Journal of Phonetics 66 (2018) 242–251

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

High variability identification and discrimination training for Japanese speakers learning English /r/–/l/ ☆

Yasuaki Shinohara [a,*], Paul Iverson [b]

[a] Waseda University, 1-6-1, Nishi-waseda, Shinjuku-ku, Tokyo 169-8050, Japan
[b] University College London, Department of Speech, Hearing and Phonetic Sciences, 2 Wakefield Street, London WC1N 1PF, United Kingdom

ARTICLE INFO

Article history:
Received 15 March 2017
Received in revised form 15 November 2017
Accepted 16 November 2017
Available online 7 December 2017

Keywords:
Speech perception
Speech production
Second language
Phonetic training

ABSTRACT

Second-language (L2) learners can benefit from exposure to phonetically variable speech during computer-based training. Moreover, this training can be effective even for L2 learners who have extensive exposure to their L2 in daily life, suggesting that there is something specific about the training task that aids learning. The present study compared traditional identification training with discrimination training to evaluate whether discrimination training could be effective, and whether different types of focused attention (i.e., on categorization vs. perceptual differences) could combine to provide a greater increase in learning. Adult Japanese speakers were given 10 sessions of identification and discrimination training, with pre/mid/post tests of identification, auditory discrimination, category discrimination, and /r/–/l/ production. The results demonstrated that both identification and discrimination training increased accuracy of Japanese speakers' perception and production of English /r/–/l/ in similar ways, but that there was little added benefit to using the two training methods in combination. It thus appears that identification and discrimination training have similar effects in second-language learners, as long as both training methods incorporate high variability.

1. Introduction

It is well established that learning second-language (L2) phonetic contrasts in adulthood can be difficult, particularly when the phonetic contrasts mismatch the processing and representations that have been developed for one's first language (L1; e.g., Best, 1995; Flege, 1995; Iverson & Kuhl, 1995). That being said, learning is possible for difficult contrasts after many years of real-world experience (Flege, Takagi, & Mann, 1996; Ingvalson, McClelland, & Holt, 2011; MacKain, Best, & Strange, 1981; Takagi & Mann, 1995) or intensive exposure to naturalistic variability in computer-based phonetic training programs (e.g., Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991). One could assume that high-variability phonetic training works because it simulates the kind of exposure to phonetically variable speech that individuals receive during real-life experience. However, even the earliest studies hypothesized that there were specific aspects of the high-variability training methods, such as the opportunity for focused attention, that likely affected learning (e.g., Logan et al., 1991). Subsequent work has demonstrated that phonetic training improves perception even for experienced L2 speakers who use their L2 daily (e.g., Iverson, Pinet, & Evans, 2012), which would only happen if phonetic training made a unique contribution to the learning process beyond exposure to phonetically variable speech.

The exact contribution of phonetic training to L2 learning is still unclear, but it tends to have more focused effects than real-world learning. For example, individual differences in perception and production of L2 contrasts tend to be highly correlated among language learners (i.e., better perceivers tend to be better speakers), but phonetic training often improves production and perception more idiosyncratically, with low individual-differences correlations in improvement (e.g., Bradlow, Pisoni, Yamada, & Tohkura, 1997; Bradlow, Akahane-Yamada, Pisoni, and Tohkura, 1999; Huensch & Tremblay, 2015; Iverson et al., 2012); phonetic training thus seems to produce less global improvements in learning. Sadakata and McQueen (2013) have suggested that high-variability phonetic

☆ A portion of this work was presented in "Perceptual training effects on production of English /r/–/l/ by Japanese speakers" at the Phonetics Teaching and Learning Conference, London, UK, 8–10 August 2013.

* Corresponding author.
E-mail address: y.shinohara@waseda.jp (Y. Shinohara).

https://doi.org/10.1016/j.wocn.2017.11.002
0095-4470/© 2017 The Authors. Published by Elsevier Ltd.

training specifically improves abstract category representations rather than tuning auditory processing for speech. Iverson, Hazan, and Bannister (2005) have likewise argued that high-variability phonetic training does not change lower-level processing, but have further suggested that training improves how automatically listeners can apply their existing category knowledge and processing to real speech, rather than fundamentally changing category representations (e.g., cue weightings).

The aim of the present study was to produce larger improvements in L2 phonetic processing by combining different training methods (i.e., identification and discrimination) that may affect different underlying processes involved in phonetic perception and production. For the most part, improvements after standard high-variability phonetic identification training, particularly for Japanese adults learning to distinguish English /r/–/l/, have been limited to about 15 percentage points, suggesting that improvements due to this method alone can reach a ceiling and that other processes which underlie L2 phonetic perception and production need to be improved (Bradlow et al., 1999, 1997; Iverson et al., 2005; Lively et al., 1993, 1994; Logan et al., 1991; MacKain et al., 1981). Beyond searching for practical improvements in learning, we aimed to compare discrimination and identification methods to examine to what extent the different types of focused attention required during these tasks are able to improve different aspects of L2 phonetic processing.

