Recent Research in Second Language Phonetics/Phonology

Recent Research in Second Language Phonetics/Phonology: Perception and Production

Edited by

Michael A. Watkins, Andreia S. Rauber and Barbara O. Baptista

CAMBRIDGE SCHOLARS

PUBLISHING

Recent Research in Second Language Phonetics/Phonology: Perception and Production,
Edited by Michael A. Watkins, Andreia S. Rauber and Barbara O. Baptista

This book first published 2009

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-4438-1125-4, ISBN (13): 978-1-4438-1125-5

TABLE OF CONTENTS

Part IV: Beyond the Syllable

PREFACE

This volume is a selection of recent research papers by some of the leading practitioners in the field of second language speech production and perception. All the papers were presented at *New Sounds 2007: Fifth International Symposium on the Acquisition of Second Language Speech*, held in Florianópolis, Brazil, in November 2007. The collection is divided into four sections: the first two dealing with segments, the third with syllables, while the fourth looks at features above syllable level and some general issues.

In Part I, two papers stress the importance of phonetic training for the improvement of English vowel perception and production. Aliaga-Garcia and Mora worked with Spanish-speaking learners, while Lacabex, Lecumberri and Cook report findings with bilingual Catalan/Spanish learners of English. The third study, by MacKay and Fullana, also involved Catalan/Spanish learners of English, but focused on the variables "age of learning" and "formal learning context" to explain differences in the production of English vowels.

Part II, dealing with consonants, brings in some other languages: Celata tested the perception of nonnative (Russian) post-sonorant contrasts by native speakers of Tuscan and Italian; Eckman, Iverson, Fox, Jacewicz and Lee investigated the perception and production of English consonants by native speakers of Japanese and Korean; Fullana and Mora deal with the perception and production of English word-final obstruents by native bilingual speakers of Catalan and Spanish; John and Cardoso investigated the production of English /h/ by Canadian francophones; and Kluge, Reis, Nobre-Oliveira and Bettoni analyzed the perception of English syllable-final nasals by native speakers of Brazilian Portuguese.

In Part III, three papers focus on the production of English /s/ clusters by native speakers of Brazilian Portuguese: Bettoni and Koerich investigated the effects of perceptual training on the production of initial clusters; Cardoso and Liakin carried out a corpus-based analysis to investigate the production of /s/ clusters; and Cardoso, John and French investigated the acquisition of English /s/ clusters by Brazilians with no knowledge of English. Still in Part III, Cebrian reports on the different syllabification strategies used by native speakers of Catalan and English, while Yavaş analyzed the voice onset time (VOT) of English stops by native speakers of Spanish.

In the final section, Major and Baptista report on a form of language attrition, describing differences between native speakers of Brazilian Portuguese who live in Brazil and others living in the USA with regard to their ability to detect foreign accent in the Brazilian Portuguese spoken by Americans. Silveira reports a case study of the influence of orthography on the production of English word-final consonants by a native speaker of Brazilian Portuguese. Wrembel describes the impact of a second language (German) on phonetic performance in a third language (English) by native speakers of Polish. The last paper, by Van Dommelen and Husby, reports an experiment to test the perception of Norwegian word tones by native speakers of Chinese and German.

All the papers describe careful empirical research, and as such will be of great interest to anyone working, or intending to work, in the specific field of second language phonological acquisition. However, given that speech production and perception are highly complex skills, the research findings in this volume will also be relevant to those with a broader interest in language learning or cognition in general.

