

# Development of Perceptual Training Software for Realizing High Variability Training Paradigm and Self Adaptive Training Paradigm

**Ruining Yang**

Graduate School of Human and  
Environmental Studies  
Kyoto University  
yang.ruining.48z@  
st.kyoto-u.ac.jp

**Hiroaki Nanjo**

Academic Center for  
Computing and Media Study  
Kyoto University  
nanjo@media.  
kyoto-u.ac.jp

**Masatake Dantsuji**

Academic Center for  
Computing and Media Study  
Kyoto University  
mdantsuji@media.  
kyoto-u.ac.jp

## Abstract

The paper describes a development of perceptual training software which realizes a perceptual training paradigm to aid Japanese Mandarin Chinese learners with identifying nasal codas “-n” and “-ng”. This training paradigm is novel from the viewpoint that it retains the primary advantages of High Variability Training Paradigm (HVTP) while adding a Self-Adaptive Training Paradigm (SATP), which can measure items that users are not good at and form personalized training for each user. Our developed educational software is based on the blend of the two training paradigms. The development of the pedagogical software with novel training paradigm is promising for a contribution of bridging the gap between research and practice in second language phonetic training.

## 1 Introduction

It is difficult for second language (L2) learners to perceive and pronounce non-native speech sounds. Plenty of studies have proven this phenomenon. For example, Miyawaki et al. (1975) indicate that it is challenging for Japanese listeners (Japanese native speakers who learn other languages) to identify alveolar approximate / r / and lateral approximate / l / in English. Nasal codas “-n” and “-ng” in Mandarin Chinese is also one of the most difficult phonetic contrasts for Japanese listeners to distinguish (Wang, 2002). Even in cases in which the duration of the study was extended, learners did not make considerable improvement in perception (Yang et al., 2016). Previous researches on

nasal codas of Mandarin Chinese usually focus on the acoustic characteristics and error patterns presented by learners. A practical method to help Japanese listeners to establish the perceptual category for nasal codas is rarely mentioned.

The SLM (Speech Learning Model) posits that development of accurate information about the phonetic properties of L2 speech sounds and the establishment of the phonetic categories requires a substantial amount of input (Flege, 2003). Learners living in the target language country have numerous opportunities to immerse in a native language environment, while it is difficult for learners who are learning foreign languages in their home country to obtain a significant amount of input. With the development of CALL (Computer-Assisted Language Learning) and ICT (Information and Communication Technology), more and more solutions to this issue have been proposed (Dantsuji and Tsubota, 2005). With regard to L2 perceptual training, a High Variability Training Paradigm (HVTP) is widely applied to help learners to identify L2 confusing contrasts. No matter in segmental features or suprasegmental features, the effectiveness of the training has been observed via vast experiments. However, Perrachione (2011) indicates that HVTP enhanced learning only for individuals with strong pretraining aptitude, and for the individuals with low pretraining aptitude, the HVTP may weaken their identification ability. Yang et al. (2018) applied HVTP to train Japanese listeners to identify nasal codas “-n” and “-ng” in Mandarin Chinese. The overall accuracy in the experimental group increased, however, because of the individual difference, some Japanese listen-

ers show no improvement in their perceptual ability. The result implies that the HVTP is not necessarily effective for everyone. Recently, more and more training paradigms have been developed to meet the various needs of users, for example, audio-visual training method, perceptual fading training method, and adaptive training method.

Moreover, most of the current training paradigms stay at the stage of laboratory research. Participants usually have to conduct the training in a set environment. In Yang et al. (2017), participants also had to come to a classroom to attend the training. The software of Yang et al. (2018) was used via a CD-ROM, participants had difficulty in collecting the real-time results they had selected. Web-based applications usually show the final results of the test and training, when the participant wishes to review the selection they have made, it becomes difficult.

