Journal of Phonetics 54 (2016) 151–168

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

# Imitation of second language sounds in relation to L2 perception and production

Yen-Chen Hao [a,*], Kenneth de Jong [b]

[a] Department of Modern Foreign Languages and Literatures, University of Tennessee Knoxville, TN, USA
[b] Department of Linguistics, Indiana University Bloomington, IN, USA

ARTICLE INFO

*Article history:*
Received 9 December 2014
Received in revised form
28 October 2015
Accepted 29 October 2015

*Keywords:*
Second language imitation
Perception
Production

ABSTRACT

This study reports findings from two experiments on second language learners, comparing their performance in an Imitation task to that in Identification and Read-Aloud tasks. Experiment 1 targeted English speakers' learning of Mandarin tones, while Experiment 2 investigated Korean speakers' learning of English consonants. The results of Experiment 1 showed that the Imitation task was generally easier for English speakers than the Identification and Read-Aloud tasks, suggesting that imitation was performed without some of the skills required by the other two tasks. As for Experiment 2, the Koreans were consistently less accurate in Imitation than in Read-Aloud, while their Imitation was more accurate than their Identification when the L2 sounds have a close counterpart in Korean. The results from both experiments revealed that the accuracy in Imitation was not always constrained by that in the Identification and Read-Aloud tasks. Hence it can be inferred that L2 imitation may not involve all the skills required by the other two tasks and probably bypasses some aspect of phonological encoding. More detailed predictions of the error patterns in the Imitation task based on Perception, Production, and Cascade models were compared. It was found that English speakers' confusion patterns in imitating Mandarin tones correlated to the same degree with the predictions of all three models, which may be because the learners' difficulty in perception and production was largely similar. On the other hand, Korean learners' errors in imitating English consonants were overall more accurately predicted by a Cascade model, suggesting that both perceptual and production imprecision aggravated imitation performance. The comparison of these two experiments corroborates that various factors may affect the relationships between L2 imitation, perception, and production.



## 1. Introduction

### 1.1. L1 imitation

One distinctive aspect of speech and language is that perceivers are also normally producers. In regular communication speakers and listeners are constantly changing their roles and they act cooperatively, with the speakers trying to be understood by the listeners and the listeners trying to understand the speakers. Even though speaking and listening obviously require very different skills, speech perception and production must be connected in a manner that permits congruency between these two modalities. To explore the link between these two domains, many studies have employed a speech imitation task, in which the performers listen to an audio stimulus and reproduce it (Flege & Eefting, 1988; Fowler, Brown, Sabadini, & Weihing, 2003; Mitterer & Ernestus, 2008; Mitterer & Müsseler, 2013; Shockley, Sabadini, & Fowler, 2004). By comparing the target stimulus and the performers' reproduction, these studies have attempted to conceive a processing route from speech perception to production and establish the relationships between them.

One main question concerning the processing mechanisms involved in speech imitation is whether audio perception and motor production are mediated by higher-level phonological encoding. If so, the reproduction should conform to the participants' linguistic categories, and sub-phonemic differences in the stimuli should be filtered out in the phonological module and should not be imitated.

* Corresponding author. Tel.: +1 865 974 7162.
*E-mail address:* yhao6@utk.edu (Y.-C. Hao).

0095-4470/$ - see front matter © 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2015.10.003

If, on the other hand, speech imitation bypasses phonological categorization and operates on a more direct link from perception to production, participants should be able to faithfully imitate the sub-phonemic details of the audio stimuli.

A number of researchers supported the latter proposition by showing that their participants could imitate fine phonetic differences that are not contrastive in their native language. For example, Fowler et al. (2003) and Shockley et al. (2004) found that English speakers lengthened the VOT of their stop consonants when imitating stimuli with artificially-lengthened VOT values. Lehiste and Shockey (1980) showed that English speakers' repetition of a synthesized *bead-beat* continuum closely mimicked the vowel duration increment in the stimuli, in contrast to the strong categorical trend in their labeling of the continuum. Other studies such as Chistovich, Fant, deSerpa-Leitfio, and Tjernlund (1966) and Repp and Williams (1985) also provided partial evidence that imitation is not mediated by phonological categorization, finding categories in their imitation that did not correspond to any of their linguistic categories. These findings, together, suggest that the perceived signal is probably not categorized into a phonological prototype before it is reproduced, because sub-phonemic differences in the stimuli are often preserved in the imitation.

