Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 69 (2015) 17–30

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

A quantitative model of first language influence in second language consonant learning

Jian Gong [a,b,*], Martin Cooke [c,b], Maria Luisa García Lecumberri [b]

[a] School of Foreign Languages, Jiangsu University of Science and Technology, 2 Mengxi Road, 212003 Zhenjiang, China
[b] Language and Speech Laboratory, Universidad del País Vasco, Paseo de la Universidad 5, 01006 Vitoria, Spain
[c] Ikerbasque (Basque Science Foundation), Spain

Received 2 July 2013; received in revised form 13 January 2015; accepted 5 February 2015
Available online 14 February 2015

Abstract

Theoretical models argue that listeners' perception of second language sounds is heavily influenced by their native language phonology, a prediction borne out by behavioural studies. However, we lack quantitative models capable of making more precise predictions of the way in which the first and second language sound systems interact. The current study introduces a computational modelling framework that permits comparison of different second language learning strategies which vary both in the degree of first language influence as well as in the manner in which second language input is combined with existing first language knowledge. Six different model variants were evaluated by comparison with behavioural data on a task involving the identification of intervocalic consonants of Castilian Spanish by Mandarin Chinese listeners. All approaches demonstrated a similar pattern of rapid improvement with exposure to that observed in listeners. However, approaches that made use of independent first and second language models made the best predictions. An approach that excluded first language influence both predicted lower listener identification levels in the initial stages of learning and higher scores in later stages, demonstrating that first language experience helps to bootstrap second language sound learning but ultimately hinders identification. However, modelling outcomes also demonstrate that no single approach can account for the identification patterns for all consonants, suggesting that learners deploy different approaches to the learning of individual sounds.

Keywords: Foreign language sound acquisition; Consonant identification; Rapid learning; Computer model

1. Introduction

Acquiring the sounds of a new language as an adult is different in at least one crucial respect from the linguistic situation that confronts us as infants: as adults, we already possess a well-developed phonological system. Investigation of how the first (L1) and second (L2) language sound systems interact to influence the development of phonetic competence in a second language is an active area of study, and the degree to which adult learners benefit – or do not – from prior experience remains an important issue both in broadening our understanding of second language learning and in the study of general phonological representations and processes.

Besides the age of acquisition, the influence of the first language is probably the single strongest factor in non-native sound acquisition. Theoretical models (e.g., Kuhl, 1993; Best, 1995; Flege, 1995) agree that perceived similarities between native and non-native phonetic categories play a crucial role in non-native sound perception. Learners may process non-native sounds in terms of their L1

* Corresponding author at: School of Foreign Languages, Jiangsu University of Science and Technology, 2 Mengxi Road, 212003 Zhenjiang, China.

E-mail addresses: habaogj@hotmail.com (J. Gong), m.cooke@ikerbasque.org (M. Cooke), garcia.lecumberri@ehu.es (M.L. García Lecumberri).

http://dx.doi.org/10.1016/j.specom.2015.02.003
0167-6393/

categories given sufficient similarity between the two in a process variously known as 'equivalence classification' (Flege, 1995) or 'perceptual assimilation' (Best, 1995), which can prevent category formation for non-native sounds. Best and Tyler (2007) suggest that L2 learning is a process of fine-tuning and possibly 're-phonologizing', in which learners might stretch or modify their existing L1 categories to accommodate L2 perception (see also Bundgaard-Nielsen et al., 2011).

While theoretical models have provided valuable hypotheses and insights concerning the influence of the L1 phonological system on the development of non-native sound perception and production, what these models lack is the ability to predict in a quantitative manner how the established L1 system and the evolving L2 system interact. For instance, it is difficult to answer questions about the nature of the relationship between the amount of exposure to second language sounds and the rate at which L2 category identification improves.