This study proposes a novel training paradigm which applies HVTP and self-adaptive training paradigm (SATP) together to a training software. It emphasizes more individual difference on the stimuli presented to the user. Users can only focus on the items that they are not good at. This training software aims to help Japanese listeners to identify nasal codas “-n” and “-ng” in Mandarin Chinese.

## 2 Previous Research

### 2.1 Previous research on nasal coda

There are two types of nasal codas in Mandarin Chinese, “-n” and “-ng”. They are one of the most difficult phonetic contrasts in Mandarin Chinese for Japanese learners. As with nasal codas, they only occur with vowels, and can not make a syllable by themselves. In Mandarin Chinese, two utterances which differ only in nasal codas will be recognized as different utterances. Correspondingly, in Japanese, the mora nasal / N / can occur in the syllable-final position, and the articulation is different depending on the following consonant (Hattori, 1984). In spite of the various articulations due to the following consonant, the meaning of such utterances does not change. As a result, it is difficult for Japanese listeners to identify the two nasal codas.

### 2.2 Previous research on perceptual training paradigms

#### 2.2.1 HVTP

SLM proposes that the phonetic systems used in the production and perception of sound remain adaptive over the lifespan (Flege, 1995). Perceptual training is an excellent method to help L2 learners to establish a good phonetic category of confusing L2 contrasts. The most successful phonetic training paradigm that has been posited may be the HVTP. HVTP directs learners attention towards relevant phonetic cues by presenting them with various stimuli in different phonetic contexts (Aliaga-Garcia and Mora, 2009). The first research that applied HVTP to L2 phonetic training is Logan et al. (1991). This research verified whether perceptual practice involving speech sound variability might help Japanese English learners to discriminate the phonetic contrast / r / and / l /. They applied six native speakers’ voices to the training and emphasized the variability of phonetic contexts (i.e. minimal pairs with / r / or / l / at the position of initial, intervocalic and final). All the stimuli used in training are natural productions other than synthetic stimuli. Japanese English learners first took a pre-test, conducted a training on the contrast, and then took a post-test. The results of the pre-test and post-test show that all the participants obtained improvements without exception.

Lively et al. (1993) and Lively et al. (1994) continued to use the procedure in Logan et al. (1991) and extended the training paradigm. Lively et al. (1993) verified that variability played an important role in perceptual learning and category formation. Lively et al. (1994) set a delayed post-test to verify whether the training effect retained after the training had stopped. Results show that three months later, the accuracy of delayed post-test was still higher than that in the pre-test. After six months without training, participants’ accuracy was still 4.5% above pretest levels. These results show that the training effect retains long-term after perceptual training.

These studies lay the foundation of HVTP and after that various researches began to make attempts to train L2 learners to perceive L2 phonetic contrasts. For example, Lambacher et al. (2005) instructed Japanese participants to identify American

English mid and low vowels. Iverson et al. (2012) trained experienced listeners and less experienced listeners on English vowels via HVTP. Participants can benefit from the training and obtain an improvement on the perceptual ability. Moreover, training also extends to suprasegmental features and proved to be effective, for example, in Mandarin Chinese tones (Wang et al., 1999), and in Japanese pitch accent (Shport, 2016).

HVTP helps listeners to establish good perception by engaging listeners with diverse natural materials that contain a variety of variations by using a variety of stimuli. The stimuli are provided by various talkers in multiple speech contexts. The improvement in perception is also proven to be transferred to production. However, this kind of “one size fits all” training paradigm ignores the individual difference. It is essential to consider the individual difference in pretraining aptitudes when evaluating the efficacy of any perceptual training paradigm. Yang et al. (2018) verified that HVTP is not necessarily competent to every participant. Ingvalson et al. (2014) also indicated that learners will show the most significant benefit from training if training is designed to take advantage of their individual abilities. As a result, more paradigms of training should be considered.

