The role of phonetic training in L2 speech learning

Cristina Aliaga-García
University of Barcelona

1 Introduction The results from the literature on L2 speech learning suggest that factors other than age of learning also have a significant influence on the accuracy of Second Language (L2) learners to perceive and produce non-native speech segments. Whereas it appears reasonable to introduce foreign languages (FLs) as early as possible, successful L2 pronunciation is also dependent on a wide variety of variables such as L1 background, amount of L1 and L2 use, length of residence in a L2-speaking environment, gender, language learning aptitude and motivation. Within the foreign language (FL) classroom context, quantity and quality of L2 input, together with concern for pronunciation, seems today a necessity as well as a significant predictor of the accuracy to pronounce and perceive the sounds of an L2 (Flege *in press*).

The fact that spending many years in a FL classroom will not guarantee itself successful L2 pronunciation -especially when pronunciation still receives little attention in most FL classrooms in Spain- shows the need for new learning environments with massive exposure to high-quality input. Within this context, phonetic training (Catford & Pisoni 1970; Moyer 1999) constitutes a research paradigm for exploring input effects on L2 speech learning and provides a context that may help to maximize the effectiveness of FL teaching.

The present paper reports on the results of a project investigating the short-term effects of a six-week phonetic training on the perception and production of a particular set of non-native sounds that are difficult for speakers of Romance languages, and discusses the implications of using specific phonetic training methodology for guaranteeing the success of pronunciation teaching.

Cross-language voice onset time (VOT) duration differences in the production of oral stops, as well as differences in spectral and duration characteristics of vowels, have shown clear limitations in the accuracy with which L2 phonetic segments are perceived and produced by native speakers (NSs) of Romance languages. On the one hand, speakers whose L1 has short-lag voiceless stops (VOT of 0-30ms, such as Catalan and Spanish) have been found to produce English voiceless stops inaccurately, with values that fall short of the typical 40-80ms VOT range of English monolinguals (Flege *et al.* 1998). On the other hand, L2 learners' failure to detect subtle spectral differences between contrasting vowels which overlap a single L1 category explains their tendency to produce them without substantial durational or spectral difference, resulting in a merged L1-based category (single-category assimilation), or their over-reliance on durational cues as a non-native strategy to produce vowel contrasts (Cebrián 2006), along with L2 vowels produced with formant frequency values intermediate to the values of the L1 and L2 (two-category assimilation) (Flege *et al.* 1997).

Based on this evidence, the purpose of this study was to assess the effects of phonetic training on two domains of Catalan/Spanish learners' phonetic accuracy, voice onset time (VOT) in oral stops and spectral and duration characteristics of contrasting vowels. After intensive exposure to high-quality input, specific practice on the perception and production of the target sounds and auditory-visual feedback, learners were predicted to improve their accuracy (1) in oral stop perception and production (through increased VOT durations) and (2) in vowel discrimination and production (through increased English-like sensitivity to cue weighting).

2 Methodology

2.1. Subjects In order to assess the effects of phonetic training, and attribute the possible gains found in L2 perception and production to the training itself, thirty-six participants took part in a pretest-posttest experiment. Two groups of bilingual Catalan/Spanish undergraduate students of English Philology (NNS; *N*=29) at the University of Barcelona –randomly classified into experimental (*N*=18) and control (*N*=11) – and a control group of NSs of British English (*N*=7) who provided base-line data. Only the experimental group went through a six-week phonetic training period, after which all groups did the same perception and production tasks again (post-test).

2.2. Phonetic Training The experimental group participated in six two-hour training sessions specifically dealing with the articulatory and distributional properties of English oral stops (/p t k b d g/) and the English vowel system, particularly the spectral and durational dimensions distinguishing /iː/-/ɪ/, /æ/-/ʌ/-/ɑː/ and /uː/-/ʊ/. Intensive practice based on various perceptual and productive tasks was preceded by an introductory theoretical part consisting of articulatory visual description, exposure to NS models and contrastive analysis. The learners received *immediate* or *trial-by-trial* feedback during the sessions, *cumulative* feedback at the end, and *weekly* feedback. Finally, group sessions were complemented with individual 15-minute sessions based on computer-based visual feedback.

