PERCEPTION AND PRODUCTION OF ORAL STOPS BY CATALAN/SPANISH LEARNERS OF ENGLISH: A PHONETIC TRAINING EXPERIMENT

CRISTINA ALIAGA-GARCÍA
JOAN C. MORA
Universitat de Barcelona

ABSTRACT. *Cross-language Voice Onset Time (VOT) duration differences in the production of oral stops have proved a useful phonetic cue in determining the extent to which the phonetic systems of native speakers of Romance languages approximate that of native speakers of English in L2 speech learning. The present study investigates Catalan/Spanish late learners' ability to modify their English /p/-/b/ and /t/-/d/ perceptual category boundaries and VOT durations in the production of word-initial /p, b, t, d/ through a six-week phonetic training period. A lexical identification task based on VOT continua and VOT duration measures in oral stops elicited through a delayed sentence repetition task were used to assess learners' gains in perceptual and productive ability, respectively, before (pre-test) and after (post-test) training. VOT durations were found to increase slightly for /t/ and significantly for /p/, whereas learners' perceptual boundary locations along VOT continua were not significantly affected by phonetic training. These results suggest that accuracy in L2 sound production, but not perceptual category boundaries, may be improved through a short period of intensive phonetic training.*

KEY WORDS*: Voice Onset Time (VOT); oral stops; phonetic training; L2 speech learning; perception; production.*

RESUMEN**.** *La duración del VOT de las oclusivas constituye un rasgo fonético útil para determinar el grado de aproximación del sistema fónico de una lengua romance al de la lengua inglesa durante el proceso de adquisición fonológica. Este estudio investiga los efectos de un periodo de entrenamiento fonético de seis semanas sobre la capacidad de aprendices adultos de inglés de modificar las fronteras perceptuales entre las categorías fonéticas inglesas /p/-/b/ y /t/-/d/ y las duraciones del VOT en la producción de /p, b, t, d/ en posición inicial de palabra. Las posibles ganancias perceptules y productivas se estimaron a partir de dos pruebas realizadas antes y después del entrenamiento fonético: una tarea de identificación léxica basada en un continuo de VOT y medidas de duración de VOT en oclusivas obtenidas mediante una tarea de repetición retardada de frases. Los resultados indican que la duración del VOT aumenta discretamente para la /t/ y significativamente para la /p/, mientras que no se observan cambios significativos en la localización de fronteras perceptuales. Estos resultados sugieren que un periodo relativamente breve de entrenamiento fonético puede mejorar aspectos productivos de la pronunciación de las oclusivas pero no modifica significativamente los límites perceptules entre las mismas.*

1. INTRODUCTION

Studies on second language (L2) speech learning in naturalistic settings have shown that learners are not able to produce L2 speech authentically across a number of segmental (and suprasegmental) phonetic dimensions. Approaches to phonological acquisition such as Flege's (1995) Speech Learning Model, have shown that one of the factors determining the degree of accuracy in L2 sound perception and production (see also Best 1995) is the nature of the interaction between the learner's L1 and L2 phonetic systems, characterised by phonetic category assimilation and dissimilation (Flege *forthcoming*). Age is often also acknowledged a central role in explaining ultimate attainment in L2 speech. A major finding stemming from this research programme is that learners who start L2 speech learning before the passing of a critical period (around the age of 12 years), i.e. early learners, tend to outperform late learners in the perception and production of L2 vowels (e.g. Piske *et al.* 2002) and consonants (e.g. MacKay *et al.* 2001), and are judged to have less strong foreign

accents (e.g. Flege *et al.* 1995). An explanatory account of the early-start advantage based on general neurological maturation (e.g. Scovel 1988), however, faces an important limitation in the fact that many contextual factors affecting L2 speech learning, such as amount and quality of L2 input or amount of L2 and L1 use, are confounded with the age at which learning began. Within this research context, amount of L2 input may be operationalized as "years of residence in the L2 speaking country", but this variable can only roughly quantify the actual amount of L2 speech an individual is exposed to (Flege *in press*). Studies measuring the differential effects of a stay-abroad term and an at-home period of formal instruction on accuracy gains in L2 sound perception and production (e.g. Díaz-Campos 2004; Mora *in press*) provide a context where input effects may be observed in the short term; however, the input a learner receives in an immersion situation is not objectively quantifiable either.