PART I:

STUDIES ON VOWELS

CHAPTER ONE

ASSESSING THE EFFECTS OF PHONETIC TRAINING ON L2 SOUND PERCEPTION AND PRODUCTION

CRISTINA ALIAGA-GARCÍA
AND JOAN C. MORA
UNIVERSITY OF BARCELONA, SPAIN

1. Introduction

Based on the assumption that perception leads production in L2 speech learning (Flege 1995) and that greater perceptual accuracy eventually leads to changes in production, phonetic training, and perceptual training in particular, has been shown to result in improved performance by L2 learners in the discrimination and identification of L2 sounds (Flege 1989, Iverson and Evans 2007, Iverson, Hazan and Bannister 2005, Lambacher, Martens, Kakehi, Marasinghe and Molholt 2005, Lively, Logan and Pisoni 1993, Logan, Lively and Pisoni 1991, Logan and Pruitt 1995, Pisoni and Lively 1995), as well as in L2 speech production (Bradlow, Akahane-Yamada, Pisoni and Tohkura 1999). However, the effectiveness of phonetic training is not a direct consequence of the amount of exposure to L2 sounds that learners receive or whether that exposure is more or less intense in time. In fact its effectiveness depends to a great extent on how successful it is in directing both the attention and L2 sound-processing ability of learners to the phonetic cues that native speakers attend to.

The phonetic training paradigm that has proved most successful in doing this is the High Variability Phonetic Training (HVPT) method, which directs learners' attention towards relevant phonetic cues by presenting them with high-variability stimuli in different phonetic contexts (Bradlow, Pisoni, Yamada and Tohkura 1997, Lively et al. 1993) obtained from natural words produced by different speakers. However, although phonetic learning appears to be enhanced by exposure to acoustic variability, it has also been shown (Jongman and Wade 2007) that phonetic training with minimal variability focusing on target phonetic prototypes may be effective in promoting the acquisition of difficult sound contrasts. Because the nature of the interaction between the learners' L1 and L2 phonetic systems may affect the effectiveness of the treatment, phonetic training methods may have to be adapted to the specific L2 sound contrasts to be acquired by learners.

The present study explores this line of research further by investigating the effects of six two-hour phonetic training sessions on the perceptual and productive competence of a group of advanced Catalan/Spanish bilingual learners of English who have been exposed to the L2 mainly through formal classroom instruction in an EFL context in Catalonia, Spain. In this learning context L2 input is very limited and often strongly foreign-accented, so at the time of testing the participants were very unlikely to have developed phonetic categories for the L2 sounds the training focused on. This was evident in the degree of perceived foreign-accentedness of their speech. The focus of the phonetic training was on four L2 sound contrasts with which Catalan/Spanish learners of English have been reported to have difficulty (Cebrian 2002, 2006, Mora 2008, Mora and Fullana 2007), namely /p/-/b/ and /t/-/d/ in word-initial position, as well as the vowel contrasts /iː/-/ɪ/ and /æ/-/ʌ/. The approach was of the HVPT type, mainly (but not only) focusing on the target sound contrasts. The phonetic dimensions on which the training was based presented crucial cross-linguistic differences. In English, the word-initial voicing contrast in /p/-/b/ and /t/-/d/ is realized mainly through VOT duration differences, and the presence/absence of vocal cord vibration during closure is a secondary cue often missing in connected speech (Gimson and Cruttenden 1994). By contrast, in Catalan and Spanish closure voicing rather than VOT differences is used primarily to implement the contrast. As regards vowels, the pairs /iː/-/ɪ/ and /æ/-/ʌ/ present durational differences but the contrasts, which are lacking in Catalan and Spanish, are mainly qualitative, with Catalan and Spanish /i/ and /a/ occupying, respectively, a portion of the vowel space occupied by /iː/-/ɪ/ and /æ/-/ʌ/. Consequently, English /iː/-/ɪ/ and /æ/-/ʌ/ are typically assimilated to the Catalan and Spanish /i/ and /a/ vowel categories, respectively. Vowel duration is not used contrastively in Catalan and Spanish, but previous research has shown that Catalan/Spanish learners of English often overrely on temporal cues in the production of these vowels (García-Lecumberri and Cenoz 1998, Mora and Fullana 2007). The phonetic training method used in the present study was specially designed to enhance the use of VOT duration rather than closure voicing in the production of word-initial oral stops, and the use of spectral rather than durational differences in the production of the /iː/-/ɪ/ and /æ/-/ʌ/ vowel contrasts.

