A COMPUTATIONAL MODELLING APPROACH TO THE DEVELOPMENT OF L2 SOUND ACQUISITION

Jian Gong[a], *Martin Cooke*[a,b] *& Maria Luisa Garc ú Lecumberri*[a]

[a]Language and Speech Laboratory, University of the Basque Country, Spain; [b]Ikerbasque, Basque
habaogj@hotmail.com; m.cooke@ikerbasque.org; garcia.lecumberri@ehu.es

ABSTRACT

A computational framework for quantitative studies of sound acquisition in a second language is presented. The framework supports the exploration of issues such as the effect of the amount and type of exposure to L2 categories, interactions between L1 and L2 sound systems as well as L1 attrition. The model is illustrated by simulating a Chinese listener's responses to English vowel-consonant-vowel exemplars. The simulations suggest that even very small quantities of L2 material can lead to rapid improvements in recognition of L2 target consonants and that with balanced amounts of exposure to the two languages, L2 models can contribute to the recognition of L1 tokens.

Keywords: computer model, L2 acquisition development, L2 exposure, attrition

1. INTRODUCTION

The study of L2 sound perception and its development is largely based on qualitative comparisons between the L1 and the L2 sound systems to decide whether two sounds are similar, different or equal [5] using criteria such as listeners' identifications, phonetic symbols and, in some cases, acoustic distances. Nevertheless, some attempts have been carried out to quantify the degree of perceived similarity between L1 and L2 sounds, for instance by means of listeners' goodness ratings [1, 2, 9] or multi-dimensional scaling techniques [9].

Recently, L2 sound acquisition has been tackled using statistical and computational modelling techniques. Statistical pattern recognition models have been used to compare the similarity between Chinese and English vowels [10]. Automatic speech recognition and information theory techniques were used in [7] to measure the distance between Chinese and English consonants. Machine learning and computational linguistic techniques have been employed to simulate native Spanish learners' Dutch vowel space development [4]. These studies suggest that quantitative methods can complement theoretical models by improving the precision of their predictions.

The purpose of the current study is to extend the domain of quantitative approaches to issues which have received little attention from modellers, in particular, issues related to the quantity of input, which has been estimated subjectively because of practical or ethical limitations [6]. These include (i) the degree of L2 exposure; (ii) the amount of L2 exposure relative to the L1; (iii) the interaction between new data and the existing L1 sound system; and (iv) L1 attrition following L2 learning [11]. These issues are susceptible to a quantitative approach as they naturally involve continuous changes in the amount of data to which the model has access during its simulation of development. Also, by adjusting the relative balance of L1 and L2 data employed during learning, both adult second language learners and balanced bilingual learning can be accommodated within the same approach.

This kind of quantitative approach aims to tackle questions such as: how little L2 exposure is needed to produce improvements in L2 sound identification, and at what point does further exposure result in little benefit? How does the reduced quantity of data in each language due to exposure to two languages affect performance? What is the effect of merging L2 data into existing L1 categories for similar/near-identical sounds? At what ratio of L2:L1 input do we observe L1 attrition?

Our long term goal is to build a computational account of the development of L2 consonant acquisition which can be used to make testable predictions of a learner's L2 confusions and their possible resolution at different stages of acquisition. In the current study, we describe a computational framework which lends itself to an exploration of the issues outlined above and show results from simulations of Chinese learners of English vowel-consonant-vowel (VCV) tokens.

2. METHODS AND MATERIALS

2.1. Modelling framework

The development of phonemic categories was simulated using Hidden Markov Models (HMM), which have many benefits in quantitative simulations. They use powerful statistical techniques for learning from data and have been optimised for speech recognition; they capture both spectral and temporal features of sounds; and they possess great flexibility in model formation and use, permitting an exploration of L2 material incorporated at the level of existing or new models. Continuous density HMMs were trained using the HTK toolkit [12] on standard speech parameters (a 39-component vector consisting of 12 Mel-frequency cepstral coefficients plus energy, and their first and second temporal derivatives) computed every 10 ms. Individual vowels and consonants were modelled as 3-state HMMs and combined during recognition into VCVs. Within each state, a mixture of Gaussian distributions represented speech observations deemed to belong to that state. A limit of 4 mixture components was determined as the best tradeoff between model accuracy and use of the available training data.