Formal instructional settings offer the possibility of measuring L2 input quantity effects more reliably in terms of "hours of instruction" by comparing groups of learners of the same age differing in the amount of instruction received (e.g. Fullana 2005), but controlling for input quality effects would require evaluating speech data that is not normally available for analysis. Phonetic training studies constitute a research paradigm that has produced interesting results as regards input effects on L2 speech learning, often reporting the effectiveness of laboratory training in improving L2 pronunciation (e.g. Catford & Pisoni 1970; Moyer 1999). Such results have a bearing on crucial issues in L2 speech learning, such as the ability of late learners' perceptual and articulatory systems to remain adaptive to linguistic experience. The present paper, which is part of a larger project investigating phonetic training effects on L2 pronunciation, further explores this line of research by measuring phonetic training effects on advanced Catalan/Spanish learners' perception and production of English oral stops.

2. METHOD

Twenty-one participants took part in the present study, two groups of bilingual Catalan/Spanish 1st year undergraduate students of English philology (NNS) at the University of Barcelona (see Table 1): experimental (*N*=9) and control (*N*=5), and a control group of native speakers of British English (NS; *N*=7) who provided base-line data. The results obtained from the speech perception and production tasks conducted on NNS participants before phonetic training (pre-test) were used to create two homogeneous groups through paired sampling, which were then randomly labelled as experimental and control. 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).

Participants		Pre-Test (T1: March 2006)	N	Phonetic Training (April-May 2006)	Post-Test (T2: June 2006)	N
NNS	Experimental	✓	9	✓	✓	9
	Control	✓	9	×	✓	5
NS Control		×	7	×	✓	7
Total			25			21

Table 1. *Experimental design and data collection times (T1 and T2)*

2.1. *Perception and production tasks*

Learners' accuracy in oral stop perception and production was assessed by means of a forced-choice lexical identification task based on VOT continua and VOT duration measures of oral stops elicited through a delayed sentence repetition task.

In the identification task the participants were asked to perceptually identify one member of a minimal pair contrasting voicing. Two 15-step VOT continua per contrast (p/b and t/d) ranging from 0 to 70 ms was used to generate 15 modified instances of each word, which were randomly presented twice to participants for identification (see Table 2). Voiced-to-voiceless VOT continua (e.g. *bin* to *pin*) were obtained by inserting a 70-ms period of voiceless breath between the release burst of a voiced oral stop and the onset of the following vowel at 5 ms steps. Participants were thus presented with 120 randomized stimuli for identification (2 contrasts x 2 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.

A delayed sentence repetition task (Flege *et al.* 1995) was used to elicit and record learners' production of word-initial /p t b d/ followed by /□.]□□ ℘/ in a carrier phrase:

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

	PERCEPTION: identification task			PRODUCTION: delayed sentence repetition task			
	Items for ID	Repetitions	Total	Contexts	Words	Repetitions	Total
p/b	pack/back (2)	15	30	4	2	3	24
	pin/bin (2)	15	30	4	2	3	24
t/d	Tan/Dan (2)	15	30	4	2	3	24
	teans/deans (2)	15	30	4	2	3	24
	Stimuli for identification:		120	Elicited oral stops for VOT analysis:			96

Table 2. *Items in the perception and production tasks*

Early and late learners of English in immersion contexts whose L1 have 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 & Eefting 1987; Flege *et al.* 1998). Studies using VOT continua in a perceptual identification task have also found robust cross-linguistic differences. For example, Spanish-English bilinguals have been found to identify stimuli as instances of the /p/ category at shorter VOT durations than English monolinguals, suggesting that the VOT-based category boundaries of NSs and NNSs are placed at different locations along a VOT continuum (Flege & Schmidt 1995, Flege *et al.* 1996). However, it has also been shown that short-term laboratory training may be effective in modifying NNS VOT-based perceptual boundaries between English voiced and voiceless oral stops (Bohn & Flege 1993). Based on this evidence, the present study sets out to explore the extent to which accuracy in the perception and production of oral stops by Catalan/Spanish late learners of English may be modified through a six-week phonetic training period, which was predicted to (1) improve learners' accuracy in oral stop production through increased VOT durations and (2) shift the learners' perceptual boundary locations (between voiced and voiceless stops) towards longer durations along a VOT continuum.