### 2.2.2 SATP

Another training paradigm widely applied is adaptive training paradigm. In L2 perceptual training studies, some researchers call perceptual fading training method as an adaptive training method (Protopapas and Calhoun, 2000; Zhang, 2009; Feng et al., 2014). Perceptual fading training method highlights the acoustic cues of confusing phonetic contrasts via speech synthesis techniques, and trains learners to notice these cues. During the training, they gradually reduce the differences in these acoustic cues, so that learners rely more on the most important cue to identify the contrast. In our study, we follow the definition of adaptive training in Barrett et al. (1977). “Adaptive training is a training technique in which the trainee’s accuracy provides feedback for subsequently changing the nature of the task. If large errors occur, the task is simplified so that the trainee can proceed at a level of difficulty commensurate with his abilities”. Concerning this meaning of adaptive, the application of adap-

tive training is still little discussed. The adaptive trainings in previous studies rarely adjust with the performance of the user. Qian et al. (2018) developed a system that combines utilizing corpus-based word frequency lists, high variability phonetic input, and text-to-speech technology to automatically create discrimination and identification perception exercises customized for individual learners. In their system, an adaptive high variability perceptual training paradigm is applied. The points of their adaptation lie in a) the selection of training tasks: discrimination exercise or identification exercise, b) their adaptive criterion, that is, if the accuracy of one question exceeds 80%, training on this kind of item will not be conducted. If the accuracy of one question is lower than 80%, training will begin and participants have to practice the same kind of item until the accuracy exceeds 80%. However, for the training item, it is not adaptive. Barriuso and Hayes-Harb (2018) point out that the importance of HVTP in the future and insist that “materials and procedures ought to be tailored to aspects of the particular language-learning scenario”. Not only the procedure, but the materials should be adaptive as well.

Barrett et al. (1977) also defined self-adaptive training. Self-adaptive training is a training method in which the trainees could control their own pace of learning by indicating when they felt they had mastered the problem. Wang and Munro (2004) developed a computer-based, learner-centered program for English vowel contrasts learning. Their study did not prescribe a rigid schedule or amount of training for the participants. Instead, they allowed participants to decide the time and the amount they conduct the training. After two months’ training, participants who were assigned to the experimental group showed an increase in the identification of vowels, while participants in the control group showed no improvement. However, sometimes users do not have an accurate judgment of their ability to distinguish the confusing contrasts, leading to situations where they often overestimate or underestimate their abilities (Yang et al., 2018). Users need a relaxed training environment, while some tests to understand their current ability are also essential. Moreover, if a system can automatically identify the items that users are good at and not good at, the users can practice in a targeted manner to improve learn-

ing efficiency.

The current study focuses on the following issue: Can we develop a training software that combines the effective training paradigms to aid L2 learners in perceiving the confusing phonetic contrasts? To answer this question, we developed a training software which includes HVTP and SATP in the software. This software is expected to provide more possibilities for Japanese listeners to learn nasal codas in Mandarin Chinese, especially for those listeners that HVTP has little effect on their perceptual ability.

### 3 Training Software

Yang et al. (2017) and Yang et al. (2018) developed a training software for Japanese listeners to identify “-n” and “-ng” in Mandarin Chinese. The training software runs on the Windows operating system as a GUI program. Users are instructed to listen to a Chinese syllable and identify the final consonant in each syllable. A two-alternative forced-choice identification task is leveraged in order to parallel the identification task in the HVTP. A test mode and a practice mode are set in the software. In the test mode, users are first asked to click the “Play” button, then to listen to a monosyllable which ends either with “-n” or “-ng”, and finally, to select an answer by clicking a button displaying either “-n” or “-ng”. After the answer selection, users proceed to the next test item. In the practice mode, users are also required to click the “Play” button and to select an answer. Immediate feedback to understand whether the sound is “-n” or “-ng” is displayed in characters after participants’ every selection.

In our study, the training software is developed with C# language and also runs on a Windows operating system as a GUI program. The primary interfaces and button functions are similar to Yang et al. (2017). In contrast to their software, our software is installed on a USB flash drive, and users can take them home and conduct the training at any time and anywhere off-line. The USB flash drive also enables the data of each selection made by the user to be recorded and saved timely. There are three modes in the software: a learning mode, a test mode, and a training mode. All the modes only allow users to click buttons, and the sounds in all the modes can be replayed as many times as the users need by means

of clicking the “Play” button.

