

Nihongo Speech Trainer: A Pronunciation Training System for Japanese Sounds

Tanporn Trakantalerngsak
Mahidol University

Correspondence concerning this article should be addressed to Tanporn Trakantalerngsak, Faculty of Liberal Arts, Mahidol University, 999 Phutthamonthon Sai 4, Nakhonpathom 73170, Thailand.
E-mail: tanporn.tra@mahidol.edu

This article will present the methodology, as well as the results, of a pilot study of the ‘Nihongo Speech Trainer’ aimed at helping Thai learners improve their ability to identify Japanese contrasts. The pilot study was performed on 15 participants. The tool focuses on specific contrasts that are problematic for Thai learners such as Japanese fricatives and affricates. Perceptual training uses a high-variability phonetic training method (hereafter referred to as “HVPT perceptual training”). Each training session included 90 minimal pairs in which the target contrasts were embedded in initial, medial and final positions. The training stimuli were produced by seven Japanese native speakers. The results of the pilot study showed that the use of the Nihongo Speech Trainer can lead to better perception of the trained Japanese sounds. The results of a questionnaire among the participants also showed that the system helped to improve their perception and production ability. However, despite these positive results with the use of the Nihongo Speech Trainer, there is room for improvement, which may lead to better training results.

Keywords: Nihongo Speech Trainer, HVPT perceptual training, L2 speech perception, perception training

Introduction

Perceptual training using a high variability phonetic training method has been shown to be the most effective tool in improving learners’ abilities to accurately perceive L2 consonants, vowels and suprasegmentals, such as pitch and tone. Furthermore, the improvements gained from this type of training have also been shown to have been generalised to new tokens and new talkers, and these improvements have been retained in the long-term (Lively et al., 1993; Bradlow et al., 1999; Hirata, 2004; Iverson et al., 2012). Some studies have also reported that perceptual improvements have successfully generalised to production (Bradlow et al., 1999; Lambacher, et al., 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 application 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). For this reason, for the purposes of this current project, a web-based tool called the “Nihongo Speech Trainer” has been created, aiming to provide a freely available website to Thai learners who can improve the Japanese sounds they wish to work on by adopting theoretical HVPT perceptual training within a computer-assisted pronunciation training application. The development of the Nihongo Speech Trainer was funded by Mahidol University as part of a year-long project based at the Faculty of Liberal Arts of Mahidol University. This article provides an overview of the Nihongo Speech Trainer – a web-based online pronunciation training program designed for Japanese pronunciation training, as well as the results of the pilot study aimed at detecting potential problems that may occur in the main study.

Method

Participants

The pilot study was conducted with a group of 15 Thai learners: undergraduate students studying Japanese at Mahidol University. All were female, aged from 19-24 years old and reported having normal hearing. None had lived in Japan, while all had been studying Japanese for at least six months. They reported that they were able to read Hiragana. Three participants had passed the N5 level of the JLPT certification. However, two participants were removed from the study because they did not complete the tasks correctly. Each participant was paid 100 baht to participate in the pilot study.

Procedure

This section will outline the design and structure of the Nihongo Speech Trainer. Seven Japanese native speakers produced 75 minimal pairs of the target contrasts (11 contrasts x 2 minimal pairs x 75 tokens = 1,650 items). The pre/post-test stimuli were produced by a female Japanese speaker (11 contrasts x 20 tokens = 220 items). The recording was carried out in a soundproof recording room with a high-quality recorder at the Faculty of Liberal Arts of Mahidol University. The settings of the recorder were set so that recording was conducted using a 32-bit mono channel and a sampling frequency of 44.1 kHz. The stimuli were presented via Microsoft Power Point, one by one in a randomized order. Regarding the procedure, all participants were given a username and a password to log on to their account on the application webpage. They were asked to perform the training individually and given instructions through “Line”, a freeware application used for instant messaging produced by Line Corporation. The training was structured as follows:

Introducing the training to the participants. Nihongo Speech Trainer was self-paced and completed outside of class time. Users were given a username and password with which to log in to their own account. 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)ʒ, 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 (See Figure 1.).

Pre/post-test. After the participants chose their problematic contrast to train on the main page, they were given a pre-test, the training itself and then a post-test. The pre/post-test were conducted in order to measure and compare possible improvements in perception and production ability. The post-test and the pre-test were identical. There were 20 words with at least two choices to select for each contrast. There was no feedback provided in this section. After users finished the pre-test phase they were then subsequently directed to the training phase.