2.2. *Phonetic Training*

The experimental group participated in 6 two-hour training sessions specifically dealing with the articulatory and distributional properties of English orals stops. 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.

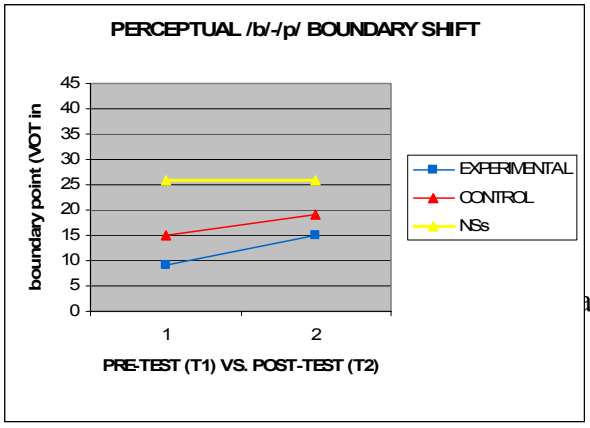
3. RESULTS AND DISCUSSION

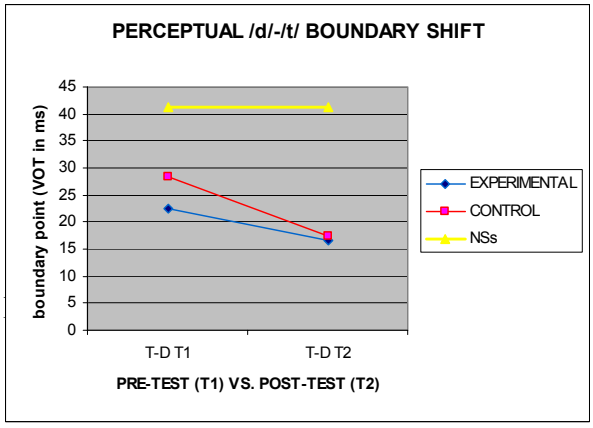
3.1. *Perception*

Mean category boundaries were computed for each subject by linear interpolation of the 50% cross-over point for each VOT continuum (see Table 3). As expected, the perceptual boundaries were located at lower VOT values for the Catalan/Spanish learners than for the NS control group at T1, which suggested that stimuli were more often identified as /p t/ by NNSs than by NSs. A t-test revealed no significant differences between the NNS experimental and control groups at T1. At T2, after phonetic training, the mean p/b category boundaries of the experimental and the control group shifted to longer VOT values, but the t/d boundary unexpectedly shifted to shorter VOT (see Figure 1). T-tests at T2 did not show significant differences between the NNS experimental and control groups but continued to reveal significant differences between NSs and NNSs, as regards p/b (p= .040) and t/d (p= .024) boundaries, suggesting that phonetic training did not have much effect on non-native patterns of perceptual identification towards English-like use of VOT as a cue for stop identification. Unlike NSs, NNSs still identified short-lag stimuli (0-30 ms) as p/t tokens after training, showing that they still judged the VOT continua in terms of native-like distinctive VOT values.

		p/b boundary		t/d boundary	
		T1	T2	T1	T2
NNS	Experimental Group	9.17	15	22.5	16.67
	Control Group	15	19	28.33	17.50
NS	NS control group	25.80		41.33	

Table 3. *Mean boundary points (ms) at T1 and T2.*





3.2. *Production*

Accuracy in the learners' production of oral stops was assessed by means of 2688 VOT measurements (96 words x 14 subjects x 2 data collection times). VOT values (in ms) were obtained by measuring the distance between the onset of the release burst and the onset of vocal fold vibration in wide-band spectrograms. When /b d/ were prevoiced, *lead* VOT was measured from the beginning of low-frequency periodicity to the onset of the release burst, and assigned negative values.

