displayed, and they could listen to the correct and the incorrect stimulus again and until they could choose the correct sound. Moreover, the participants were also asked whether they wanted to train on the tokens they misperceive again. After they finished the training, they were then given the post-test to measure whether an improvement occurred subsequent to the training.

5. Questionnaire

After the post-test, the participants were directed to a Google form to complete the questionnaire. The questionnaire aimed to gather more insightful data from the participants. The questions were as follows:

1) What do you think of "Nihongo Speech Trainer"?

2) Do you think "Nihongo Speech Trainer" help listening and speaking skills? Or one another?

3) Which part of the "Nihongo Speech Trainer" do you think is helpful? a) the video tutorial b) the perceptual training c) both trainings

4) What do you think about the design of "Nihongo Speech Trainer"?

5) Do you think the time spent for each section too long?

6) Apart from the 11 contrasts used by "Nihongo Speech Trainer" Do you have any other Japanese contrasts you want to train on?
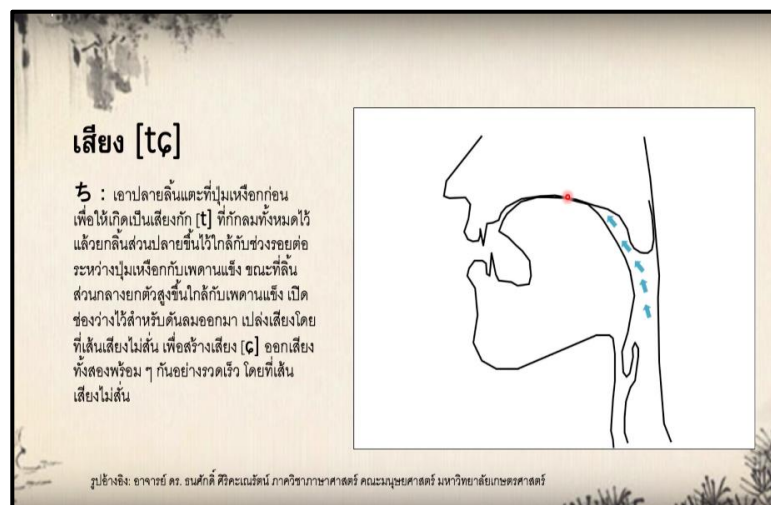
(a)

(b)

Figure 4. The video tutorial (a) the minimal pair (b) a vocal tract diagram

Conclusions and Future Work

In this paper, Nihongo Speech Trainer using HVPT method perceptual training and video tutorial was introduced. Moreover, the theoretical and pedagogical features were also discussed. The main goal of Nihongo Speech Trainer is to provide a tool for improving Japanese pronunciation. Nihongo Speech Trainer is currently being used to investigate the effect of the system on learning among Thai learners. The empirical evidence of the training's efficacy will also be discussed. Moreover, it is strongly hoped that Nihongo Speech Trainer can expand the target contrast and be conducted among other target learners.

References

Bradlow, R. A., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977-985.

Barriuso, T.A., & Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *CATESOL Journal*, *30*(1), 177-194.

Carlet, A. F. (2017). *L2 perception and production of English consonants and vowels by Catalan speakers: The effects of attention and training task in a cross-training study* (Doctoral dissertation). Retrieved from http://hdl.handle.net/10803/403758

Celce-Murcia, M., Brinton, D., & Goodwin, J. (1st ed.: 1996, 2nd ed.: 2011). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. New York: Cambridge University press.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins.

Gilakjani, A. P. (2017). A review of the literature on the integration of technology into the learning and teaching of English language skills. *International Journal of English Linguistics*, *7*(5), 95-106.

Hirata, Y. (2004). Computer assisted pronunciation training for English speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning*, *17*(3-4), 357-376.

Iverson, P., Pinet, M., & Evans, G. B. (2012). Auditory training for experienced and inexperienced second-language learners : Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(1), 145-160.

Kawano, T. (2014). *Onsei kyouiku no jissen* [A practice of speech learning]. Tokyo: Kuroshio.

Lambacher, S., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227-247.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II. The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(2), 1242-1255.

Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 351-377). Baltimore, MD: York Press.

Matsuzaki, H. (2002). Learners' mind during repetitious pronunciation practice. *Hiroshima Daigaku Nihongo Kyoiku Kenkyu* [Hiroshima University Department Bulletin Paper], *12*, 33-41.

Sukegawa, Y. (1993). Bogo betsu ni mita hatsuon no keikou – ankeito chousa no kekka kara- [A pronunciation trend as analysed by L1 language]. *Nihongo onsei to nihongo kyouiku monushou juuten ryouiki kenkyuusho kenkyuu seika houkokusho*, 187-222.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, *28*(3), 744-765.

Tsai, P. H. (2015). Computer-assisted pronunciation learning in a collaborative context: A case study in Taiwan. *The Turkish Online Journal of Educational Technology*, *14*(4), 1-13.

Yan, H., Lin, J., & Liu, Y. (2018). EFL pronunciation training with computer-assisted adaptive peer review. *English Language Teaching*, *11*(11), 74-82.