Lesson and Contents Planning. The experimental group was divided into eight small groups (of two and three people) according to the level estimated after the pre-test and their time availability. The training sessions were administered on Days 1 to 6 and followed a similar structure:

1. INTRODUCTION: interesting readings and/or revision of contents.
2. THEORY: Articulatory (visual) description of sounds; tips for learning pronunciation; NS models; Contrastive Analysis: Spanish/Catalan vs. English; w*hat we must and mustn't do*.
3. 45-MINUTE PRACTICE ON CONSONANTS
   Focus on aspiration and voice onset time as a cue for contrasting voiceless vs. voiced stops in English; closure, hold and release phases for the articulation of plosives; contexts of aspiration and distribution of English stops; cross-linguistic differences; spectrographic feedback.
   3.1. Work on perception, using a variety of tasks different from the pre-test/post-test; multiplicity of contexts, NS voices and accents of English; immediate feedback; work in pairs to encourage peer-correction; critical listening to other non-native accents.
   3.2. Work on production, using a variety of tasks (i.e. reading aloud, listen and repeat…); immediate feedback; recording oral performance in class using microphones and digital recorders.
4. 45-MINUTE PRACTICE ON VOWELS
   Study of the English vowel system (high-mid-low, front-central-back, open/mid-open/mid-close/close, tense vs. lax, rounded vs. unrounded) with main focus on the vowel contrasts /iː/-/ɪ/ and /æ/-/ʌ/, to learn the tense-lax and front-central distinctions respectively. /uː/-/ʊ/ was introduced to add further details to the tense-lax dimension; /ɑː/ was contrasted with the front-central vowels. Description of the tongue movement and lip position.
   Perception and *production* (see above for details).
5. FEEDBACK

Key features and methodological issues. A multiplicity of tasks were used in order to: a) develop the perceptual and productive abilities of the participants, b) modify their performance on certain pronunciation aspects (i.e. unaspirated stops in word-initial position, overuse of the length cue to distinguish the tense-lax vowel contrast) and, finally, c) permit generalisation or transfer to novel stimuli or tasks outside the training. However, identification tasks were preferred over discrimination tasks, since they are the most widely used training method of stimulus presentation (Pisoni & Lively, 1995), in order to learn to perceive non-native categories and promote generalisation to novel stimuli. The reason may be that the learner is required to identify the stimulus item's characteristics rather than compare the stimuli with each other.

Individual training sessions with EyeSpeak. Subject-controlled stimulus presentation followed the mainstream training sessions. The main advantage is the learners' possibility of focusing on stimuli that are particularly difficult to perceive, or even have an increased number of presentations, but also keep a record of the performances. Individual 15-minute sessions took place once a week in a quiet room with the visual pronunciation software EyeSpeak (Ferguson 2005; www.eyespeak.info), which provided learners with the opportunities to use a visual approach to learn and practice vowels, that allowed them to see the position of the vowel in a graph and its degree of highness/backness, check the vowel length and obtain immediate feedback (score) and opportunities for self-correction. It also allowed for work on self-monitoring aspects of learning, going as further as possible according to one's level and interests, and taking one's own decisions during the learning process.

Some other features related to the stimuli used in the training were: use of multiple talkers and natural tokens in multiple acoustic and phonetic contexts; use of natural tokens rather than unnatural synthetic stimuli*;* gradual transitions from easily identified stimuli to more difficult stimuli; introduction to the IPA phonetic symbols; attention to individual differences through work in groups; participants' familiarisation with technology such as: headphones, microphones, interactive visual feedback, computer-generated spectrograms, etc.

2.3 Assessment: Perception and Production Tests. Categorical perception was assessed through an oral stop (/b-p/; /d-t/) identification task based on stimuli from [b]-to-[p] and [d]-to-[t] VOT continua and a minimal-pair AX vowel discrimination task (/iː-ɪ; æ-ʌ/).