	Mean VOT (ms)										
	/p/			/b/			/t/			/d/		
	T1	T2	*p*=	T1	T2	*p*=	T1	T2	*p*=	T1	T2	*p*=
Experimental group	33.68 (14.59)	55.79 (30.16)	*.037	-54.69 (34.89)	-62.47 (24.56)	.413	59.11 (19.41)	68.96 (27.21)	.069	-42.72 (24.52)	-55.82 (26.18)	.071
Control group	33.46 (17.90)	32.70 (19.31)	.704	-34.64 (42.10)	-33.86 (28.52)	.937	45.87 (10.74)	54.78 (15.32)	.259	-57.18 (59.79)	-56,63 (57.27)	.957
NS group	76 (10.20)		-	-39.75 (49.06)		-	86.18 (13.71)		-	-21.86 (45.72)		-

Table 4. *Mean VOT duration (ms) at T1 and T2 and significant differences at α=.05* .

Mean VOT values obtained at T2 were higher than those at T1 (see Figure 2). Paired-samples t-tests revealed no significant T1-T2 differences for the control group, but the experimental group showed a significant increase in VOT for /p/. The significant differences found between the experimental and the NS group at T1 as regards /p/ (*p*= .000) and /t/ (*p*= .006) disappeared after training according to the t-tests conducted at T2, which confirms the positive impact of phonetic training upon mean VOT duration. Production of alveolar /t/ −which is dentialveolar in their L1 and therefore more dissimilar than /p/ cross-linguistically− showed, however, a lower degree of improvement at T2 as it approximated the English phonetic norm already before training.

However, the mean VOT values obtained at T2 are still far from the native-like average: the mean VOT for /p/ and /t/ in the present study was almost intermediate to the means observed for monolingual Spanish adults (26 ms) in Caramazza's study (1973), on the one hand, and NS values from the present study, on the other.

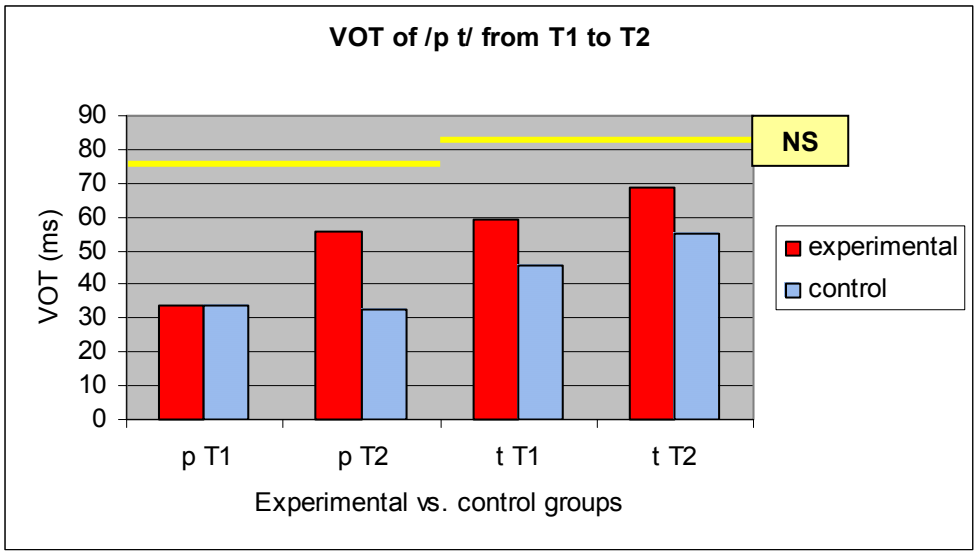


Figure 2. *Mean VOT (ms) of /p/ and /t/ (T1 and T2).*

A paired-samples t-test confirmed that the production of /b d/ did not benefit from such training (cf. Mackay *et al.* 2001): the experimental group showed a higher amount of /b d/ prevoicing than NSs of English at T2, with no significant differences between the mean VOT durations of /b/ and /d/ at T1 and T2 (see Table 4). Similarly, independent-samples t-tests showed no significant advantage of the experimental group over the control group in this respect.