Training. In each training session, 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 were produced by seven different speakers. The order was randomly chosen by the application’s software. Target sounds were 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 choices of stimuli quantity in the training of this study to see the effect of stimuli volume (45, 60 or 75 tokens). The quantity and length of the training varied according to the participants’ training performance, varying from 10 minutes to 15 minutes for each contrast. In the training phase, participants were asked to press the “play” button to listen to the item and that item was then played twice automatically. Participants identified the sounds and were given immediate feedback regarding the correct answer after each attempt. 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 until they managed to choose the correct sound. Moreover, the participants were also asked whether they wanted further training for the tokens they misperceived. After they finished the training, they were then given the post-test to measure whether or not an improvement had occurred subsequent to the training.

Questionnaire. After the post-test, the participants were then directed to a Google form to complete the questionnaire. The questionnaire aimed to gather more insightful data from the participants.

Results and Discussion

Training efficacy

Participants' pre-test and post-test scores in each contrast were compared in order to examine whether the training facilitated improvements in the participants' perceptual skills in identifying trained contrasts. Table 1 displays the number of training sessions which occurred. Figure 3 illustrates participants' pre-test and post-test identification accuracy scores for each contrast. According to Figure 3, there was no attempt at training /d/ and [(d)z] made by any participant. Positive gains were observed in /z/, /ts/, /tɕ/ /g/ and sokuon. Negative gains were observed in /ɕ/ and the long-short vowel. No gain was observed in /b/ and yoon.

Participants used the "Nihongo Speech Trainer" to train with an average number of 2 contrasts per person.

The pilot study has shown that the training was effective in enhancing the participants' ability to perceptually identify the target contrast in specific contrasts. Nevertheless, the results cannot be said to be generalisable since the number of participants was not big enough to run any statistical analysis. To assess the effectiveness of the training, a higher number of participants would be needed.

Regarding the volume of the training, through 19 attempts 45 tokens were chosen to train with, while three other participants chose to train with 75 tokens.

Table 1
Number of training (N)

Contrast	Numbers of training
/ts/	5
/z/	4
/ɕ/	4
Long-short vowel	4
Yoon	2
/tɕ/	1
/b/	1
/g/	1
Sokuon	1
/z/	0
/d/	0
Total	23

Training contents

According to the questionnaire results, some participants reported that they expected to receive more tips and suggestions on how to learn the contrast, and the characteristics of the contrast they were trained in, such as through a tutorial video.

Some participants reported that being exposed to only listening practices made them feel uncomfortable. They also said that they still wanted to rely more on their teachers' instructions.

Tool system

Some functions were not easy to understand, such as the right or wrong symbols. More visual aids need to be used.

The system sometimes ran slowly. Some participants reported problems with the host connection.

Many participants expressed a desire to use the website on their smartphone.

Two of the participants dropped out because they did not perform the post-test.

Questionnaire

A questionnaire survey was conducted among 15 participants to gauge the effectiveness of the Nihongo Speech Trainer, and to get comments and suggestions from the participants. The survey was carried out through Google Form (See details of the questionnaire in Appendix A).

All of the participants believed that their listening skills improved after using the Nihongo Speech Trainer.

13 out of 15 (87%) reported that the Nihongo Speech Trainer was useful and interesting.

50% thought that the Nihongo Speech Trainer was useful in improving their listening skills and 47% thought that the Nihongo Speech Trainer was useful in improving both speaking and listening skills. 3% thought that the Nihongo Speech Trainer was not useful in helping learn both skills in Japanese.

77% reported that basic explanation of the phonetic characteristics of each contrast might be useful in learning new sounds and it would help the website to be more entertaining. To be specific, they suggested a video allowing basic knowledge of the sound properties and some tips for how to learn the contrast. Moreover, 17.6% suggested that the presentation of articulation might help understand the articulation process.

89% suggested that the Japanese design of the website would help motivate the learners more. They also reported that some symbols were difficult to understand, such as the right or wrong symbols.

94% agreed that the self-paced style fit will with their learning needs because they could get access to repetitive practice sessions and practice at their own pace.

76.5% reported that the length of the training is moderate – not too long or too short. 12% thought that the length was too long. And another 12% thought that the training was too short.

They also showed a demand for training in Japanese prosody and pitch-accent (10 out of 15, 71%).