2.2. Corpora

VCV tokens used for model training and testing were derived from English and Chinese corpora collected for the Interspeech 2008 Consonant Challenge [3] and a recent modelling study [7] respectively. Each corpus contains exemplars from many talkers of the 24 English and Chinese consonants in vowel contexts based on combinations of the vowels /æ iː uː/. Here we focus on the 16 English consonants that have close counterparts in Chinese (English /p b t d k ɡ f s ʃ h m n l r j w/; Chinese /pʰ p tʰ t kʰ k f s ʃ̺ x m n l ɹ̺ j w/), since the same label can be used to assess ‘correct’ category recognition. The remaining consonants in each language (English /ʧ ʤ v θ ð z ʒ ŋ/; Chinese /tsʰ ts ʧ̺ʰ ʧ̺ ʨʰ ʨ ɕ ŋ/) were also used to allow for confusions among the full set of categories.

2.3. Modelling strategies

All simulations presented here explored a range of learners differing in the ratio of L1 to L2 tokens used in model development. Consequently, Chinese and English monolinguals occupy ends of the continuum and other points along the continuum represent L2 learners or bilinguals.

To model the effect of absolute exposure to L2 material during acquisition, differing numbers of VCV tokens (25, 50 and 100 exemplars per consonant) were employed during training. These quantities represent the *maximum* number of tokens available to learners, which occurs when the ratio of L1:L2 tokens is 1:1. At other points along the continuum, the number of L2 tokens is smaller. For example, at a ratio of 5:1, only 5 examples of each VCV in the L2 are used.

We model the situation where an adult learner with a fully developed L1 sound system is involved in the process of L2 sound acquisition. We also model the alternative situation where a learner is acquiring both L1 and L2 at the same time. Here, the total quantity of tokens available at each point along the L1:L2 ratio continuum is fixed e.g. assuming 25 exemplars per VCV are available, 5 L2 tokens now corresponds to a ratio of 4:1 as opposed to 5:1 in the former case.

The issue of how L2 material interacts with the existing L1 sound system is investigated here by comparing two strategies: direct blending of L2 data into L1 categories, versus development of separate L2 category models. The first approach is similar to [4], where the L2 speech material (Dutch) was mixed with L1 (Spanish) data during model training, to simulate an advanced learner’s L2 exposure. The second approach maintains separate models for L1 and L2 categories, even for sounds that are close counterparts. Finally, to explore any effects of L1 attrition, models trained with L1 and L2 data are tested using L1 data.

3. RESULTS

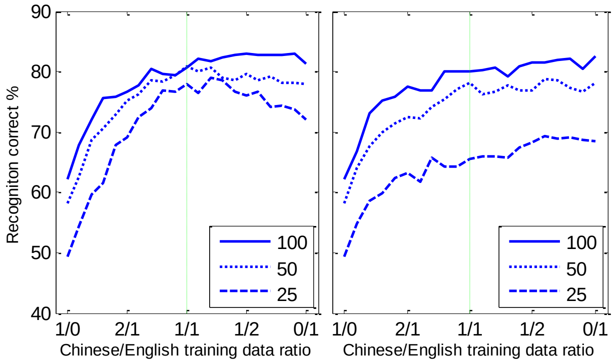
3.1. Effect of degree of L2 exposure

In this section the effect of blending L2 data into existing L1 models is analysed as a function of the quantity of L2 exposure, the ratio of L1:L2 experience, and the additive or fixed overall amount of training tokens. In all cases, the quantity reported is the percentage of L2 categories correctly identified for the 16 consonants having a close counterpart in the L1.