Computational simulations provide a useful adjunct to theoretical models, and have been used in studies of L1 sound acquisition to show how infants learn native vowel categories and to simulate the perceptual magnet effect (de Boer and Kuhl, 2003; Vallabha et al., 2007; Lake et al., 2009; McMurray et al., 2009). Statistical pattern recognition approaches have also been applied recently to simulate human perceptual assimilation tasks in order to measure cross-language category similarities directly from the acoustic data, providing quantitative formal evaluation of the predictions of theoretical models (Strange et al., 2004; Morrison, 2009; Thomson et al., 2009; Gong et al., 2010). Computational approaches have also been used in studies of the development of L2 perception. For example, in Escudero et al. (2007), machine learning and computational linguistic models were used to simulate and visualise the evolution of learners' L2 vowel spaces. Hidden Markov modelling (HMM) techniques were adopted by Gong et al. (2011) in a modelling study investigating the effect of different ratios of L1 and L2 exposure in identifying second language. These studies not only demonstrated the promise of using computational approaches to model the L2 learning process, but also highlighted a key advantage of simulation, viz. the ability to contrast and evaluate competing models while maintaining control over factors such as the degree and type of exposure. However, simulation studies to date have been somewhat limited in scope. Many existing models have been constructed using either synthetic speech (Escudero et al., 2007) or simplified and abstract speech parameters (e.g., F1/F2 values or VOTs) (de Boer and Kuhl, 2003; Vallabha et al., 2007; Strange et al., 2004; Thomson et al., 2009) while small subsets of vowels or consonants have usually been chosen as the modelling targets (de Boer and Kuhl, 2003; Vallabha et al., 2007; Morrison, 2009; Escudero et al., 2007).

The current study addresses the issue of how L1 knowledge interacts with L2 exposure at the outset of second language learning. We do so by evaluating how closely a number of computational models predict findings from a recent study of non-native consonant identification (Gong, 2013). In that study (reviewed in Section 2 below) Chinese listeners with no experience in Spanish took part in an intensive high-variability perceptual training programme of the kind found to be effective in earlier studies (e.g., Logan et al., 1991; Lively et al., 1993; Bradlow et al., 1997). Listeners were required to identify Spanish consonants drawn from the full consonant inventory, when presented in intervocalic context. Using Gong (2013) as the behavioural reference, in the current study we apply computational modelling to investigate the development of second language learning when confronted by an extensive L2 sound inventory. Our models differ in the manner in which speech material in the L1 and L2 interact during learning and consonant recognition. These models use precisely the same training data, in the same sequence, as made available to listeners in the behavioural study. As such, listeners and models had identical exposure to the consonants of the target language during training. As in Gong et al. (2011), HMMs were used to represent sound categories. An initial HMM set was trained using Chinese data and consonant categories, and subsequently retrained for Spanish consonant categories based on listeners' assimilations of Spanish sounds.

One model – Blend – is motivated by the hypothesis that it is the raw amount of L2 exposure that is the key determinant of listeners' identification of L2 sounds. Blend operates by adding in progressively larger quantities of L2 stimuli and re-learning HMM parameters *ab initio*. A second model, Adapt, is based on the idea that it is not just quantity but the sequence of exposure to new sounds that matters. Rather than re-training at each stage of learning, HMM parameters are adapted using Bayesian speaker adaptation techniques. A further model, Separate, represents the hypothesis that learners of a new sound system are capable of maintaining separate L1 and L2 representations and that they use only the latter to identify L2 speech sounds. We additionally evaluate versions of each of the Blend, Adapt and Separate approaches in which HMMs for the L1 are activated in parallel. These models – Par-blend Par-adapt and Par-separate – represent L1/L2 interaction at the level of categories.

Section 2 reviews the behavioural study of Gong (2013) and describes the Spanish and Chinese speech materials used in the current study. Section 3 describes how listener assimilation results inform the development of the initial model set, while the six modelling approaches are explained in Section 4. Simulation results are presented and compared to listeners in Section 5.

2. Behavioural study

The stimuli and human baselines used in the current study are described in Gong (2013). Here, we review the tasks, speech materials, and outcomes of that study.