78% of the participants were satisfied with the system's stability and operability. Four participants reported that problems occurred while doing the training.

In summary, the questionnaire showed a high degree of satisfaction among the participants. It seems that they considered the Nihongo Speech Trainer to be useful and viable as a tool for teaching Japanese pronunciation for Thai learners in improving their perception ability. However, there are things to be taken into consideration to enhance the effectiveness of the tool.

Application of the outcomes of pilot study

The results from the pilot study showed that the Nihongo Speech Trainer was able to raise the learners' awareness in learning Japanese contrasts and was useful in improving their identification ability in the contrasts trained. The results of the pilot study gave insightful suggestions that were beneficial for the main study design. However, to motivate and guide the participants towards a meaningful goal, some points need to be adjusted, as described below in Table 2.

Conclusion

This paper has described the design and development of the Nihongo Speech Trainer as well as reporting the pilot study's results. The results showed that the tool fostered improvements in the learners' perception ability and the questionnaire also revealed that they considered this tool to be useful for educational purposes. However, there are some points to be considered. The main limitation concerns the number of participants. Although the results here make a contribution to the study of perceptual training, they cannot be generalised as the study was carried out with only fifteen participants. However, it would not be difficult to collect data from a larger sample

Table 2
Practical considerations and directions for the main study

Pilot study	Main study
<p><i>Efficacy</i> The participants tested the training system. However, the effectiveness of the system was not sufficiently observable since the number of participants was too small.</p>	<p>➤ To investigate whether a significant improvement has occurred after the training, the main study will be conducted on a larger number of participants aiming at approximately 70 participants.</p>
<p><i>Training content</i> Perceptual training using a two-alternative forced choice identification task was used alone in the training. However, that perceptual training itself seems to have been relatively demotivating since there was no interaction given to the participants.</p>	<p>➤ Based on the questionnaire and the interview, they reported that the role of the instructor is essential for learners to learn the target sounds. Hence, to maximize the effect of the training, the production of a video tutorial focusing on perception and production techniques as well as the perceptual training will be employed in the main study. Teachers should not expect that technology can solve all the students' learning problems and can replace them. There are limitations to what any tool can do and how it can be used. 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 (Derwing & Munro, 2015; Pirani, 2004; Pi-Hua, 2015; Yoshida, 2018). Hence, it is strongly hoped that the additional tutorial video guiding the students through the sound principles of phonology and phonetics will contribute to the efficacy of the training.</p>
<p><i>Tool system</i> A few participants (22%) reported that they were not satisfied with the system's stability and operability.</p>	<p>➤ The main study will improve weaknesses in the pilot study such as the system's stability and operability to have a better connection.</p> <p>➤ It will focus on presenting Japanese styles on the website to motivate learners by using Japanese characters, backgrounds and signs.</p> <p>➤ For the main study, the website will be redesigned, especially the design of the website and elements such as the timer and other functions (pictographs etc.). Some buttons also need to be redesigned to help participants understand the meaning ("Next", "Play", "Start" etc.).</p>

in the main study. A further limitation concerns the system connectivity. We would work on the improvement of system performance for the purposes of the main study to ensure it works reliably. Moreover, the audio sounds and other designs will be modified so that they are of a smaller file size thereby enabling better fluidity of use.

References

- Barriuso, T.A., & Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *CATESOL Journal*, 30(1), 177-194.
- 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.
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam, The Netherlands: Benjamins.
- 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.
- 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 Pr.
- Pi-Hua, T. (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.
- Pirani, J. (2004). *Supporting elearning in higher education*. Educause Center for Applied Research. Retrieved from www.educause.edu/ecar/.
- Pourhossein Gilakjani, A. (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.
- Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, 28(3), 744-765.
- Yoshida, M. (2018). Choosing technology tools to meet pronunciation teaching and learning goals. *The CATESOL Journal*, 30(1), 195-212.