Figure 1 (left) depicts category identification scores for L2 (English) test tokens in the situation where L2 tokens are added to a fixed L1 training set. The leftmost part of the x-axis corresponds to a monolingual Chinese model while the rightmost part represents a monolingual English listener. As expected, identification scores are low for the Chinese monolingual model for English test tokens, though well above chance. As English tokens are added to the training set, performance initially increases rapidly, reaching a slower asymptotic growth towards native-like levels. Increasing the absolute quantity of exposure from 25 to 50 to 100 tokens per consonant leads to increased scores. The tendency to asymptote occurs at an earlier L1:L2 ratio for larger numbers of training tokens, suggesting that the absolute amount of exposure as opposed to the relative L1/L2 exposure is most important.

Figure 1 (right) shows equivalent results for the situation where the overall amount of L1+L2 data is fixed. Here, the pattern of improvement is more gradual compared to the case where the L2 material is additional to a fixed L1 training set. Further, the L1:L2 ratio at which the slower asymptote is reached is similar irrespective of the absolute quantity of training data.

Figure 1: Category identification scores for added L2 tokens (left) and fixed overall L1+L2 token quantity (right).

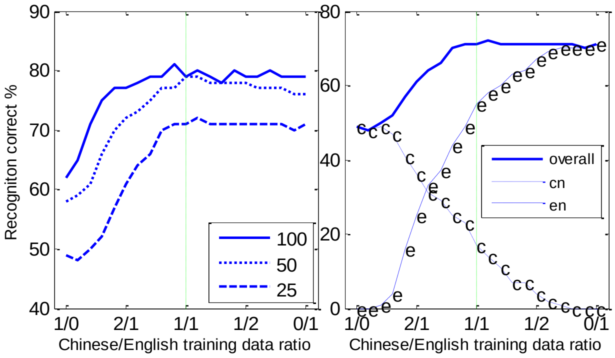


An intriguing feature is the progressive reduction in scores on English tokens that occurs in spite of the fact that English is more dominant in the model's exposure. This reduction may be due to the fact that the model is exposed to fewer tokens overall for the monolingual English model than at the mid-point of the continuum where it receives equal L1/L2 exposure. This idea is supported by the absence of such an effect in the right panel, where the overall amount of training data is constant across the continuum. The reduction is visible for small amounts of training data (25) but disappears when more training data is used (100), suggesting that in cases of data sparsity, even L2 data is better than none at all.

3.2. Effects of separate L2 categories

Figure 2 (left) illustrates the consequence of maintaining separate L2 models rather than blending L1 and L2 material into a single category, for the case of additive L2 data (i.e. corresponding to the left panel of Figure 1). While the pattern is quite similar to the blended case, some differences are evident. The initial rapid rise is delayed, and peak scores are somewhat lower for the independent models. Both outcomes may be caused by paucity of L2 exposure when material is not being incorporated into existing categories. The other key difference is the lack of reduction in scores as the L1:L2 ratio approaches the monolingual target sound model.

Figure 2: Scores for independent L2 models.

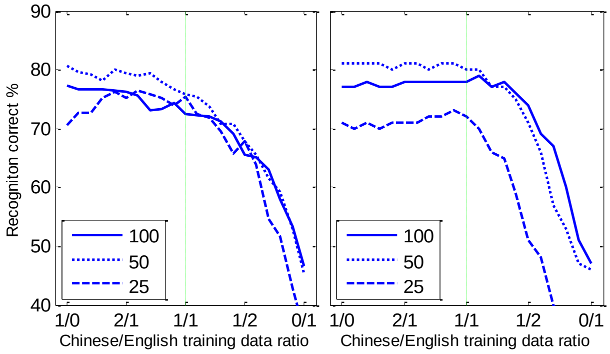


Additional insights can be obtained by depicting the extent to which L1 and L2 categories are matching the target tokens. Figure 2 (right) breaks the overall identification scores (solid curve) into correct identifications by the English and Chinese models. The example shown is for the 25 tokens case, but the others show a similar pattern. It is clear that once sufficient exposure to English tokens has been received, the new L2 models make a very rapid contribution to category identification, while the contribution of the L1 categories gradually declines. One interesting finding is that even at the mid-point of the continuum when both models have the same amount of training data, L2 categories are still contributing to the recognition of L1 tokens, a trend which continues even long after overall performance has asymptoted. While interlanguage similarity for these consonants doubtless contributes, another possibility is that the model from one language may be more sensitive to features more important in that language but not in the other one (e.g. aspiration for Chinese and voicing for English). Increasing robustness via dual-language exposure has also been demonstrated in [8] which showed that Chinese listeners can understand Chinese accented English better than native English listeners.