Since the training software combines HVTP and SATP, we introduce it by paradigm. In HVTP, three important factors should be considered: a talker variability, a stimuli variability, and a phonetic context variability. The voice of two male native speakers and two female native speakers is included in the training software. All the producers of the stimuli are from areas in which nasal codas “-n” and “-ng” are discriminated clearly. Since tone is an essential part of one syllable in Mandarin Chinese, the stimuli used in the software are monosyllables which end with nasal codas in four tones. They were digitally recorded using a TASCAM HD-P2 recorder in a sound-proof room with a sampling rate of 44,100 Hz. Wang (2002) indicates that preceding vowels before nasal coda are important factor that influence the accuracy of identifying nasal codas. We selected nine pairs of nasal codas whose preceding vowel is “a”, “e” or “i” with four tones as training stimuli respectively. As a result, the total number of stimuli is  $4$  (talker)  $\times$   $4$  (tone)  $\times$   $3$  (vowel)  $\times$   $2$  (type: n/ng)  $\times$   $9$  (item) = 864. With regard to the phonetic context variability, since nasal coda only exists in the final position of a syllable, the variability is difficult to consider in a monosyllable environment.

In SATP, in order to make users focus on the items which are especially difficult for them, and reduce the time spent on redundant or unnecessary training, we propose an adaptive training paradigm on training stimuli. The stimuli in the software are divided into four levels: talker, tone, vowel and type (“-n” or “-ng”).  $4$  (talker)  $\times$   $4$  (tone)  $\times$   $3$  (vowel)  $\times$   $2$  (type) = 96 folders are set up to store all the stimuli. Each stimulus has a naming number and is placed in the corresponding folder. When a user clicks the “Test” button on the start page, the system will select one item randomly from each level’s folder and form a 96-question test. The interface of the start page is illustrated in Figure 1.

In the learning mode, correct example sounds are provided for listeners to learn. Example sounds are the most basic monosyllables ending with nasal codas without any preceding consonant. Listeners can compare the monosyllables ending with nasal codas with the same preceding vowel. The number of times that listen to each stimulus is recorded when the button of the item is pressed. The interface of

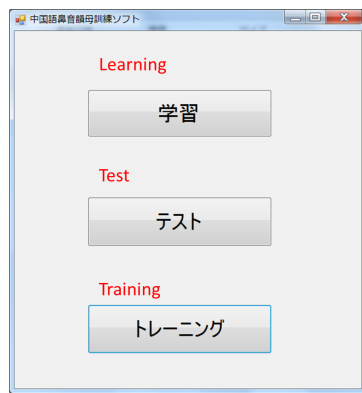


Figure 1. Interface of the start page.



Figure 2. Interface of the learning mode.

learning mode is shown in Figure 2.

In the test mode, listeners are asked to answer a profile test with 96 questions to decide the accuracy of identifying nasal codas in Mandarin Chinese. The accuracy of the test is shown to the participant after the test is completed. The software will record every choice that a participant has made, and calculate the accuracy of each item in talker level, tone level, and vowel level respectively. The Software can pick up the lowest and the second lowest accuracy in each level, and generate the training items in the training mode. The items for training are randomly selected from the lowest and second lowest accuracy folders, that is, from 2 (talker) \* 2 (tone) \* 2 (vowel) \* 2 (type) \* 9 (item) = 144 items. Figure 3 illustrates an example of the design of question selection, a user completes the profile test and the system calculates the accuracy of each level. Concerning the talker level, the accuracy of talker B and talker C are the lower ones. The accuracy of tone 2 and tone 3 are the lower ones in the tone level, and the accuracy of vowel “i” and “e” are the lower ones in the vowel

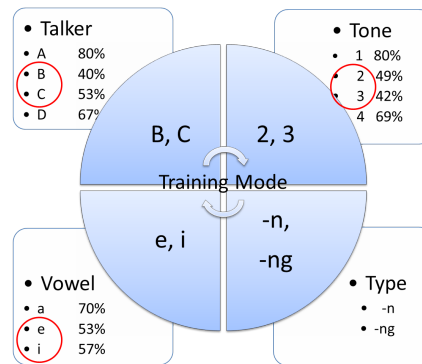


Figure 3. An example of the design of question selection.