4. CONCLUSION

The results suggest that the input administered through phonetic training did not have the same effects on the subjects' perception and production of /p/-/b/ and/ /t/-/d/: the amount of gains obtained from T1 to T2 differed considerably according to the ability and the speech sound under focus. The learners' production of Catalan/Spanish-accented /p/ was more English-like (i.e. produced with longer VOT) after the training period. However, production was not aligned with perception in this respect, since NNSs showed a considerably higher percentage of voiceless stop identification than NSs. Lack of significant improvement in some cases may be due, on the one hand, to the small pool of valid subjects participating in the present study. On the other hand, a six-week phonetic training may not be long enough to produce similar gains in perceptual and productive ability.

ACKNOWLEDGEMENTS

This research was supported by research grant HUM2004-05167/FILO from the Spanish Ministry of Science and Technology. We would like to thank the UB students who volunteered to participate in the phonetic training and did the perception and production tests analysed in the present study.

REFERENCES

Best, C. 1995. "A direct realist view of cross-language speech perception". *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues.* Ed. W. Strange. Timonium, MD: York Press. 171-204.

Bohn, O. and Flege, J.E. 1993. "Perceptual switching in Spanish/English bilinguals evidence for universal factors in stop voicing judgments". *Journal of Phonetics* 21: 267.

Caramazza, A.; Yeni-Komshian, G. H.; Zurif, E. B. & Carbone, E. 1973. "The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals". *The Journal of the Acoustical Society of America* 54: 421-428.

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

Champagne-Muzar, C.; Schenidерman, E. I. & Bourdages, J. S. (2003). Second Language accent: The Role of the pedagogical environment. *International Review of Applied Linguistics,* 31: 143-160.

Díaz-Campos, M. 2004. "Context of learning in the acquisition of Spanish second language phonology". *Studies in Second Language Acquisition* 26: 249–273.

Flege, J. E. 1995. "Second-language speech learning: theory, findings and problems". *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues.* Ed. W. Strange. Timonium, MD: York Press. 229-273.

Flege, J. E. and Schmidt, A.M, and Wharton, G. 1996. "Age of learning affects rate-depending processing of stops in a second language". *Phonetica* 53: 143-161.

Flege, J. E. *forthcoming*. "Language contact in bilingualism: phonetic system interactions". *Laboratory Phonology 9*. Eds. J. Cole and J. Hualde. Berlin: Mouton de Gruyter.

Flege, J. E. *in press*. "Give input a chance!". *Input Matters.* Eds. M. Young.Scholten and T. 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. and Schmidt, A.M. 1995. "Native speakers of Spanish show rate-dependent processing of English stop consonants". *Phonetica* 52: 90-111.

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.

Fullana, N. 2005. *Age-Related Effects on the Acquisition of a Foreign Language Phonology in a Formal Setting.* Unpublished PhD Thesis. Universitat de Barcelona.

MacKay, I. R. A., Flege, J. E., Piske, T. and Schirru, C. 2001. "Category restructuring during second-language (L2) speech acquisition". *Journal of the Acoustical Society of America* 110: 516-528.

Mora, J. C. forthcoming. "Learning context effects on the acquisition of a second language phonology". *A Portrait of the Young in the New Multilingual Spain.* Eds. C. Pérez-vidal, M. Juan-Garau and A. Bel. Clevedon: Multilingual Matters.

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

Piske, T., Flege, J. E. MacKay, I. R. A. and Meador, D. 2002. "The production of English vowels by fluent early and late Italian-English bilinguals". *Phonetica* 59: 49-71.

Scovel, T. 1988. *A Time to Speak. A PsycholinguisticInquiry into the Critical Period for Human Speech.* Cambridge, MA: Newbury House Publishers.