3.3. Attrition

To determine whether L2 exposure affects L1 performance, Figure 3 plots scores for the blended models (left) and separate L1/L2 models (right) in recognising Chinese tokens. For the blended models, there is some evidence of attrition revealed by a drop in scores up to the mid-point of the continuum, while separated models appear better able to resist attrition, perhaps due to the lack of mixing of L1 and L2 data.

Figure 3: Recognition of L1 tokens by blended data (left) and independent (right) models.



4. DISCUSSION

The current study reports on a computational framework for simulating some aspects of L2 sound acquisition. The framework can be used to explore issues such as exposure, category representations and attrition in a quantitative manner. The tool has the potential to be used both to confirm (or otherwise) observations made with L2 learners and bilinguals, and also to discover and make predictions about outcomes in L2 learning. Among the less-expected results of the simulations are the suggestions that (i) even very small quantities of L2 material (6 or 7 tokens per category) can lead to rapid improvements in recognition of L2 target consonants; (ii) in the case of minimal L1 exposure (as during the early stages of development), an L1/L2 mixture is more beneficial than single-language categories; and (iii) even with balanced amounts of training data across the two languages, L2 models can contribute to the recognition of L1 tokens.

The current framework is in need of extention and refinement in two areas. First, in the data blending strategy existing L1 categories are not adapted sequentially, as one would find in an adult learner, but rather retrained with mixed L1/L2 exemplars. Second, the approach here has been illustrated with L1 sounds having close counterparts in the L2. Further work is required to develop a model capable of quantitative predictions for those L2 sounds with a more complex mapping to native categories.

5. REFERENCES

[1] Best, C.T. 1995. A direct realist view of cross-language speech perception. In Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, 171-204.

[2] Best, C.T., McRoberts, G.W., Goodell, E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *J. Acoust. Soc. Am.* 109, 775-794.

[3] Cooke, M.P., Scharenborg, O. 2008. The interspeech 2008 consonant challenge. *Proc. Interspeech 2008* Brisbane, Australia.

[4] Escudero, P., Kastelein, J., Weiand, K., van Son, R.J.J.H. 2007. Formal modelling of L1 and L2 perceptual learning: Computational linguistics versus machine learning. *Proc. Interspeech 2007* Antwerp, Belgium.

[5] Flege, J. 1995. Second-language speech learning: Theory, findings and problems. In Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, 233-273.

[6] Flege, J. 2009. Give input a chance! In Piske, T., Young-Scholten, M. (eds.), *Input Matters in SLA.* Bristol, England: Multilingual Matters, 175-190.

[7] Gong, J., Cooke, M., Garcia Lecumberri, M.L. 2010. Towards a quantitative model of Mandarin Chinese perception of English consonants. *Proc. NewSounds 2010* Poznan, Poland.

[8] Hayes-Harb, R., Smith, B., Bent, T., Bradlow, A.R. 2008. The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *J. Phonetics* 36(4), 664-679.

[9] Iverson, P., Kuhl, P.K. 1995. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *J. Acoust. Soc. Am.* 97, 553-562.

[10] Thomson, R.I., Nearey, T.M., Derwing, T.M. 2009. A modified statistical pattern recognition approach to measuring the crosslinguistic similarity of Mandarin and English vowels. *J. Acoust. Soc. Am.* 126(3), 1447-1460.

[11] Ventureyra, V.A.G., Pallier, C., Yoo, H. 2003. The loss of first language phonetic perception in adopted Koreans. *J. Neurolinguistics* 17(1), 79-91.

[12] Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., Woodland, P. 2006. *The HTK book (for HTK version 3.4).* Cambridge: Cambridge University Engineering Department.