Figure 4. Interface of the test mode.

level. Therefore, the stimuli in the training mode are selected from the folders corresponding to talkers B and C, tones 2 and 3, and vowels “i” and “e”. The selection of training stimuli updates when the profile test is conducted. The interface of the test mode is shown in Figure 4.

In the training mode, there are 120 questions for users to practice. Questions are also provided in a random order. Feedback is given after listeners make a selection to each question. After listeners press the button of the choices “-n” or “-ng”, a warning sound was presented to help listeners judge if the listener has made a correct decision immediately. When they choose a correct answer, they hear a crisp sound. Otherwise, they hear a heavy sound, at the same time, the correct answer is also presented in characters in Pinyin. If the user makes a wrong choice, the sound of the stimulus will be played one more time automatically to help the user understand the item better. The interface of the training mode is illustrated in Figure 5.

The naming-number of stimuli and the result of



Figure 5. Interface of the training mode.

the selection of each mode is recorded in csv files respectively. The reaction time and the times of each sound played are also recorded in the files of the test mode and training mode. Users can refer to the results after the training and easily understand which kinds of sounds are their weak items.

#### 4 Conclusion and Future Work

In this paper, we developed a training software which combines HVTP and SATP to aid Japanese Mandarin Chinese learners in identifying nasal codas “-n” and “-ng”. This novel training paradigm retains the primary advantages of HVTP while adding a SATP, which can measure items that users are not good at and form personalized training for each user. This novel training paradigm which provides Japanese Mandarin Chinese learners with a multi-voice training environment can be applied to a pedagogical occasion in and out of the classroom.

Future study should be extended in the following dimensions. More talkers’ voices should be applied to the training software to increase the variability, aiding users in getting accustomed to the voice of various native speakers. The evaluation of the training software should be conducted to verify the effectiveness. Moreover, the comparison with HVTP should be made to observe the difference in efficiency.

#### References

- Athanassios. Protopapas and Barbara. Calhoun. 2000. Adaptive phonetic training for second language learners. *Proceedings of the 2nd International Workshop on Integrating Speech Technology in Language Learning*, 31–38.
- Cristina. Aliaga-Garcia and Joan C. Mora. 2009. Assessing the effects of phonetic training on L2 sound perception and production. *Recent Research in Second Language Phonetics/Phonology: Perception and Production*, Cambridge Scholars Publishing, Newcastle upon Tyne, UK.
- Erin M. Ingvalson, Marc. Ettliger, and Patrick C M. Wong. 2014. Bilingual speech perception and learning: A review of recent trends. *International Journal of Bilingualism*, 18(1):35–47.
- Gerald V. Barrett, James P. Greenawalt, Carl L. Thornton, and Thomas R. Williamson. 1977. Adaptive Training and Individual Differences in Perception. *Perceptual and Motor Skills*, 44:875–880.
- Irina A. Shport. 2016. Training English listeners to identify pitch-accent patterns in Tokyo Japanese. *Studies in Second Language Acquisition*, 38(4):739–769.
- James E. Flege. 1995. Second-language Speech Learning: Theory, Findings, and Problems. *Speech Perception and Linguistic Experience: Issues in Cross-language research, Chapter 8*, York Press, York, UK.
- James E. Flege. 2003. Assessing constraints on second-language segmental production and perception. *Assessing constraints on second-language segmental production and perception*, Mouton de Gruyter, Berlin and New York, USA.
- John S. Logan, Scott E. Lively and David B. Pisoni. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, 89(2):874–886.
- Kuniko. Miyawaki, James J. Jenkins, Winifred. Strange, Alvin M. Liberman, Robert. Verbrugge, and Osamu. Fujimura. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, 18(5):331–340.
- Linjun. Zhang. 2009. Adaptive Perceptual Training and the Categorical Perception of Aspirated/Unaspirated Contrast by Japanese Learners of Chinese. *Chinese teaching in the world*, 23(4):560–566.
- Manman. Qian, Evgeny. Chukharev-Hudilainen, and John. Levis. 2018. A system for adaptive high variability segmental perceptual training: implementation, effectiveness, and transfer. *Language Learning & Technology*, 22(1):69–96.
- Masatake. Dantsuji and Yasushi. Tsubota. 2005. The Effectiveness of CALL in Speech Learning in Second Language Acquisition. *Journal of the Phonetic Society*, 9(2):5–15.
- Paul. Iverson, Melanie. Pinet, and Bronwen G. Evan. 2012. Auditory training for experienced and inexperi-