In the identification task, the participants were presented 120 randomized stimuli for identification (2 contrasts x 2 VOT continua x 15 5-ms steps x 2 repetitions), produced by a female and a male speaker of British English, at 1-second intervals distributed into eight 15-stimuli blocks separated by 10-second pauses. They were asked to perceptually identify one member of a minimal pair contrasting voicing (p/b and t/d).

In the AX discrimination task, learners were presented 3 repetitions of 24 monosyllabic minimal pairs (e.g. *feel-fill*) pairs and 6 distractors (e.g. *wheel-wheel*) containing the vowel contrasts /iː-ɪ; æ-ʌ/ in a variety of phonetic environments (CVC, CVCC, CCVC and CCVCC), after a previous familiarization phase. Their task was to indicate whether the two stimuli in each of the randomized 90 word-pairs distributed in 6 sections of 15 trials were the *same* or *different.*

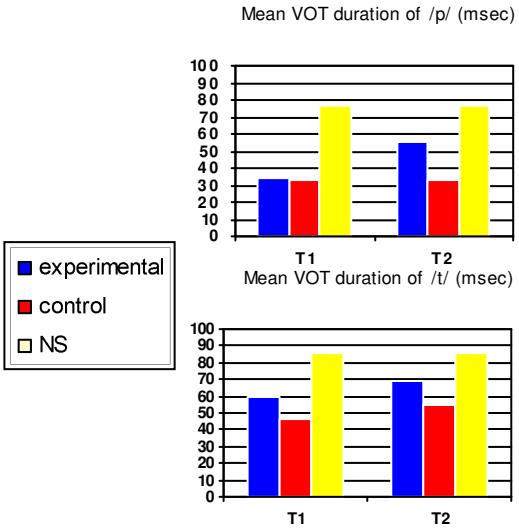
Accuracy in production was assessed by measuring VOT durations in oral stops and formant frequency and length in vowels in words elicited through a sentence repetition task (Flege *et al.* 1995):

A	What is the next word?
B	*BEACH* is the next word.
A	What is the next word?
You	_____ *is the next word.*

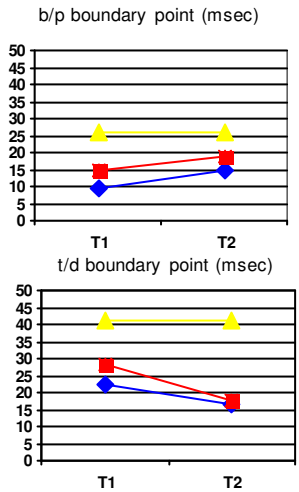
3 Results and discussion: effects of phonetic training The results revealed differential gains in perceptual and productive ability according to type of phonetic category and speech sounds under focus.

3.1. Effects of training on the perception and production of English oral stops. Mean category boundaries were computed for each subject by linear interpolation of the 50% cross-over point for each VOT continuum. Accuracy in the learners' production of oral stops was assessed by means of 6912 VOT measurements (96 words x 36 subjects x 2 data collection times). According to t-tests, the experimental group, but not the control group, obtained significant accuracy gains in the production of /p/ (p= .000), but /t/ (p= .006) improved only slightly (see Figures 1-2). However, in

perception, the location of the /b/-/p/ and /d/-/t/ category boundaries was not significantly affected by phonetic training (see Figures 3-4).



Figures 1 and 2. Mean VOT (ms) of /p/ and /t/.



Figures 3 and 4. Mean b/p and t/d perceptual boundaries.

3.2. Effects of training on the perception and production of English vowel contrasts. As regards vowels, mean percent correct discrimination scores were computed for each subject and contrast. Accuracy in L2 vowel production was assessed by means of 6912 F1-F2 and 6912 duration measurements (96 words x 36 subjects x 2 data collection times) of each vowel. As expected, t-tests revealed that the discrimination ability of trained participants improved significantly at post-test (see Figure 5), whereas in production only modest accuracy gains were obtained. Unlike NSs, Catalan/Spanish learners did not rely on quality differences to implement the /ɪ/-/iː/ contrast but did it in a non-native way to produce /æ/-/ʌ/. On the other hand, length was consistently over-used as a distinctive vowel differentiation cue by learners, especially for /ɪ/-/iː/.