In contrast, another group of researchers have shown that their participants imitated the properties that are phonologically relevant in their native language, but not necessarily the fine phonetic details. For instance, Mitterer and Ernestus (2008) examined the imitation of stop consonants with 0, −38, and −64 ms of VOT by native speakers of Dutch, in which only the presence or absence of pre-voicing is phonologically contrastive but not the amount of pre-voicing. They found that the Dutch participants only imitated the VOT difference between the 0 vs. −38 ms contrast but not between the −38 vs. −64 ms contrast, and thus the authors regarded this as evidence for the mediation of phonological categorization in imitation. In addition, studies on the imitation of English intonation have also lent support for the involvement of phonological encoding in the process. For instance, Pierrehumbert and Steele (1989) examined English speakers' imitation of a short sentence varying in the $F_0$ peak delay between the L+H* and L*+H intonation patterns. The participants' reproductions were mostly bimodal rather than continuous in terms of the peak timings, suggesting that their production patterns were strongly affected by a categorization process associated with the existence of phonologically specific intonational categories. Similar results for vowel categories have been found by Viechnicki (2002) as well.

The discrepancy in the findings of these imitation studies may not be surprising, since they differed considerably in the types of contrast, stimuli complexity, and task procedures. What is clear, though, is that researchers have been investigating the involvement of phonological categorization in speech imitation with various methods, yet have not reached a definite conclusion.

Another question often pursued in the literature is how speech imitation relates to auditory perception and motor production. From one perspective, imitation can be seen as a production task adopting auditory instead of orthographic prompts; however, one might equally view imitation as a perception task utilizing a verbal rather than a written response. Hence the comparison of an imitation task with a perception and a production task would offer insight on the relative contribution of perceptual and production factors in imitation. Some studies have suggested that the participants' imitation patterns largely resemble those in a perception task. For example, Schouten (1977) found that his participants' formant frequency clusters in a vowel imitation task closely corresponded to their perceptual categories in an identification task. In another study, Jia, Strange, Wu, Collado, and Guan (2006) showed that their participants' accuracy in a vowel imitation task and that in an identification task significantly correlated. These findings demonstrated that the participants' imitation performance is very similar to their perceptual identification. Other studies, on the other hand, showed a high degree of resemblance between an imitation and a production task. For instance, Repp and Williams (1985, 1987) found that their participants' imitation of an /i/-/æ/ vowel continuum displayed a few clusters of formant frequencies, which corresponded closely to the same participants' production of the vowel continuum without an audio model. A common limitation of these former studies, however, is that they only compared imitation to either a perception or a production task, but not to both. Hence they may not offer a complete view of the relationships between imitation, perception, and production.

In summary, at least two major questions remain unresolved in the literature on speech imitation: (1) whether phonological encoding is employed in the process of imitation, and (2) whether perception or production exerts a stronger influence on imitation. The present study aims to further investigate these two issues on a different population from the studies reviewed above: second language (L2) learners.

### 1.2. L2 imitation

L2 imitation merits thorough investigation because it not only permits an examination of both L2 perception and production but also illustrates how these two modalities are coordinated in a single act. Additionally, since L2 perception and production have often been found to diverge from each other (e.g. Baker & Trofimovich, 2006; Bohn & Flege, 1997; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Flege, 1993; Flege, Bohn, & Jang, 1997; Ingram & Park, 1997; Sheldon & Strange, 1982), a comparison of second language learners' imitation with their perception and production patterns would reveal more clearly the relative contribution of perception, production, and other processing mechanisms in speech imitation. The studies done by Flege and Eefting (1987, 1988) are among the few that closely examined L2 imitation in relation to L2 perception and production. The authors recruited Spanish speakers who started learning English before the age of 6 to imitate and label a /da/-/ta/ continuum ranging from −60 to 90 ms VOT. In addition, these learners were asked to read a list of English words starting with /t/ and /d/. The results showed that these learners did not faithfully imitate the VOT increment in the stimuli but instead produced three response concentrations indicative of the lead, short-lag, and long-lag stop categories. Hence the authors interpreted these results as support for the involvement of phonological encoding in L2 imitation, and concluded that these Spanish speakers have established a new phonetic category for the English long-lag stop.