Previous attempts at discrimination training have had limited success. For example, Strange and Dittmann (1984) gave Japanese adults same-different discrimination training along a synthesized *rock-lock* continuum. The results demonstrated that discrimination training improved identification and discrimination at the /r/–/l/ category boundary for these synthetic stimuli, but this improvement did not generalize to novel stimuli or natural recordings. In contrast, subsequent work by Logan et al. (1991) found that generalization could be obtained from identification training that used natural recordings of minimal-pair words spoken by multiple talkers. Logan et al. hypothesized that their technique was more effective because it allowed listeners to form robust categories that extend to varied phonetic contexts, whereas discrimination training primarily increases sensory resolution overall, including to within-category variation that likely interferes with categorization. However, there were many differences between the techniques of Strange and Dittmann (1984) and Logan et al. (1991), particularly related to the variability of the stimuli, making it unclear to what extent the differences in the findings were due to discrimination vs. identification training. If discrimination training was able to alter sensitivity to acoustic variation in a targeted way (e.g., improving the primary acoustic cue sensitivity at category boundaries without raising within-category sensitivity), then this might prove to be an effective supplement to identification training.

The English /r/–/l/ categories are primarily distinguished by L1 speakers in terms of F3 frequency, with a lower F3 frequency for /r/ and a higher frequency for /l/, along with secondary cues such as closure duration (shorter for /r/ and longer for /l/) and transition duration (longer for /r/ and shorter for /l/), and variation in F2 frequency related to dark/light articulation that is mostly orthogonal to the /r/–/l/ contrast (e.g., Hattori & Iverson, 2009; Underbakke, Pola, Gottfried, & Strange, 1988; Yuan & Liberman, 2011). It has been claimed that Japanese adults have difficulty learning this distinction because both English /r/ and /l/ are assimilated into their L1 tap category (Best & Strange, 1992), or because their L1 tap is similar enough to English /l/ to block the formation of a new English /l/ category (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004). However, Hattori and Iverson (2009) found that individual differences in L1–L2 category assimilation, or the distance between these categories, is poorly correlated with identification accuracy. Moreover, they found that Japanese adults are able to learn secondary cues for /r/ and /l/, but have a highly specific difficulty with learning F3. Ingvalson et al. (2011) similarly demonstrated a positive correlation between F3 cue weighting and /r/–/l/ identification, but also showed that F3 reliance is not changed as a function of length of residence in an English-speaking country, age of arrival in an English-speaking country, the amount of Japanese use, or length of student status in an English environment (English education). It has been argued that this difficulty with F3 primarily stems from a pre-categorical level of phonetic processing (Iverson et al., 2003; Iverson, Wagner, & Rosen, 2016). That is, Japanese adults tend to be less sensitive to F3 variation near the English /r/–/l/ category boundary and more sensitive to irrelevant variation, and it is this perceptual warping of the cue variance at an early level that affects how the cues are represented at a more abstract categorical level. It would thus be desirable if this earlier level of processing could be altered through training, rather than only using identification training to improve more abstract representations.

In the present study, Japanese speakers were trained on both identification (ID) and discrimination (DIS). The training for each method lasted for five sessions, and the order of the training programs was balanced across subjects so that half of the subjects performed identification training first (ID-DIS) and the other half performed discrimination training first (DIS-ID). Identification training, which was intended to improve category representations (Gerrits & Schouten, 2004; Logan et al., 1991; Sadakata & McQueen, 2013; Sjerps, McQueen, & Mitterer, 2013), used a standard high-variability technique from a previous study (Iverson et al., 2005). The discrimination training, which was intended to improve pre-categorical processing (Gerrits & Schouten, 2004; Logan et al., 1991; Sadakata & McQueen, 2013; Sjerps et al., 2013; Strange & Dittmann, 1984), used 20 stimulus continua based on signal-processed natural stimuli (four contrasts each from five talkers), as well as fully natural stimuli. Listeners performed three types of three-interval oddity tasks: auditory discrimination with natural stimuli (i.e., two natural stimuli that were identical and one that was different, where the acoustic differences were uncontrolled), auditory discrimination with signal-processed stimuli (where the different stimulus was specifically different in terms of F3), and category discrimination (three natural stimuli that were all acoustically different, but one started with a different phoneme). It was intended that the auditory discrimination tasks would improve F3 perception at the boundary in a way that would generalize to other stimuli, and that category discrimination would additionally decrease sensitivity to irrelevant acoustic variation (e.g., Flege, 2003; Gerrits & Schouten, 2004; Højen & Flege, 2006; Iverson et al., 2003, 2012; Logan et al., 1991; Sadakata & McQueen, 2013; Sjerps et al., 2013).