- enced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(1):145–160.
- Ruining. Yang, Hiroaki. Nanjo, and Masatake. Danstuji. 2016. The Perceptual Characteristics of Mandarin Chinese Nasal Codas by Japanese Speakers -with the condition of preceding consonants are plosive consonants-. *Journal of the Japan Society for Speech Science*, 17:115–132.(in Japanese)
- Ruining. Yang, Hiroaki. Nanjo, and Masatake. Danstuji. 2017. Development of ICT Educational Software for Perceptual Training of Mandarin Chinese Nasal Codas. *11th annual International Technology, Education and Development Conference,INTED2017 Proceedings*, 2664–2668.
- Ruining. Yang, Hiroaki. Nanjo, and Masatake. Danstuji. 2018. Training Japanese Speakers to Identify Nasal Codas of Mandarin Chinese. *Journal of Language Teaching and Research*, 9(1):7–15.
- Scott E. Lively, John S. Logan, and David B. Pisoni. 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(3):1242–1255.
- Scott E. Lively, David B. Pisoni, Reiko A. Yamada, Yohichi. Tohkura, and Tsuneo. Yamada. 1994. Training Japanese listeners to identify English /r/ and /l/. III. Longterm retention of new phonetic categories. *The Journal of the Acoustical Society of America*, 96(4):2076–2087.
- Shiro. Hattori. 1984. *Phonetics*, Iwanami shoten, Tokyo, Japan.(in Japanese)
- Stephen G. Lambacher, William L. Martens, Kazuhiko. Kakehi, Chandrajith A. Marasinghe, and Garry. Molholt. 2005. The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26(2):227–247.
- Taylor A. Barriuso and Rachel. Hayes-Harb. 2018. High Variability Phonetic Training as a Bridge From Research to Practice. *The CATESOL Journal*, 30(1):177–194.
- Tyler K. Perrachione, Jiyeon. Lee, Louisa Y. Y. Ha, and Patrick C. M. Wong. 2011. Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, 130(1):461–472.
- Xinchun. Wang and Murray J. Munro. 2004. Computer-based training for learning English vowel contrasts. *System*, 32(4):539–552.
- Xiaoli. Feng, Yue. Sun, Jinsong. Zhang, and Yanlu. Xie. 2014. A Study on the long-term retention effects of Japanese C2L learners to distinguish Mandarin Chinese Tone 2 and Tone 3 after perceptual training. *Proceedings of the 9th International Symposium on Chinese Spoken Language Processing, ISCSLP 2014*, 599–603.
- Yue. Wang, Michelle M. Spence, Allard. Jongman, and Joan A. Sereno. 1999. Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, 106(6):3649–3658.
- Yunjia. Wang. 2002. An experimental study on the perception and production of nasal codas by Japanese Learners of Chinese Putonghua. *Chinese teaching in the world*, 60:47–60.