The effectiveness of the phonetic training administered to the experimental group is measured in terms of how successful it is in changing learners’ perceptual ability to categorize /p/, /b/, /t/ and /d/ stimuli (drawn from VOT continua) and to correctly discriminate between /iː/-/ɪ/ and /æ/-/ʌ/. Besides improving perception, the training could be considered successful if it also improves learners’ accuracy in the production of /p/, /b/, /t/ and /d/ through the use of longer VOT duration and shorter closure voicing, as well as in the production of /iː/-/ɪ/ and /æ/-/ʌ/ through the use of spectral rather than temporal differences.

Not only do phonetic training studies have interesting pedagogical implications for L2 pronunciation instruction, but the findings obtained through this research paradigm also have a bearing on crucial issues in L2 speech learning, such as the ability of the adult perceptual system to remain adaptive to new input and to change as learning progresses. Phonetic training studies also provide us with a context for testing the role of quantity and quality of input in L2 speech learning in a controlled way, a possibility that has remained elusive in L2 speech learning studies carried out in naturalistic learning contexts.

2. Method

2.1 Participants

A total of 36 participants (32 females, 4 males) between the ages of 19 and 48 (mean age 21.62) took part (voluntarily) in a pretest (T1) – posttest (T2) experiment: 29 bilingual Catalan/Spanish undergraduate students of English Philology (NNS) at the University of Barcelona − divided into experimental (N=18) and control (N=11) groups − and a control group of native speakers of British English (NSs; N=7) who provided base-line data (Table 1-1). 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). The learners had been receiving formal English instruction for eight years and had not had any previous phonetic training or experience abroad. The age at which they began learning English ranged from 8 to 12 years (mean 9.15). They were given course credits for their participation.

Table 1-1. Design of the study

Participants		Pre-Test (T1) (Week 1)	Phonetic Training (Weeks 2-7)	Post-Test (T2) (Week 8)	N
NNS	Experimental	✓	✓	✓	18
	Control	✓	×	✓	11
NS Control		×	×	✓	7
Total					36

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 features distinguishing /iː/-/ɪ/, /æ/-/ʌ/-/ɑː/ and /uː/-/ʊ/. Intensive practice based on various perception and production 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 working sessions based on computer-based visual feedback.

The training sessions were administered on days 1 to 6 and consisted of an hour's practice on English oral stops (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) and an hour's practice on English vowels (the English vowel system; focus on the tense-lax (/iː/-/ɪ/; /uː/-/ʊ/) and front-central-back (/æ/-/ʌ/-/ɑː/) contrasts; description of tongue movement and position of lips).

A multiplicity of tasks were used (Logan and Pruitt 1995: 351) in order to (a) develop the perceptual and productive abilities of the participants; (b) modify their performance on certain aspects of pronunciation (i.e., unaspirated stops in word-initial position, overuse of the length cue to distinguish the tense-lax vowel contrast); and (c) permit generalization or transfer to novel stimuli or tasks outside the training. The perceptual training consisted of identification, discrimination and phonetic transcription tasks, whereas the productive training was based on imitation and reading aloud tasks. The stimuli used in the training were natural tokens produced by multiple native speakers presenting the target sounds in a variety of acoustic and phonetic contexts (Pisoni and Lively 1995). Attention to individual differences through team work and participants' familiarization with speech-related technology (such as microphones and speech analysis software) were among the general pedagogic features of the training sessions.

Subject-controlled presentation of stimuli followed the mainstream training sessions, which allowed learners to focus on the stimuli that they found particularly difficult to discern, to have an increased number of presentations and to keep a record of their performances. Individual 15-minute sessions took place once a week in a quiet room with the visual pronunciation software EyeSpeak (Chan 2004). This software provided learners with an opportunity to use a visual approach to learn and practice vowels by allowing them to (a) see the position of the vowel on a graph representing an F1-F2 vowel space and its degree of height/backness, (b) check the vowel length, and (c) obtain immediate feedback (score) and opportunities for self-correction. It was also useful in that it enabled the participants to work on self-monitoring aspects of learning, going as far as possible according to their level and interests, and taking their own decisions during the learning process.