APPENDIX

1. What do you think of Nihongo Speech Trainer?
a. fun b. not fun c. boring d. long e. useful
2. Do you think Nihongo Speech Trainer helped you improve the perception and production ability of the Japanese sounds?
a. not at all b. help listening skill c. help speaking skill d. help improving both skills
3. What function do you think would help enhance the effectiveness of the Nihongo Speech Trainer?
a. perceptual training b. basic explanation of the phonetic characteristics c. articulatory diagrams
4. Do you think the Japanese design of the website would help motivate learning and concentrating?
a. Yes b. No
5. What do you think of the self-paced style of learning used in the Nihongo Speech Trainer?
a. Good b. Not good
6. Do you have any specific contrast you want to train on apart from the sounds used in Nihongo Speech Trainer?
a. Pitch accent b. Intonation c. Vowels d. Other consonants e. Others please specify
7. What do you think about the time length of the training?
a. short b. moderate c. long

Figures

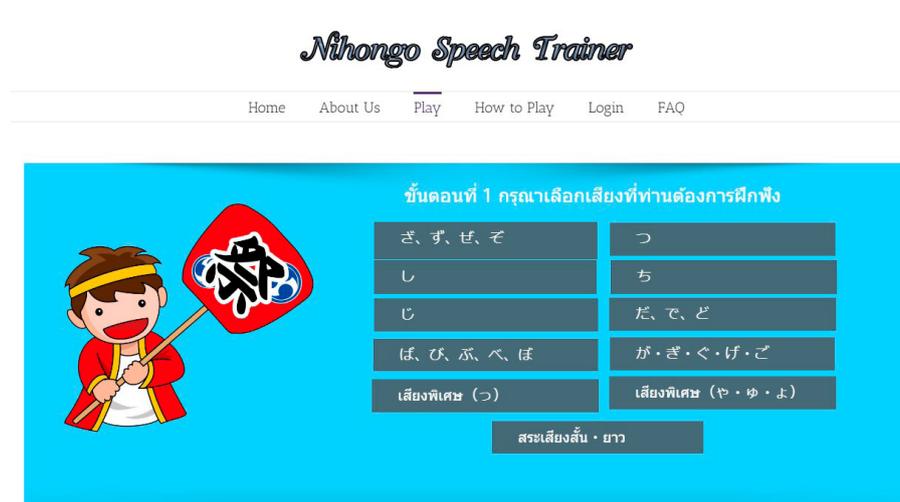


Figure 1. Screenshot of Nihongo Speech Trainer: Target contrasts

A PRONUNCIATION TRAINING SYSTEM FOR JAPANESE SOUNDS



Figure 2. Screenshot of Nihongo Speech Trainer: Two forced-choice identification task.

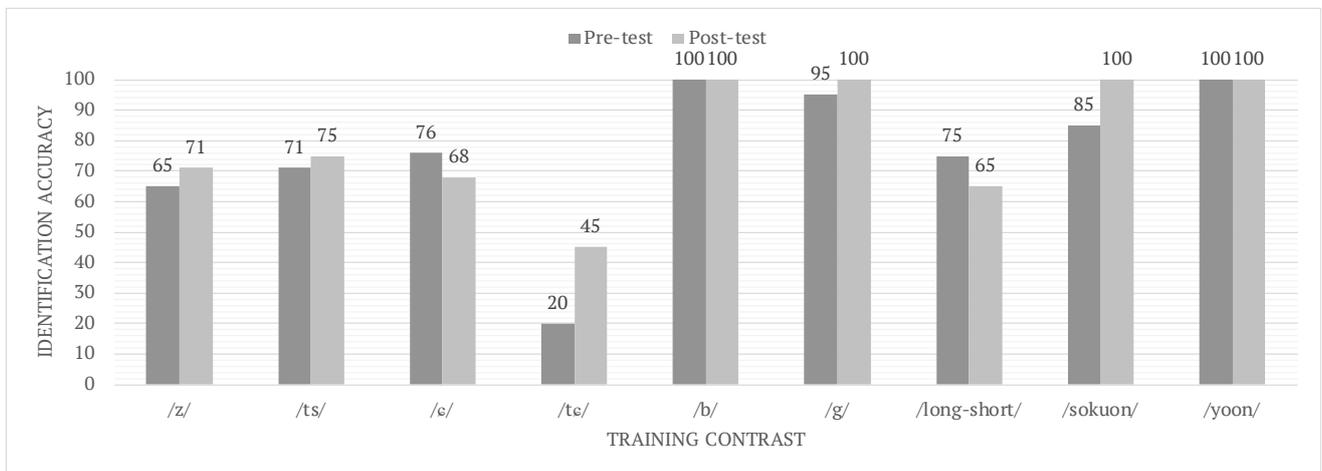


Figure 3. Mean percentage of correct identification scores in Japanese contrasts by participants.