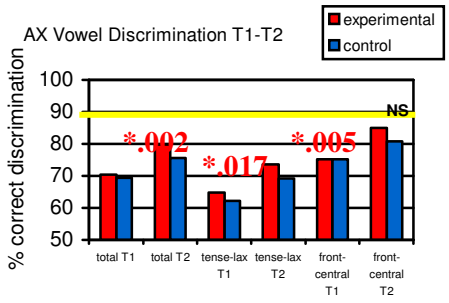


Figure 5 Mean %correct vowel discrimination.

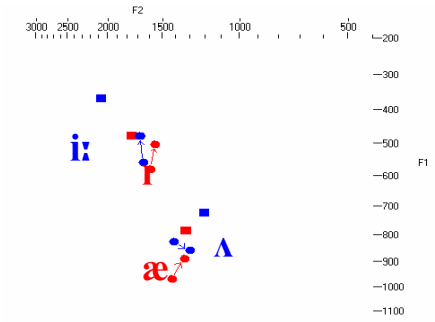


Figure 6 Vowel formant plot for the English vowels pronounced by NSs (■ ■) and learners (● ●) at T1-T2, with arrows showing the degree and direction of change in the F1-F2 values.

4 Conclusions and implications for FL teaching The results presented in this paper show that phonetic training does have some significant short-term effects on the learners' pronunciation, i.e. in the form of improved discrimination of English vowel contrasts and more English-like production of oral stops in word-initial position. One possible explanation for mixed results, the fact that training effects were not so obvious in other cases, is that learners probably need to be exposed to longer training sessions to enhance their identification of oral stops on a

VOT continuum –which probably is less malleable or susceptible to change– and their production of English vowel contrasts (on the basis of spectral difference). On the other hand, the type of assessment may have influenced the findings. Since learners showed improvement in some cases, we think that the implementation of the methodology previously described within the FL classroom would probably raise learners' awareness of some tools and tips available to improve pronunciation and, therefore, help them attain a more accurate L2 pronunciation. Today it is by all means necessary to include a powerful training component within the FL classroom, based on L1-L2 similarities/differences, and including high quality and quantity of L2 input, which encourages learners to use the FL as much as possible, since it has been widely documented L2 pronunciation is not simply picked up by beginning to study the FL from a very early age onwards. However, for an appropriate application of training methods, further research is needed to investigate the extent to which phonetic training may affect L2 learner's lower-level articulatory and perceptual dimensions of speech production as well as representations of phonetic categories.

5 References

Catford, J., and David Pisoni. 1970. "Auditory vs. articulatory training in exotic sounds". *Modern Language Journal* 54: 477-481.

Cebrián, J. 2006. "Experience and the use of non-native duration in L2 vowel categorization". *Journal of Phonetics* 34: 371-387.

Flege, J. E. *in press*. "Give input a chance!". *Input Matters.* Eds. M. Young.Scholten and T. Piske. Clevedon: Multilingual Matters. Piske. Clevedon: Multilingual Matters.

Flege, J. E., Munro, M. and MacKay, I. R. A. 1995. "Factors affecting degree of perceived foreign accent in a second language". *Journal of the Acoustical Society of America* 97: 3125-3134.

Flege, J. E.; Bohn, O. S., & Jang, S. 1997. "Effects of experience on non-native speakers' production and perception of English vowels". *Journal of Phonetics* 25: 437-470.

Flege, J.E., Frieda, E.M., Walley, A.C. and Randazza, L.A. 1998. "Lexical factors and segmental accuracy in second language speech production". *Studies in Second Language Acquisition* 20: 155-187.

Moyer, A. 1999. "Ultimate Attainment in L2 Phonology: The Critical factors of Age, Motivation and Instruction". *SSLA* 21: 81-108.

Pisoni, D. & Lively, S. 1995. Variability and Invariance in Speech Perception. A New Look at Some Old Problems in Perceptual Learning. Chapter 15 in in Strange, Winifred. (ed.) *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Research.* Timonium, MD: York Press Inc. pp. 433-459.