2.3 Perception and production tasks

2.3.1 Speech perception tasks

Two speech perception tests were administered to participants. A forced-choice lexical identification task based on VOT continua was used to assess their accuracy in the perception of oral stops, while accuracy in vowel perception was assessed by means of an AX discrimination task.

In the identification task the participants were asked to identify one member of a minimal pair containing a voicing contrast. Two 15-step VOT continua per contrast (C1: p/b; C2: t/d) ranging from 0 to 70 ms were used to generate 15 modified instances of each word, which were randomly presented twice to participants for identification (see Table 1-2). The two VOT continua were made from natural speech and with the test consonant in word-initial position (e.g., *pack-back*) using the software Praat (Boersma and Weenink 2007). 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. The result was 15 different VOT values from 0ms to 70ms voicing lag in 5ms increments for each VOT continuum, forming continua of identifiable words (see Table 1-3). Participants were thus presented with 120 randomized stimuli for identification (2 contrasts x 2 continua x 15 5-ms steps x 2 repetitions) at 1-second intervals. The stimuli were produced by a female and a male speaker of British English and were grouped into eight 15-stimuli blocks separated by 10-second pauses. Trial items were presented only once, and were preceded by 15 practice items from a k/g VOT continuum (*coat-goat*).

Table 1-2. VOT (ms) for each of the stimuli in the /b/-/p/ and /d/-/t/ continua

VOT continua															
Steps	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
Dur (ms)	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70

Table 1-3. Items in the oral stop identification test

Perception: forced-choice lexical identification task					
VOT continua		**Items for ID**	**Repetitions**	**Steps**	**Total**
C1	/p/-/b/	pack/back	2	15	30
		pin/bin	2	15	30
C2	/t/-/d/	tan/Dan	2	15	30
		teans/deans	2	15	30
Stimuli for identification					120

In the AX discrimination task (see Table 1-4), learners were presented with 3 repetitions of 24 monosyllabic minimal pairs (e.g., *feel-fill*) 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. The participants' 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.*

Table 1-4. Items in the AX vowel discrimination test

Perception: AX vowel discrimination task					
Minimal pairs		Distractors	Repetitions	Items	
/iː/ /ɪ/	11 (feet-fit)	3 (chip-chip)	3	42	
/æ/ /ʌ/	13 (cap-cup)	3 (run-run)	3	48	
Word-pairs for discrimination				90	

2.3.2 Speech production tasks

Accuracy in production was assessed by measuring VOT durations in oral stops and formant frequency and length in vowels in words elicited through a delayed repetition technique (Flege, Munro and MacKay 1995), in which elicited target words are embedded in mini-dialogues, as in Mora and Fullana (2007) (See Table 1-5).

Table 1-5. Items in the production test

Production: delayed sentence repetition task					
Word-initial oral stops + Vowels			Words	Repetitions	Total
voiceless	/p/	/iː ɪ æ ʌ/	2	3	24
	/t/	/iː ɪ æ ʌ/	2	3	24
voiced	/b/	/iː ɪ æ ʌ/	2	3	24
	/d/	/iː ɪ æ ʌ/	2	3	24
Elicited oral stops and vowels					96

2.3.3 Target sounds

Early and late learners of English in immersion contexts whose L1 has short-lag voiceless stops (VOT of 0-30 ms, 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 and Eefting 1987, Flege, Frieda, Walley and Randazza 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 and Schmidt 1995, Flege, Schmidt and Wharton 1996). 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 and Flege 1993). In order to quantify oral stop identification data more accurately, studies on the development of phonemic categorization (Hazan and Barrett 2000) and language deficits and speech perception (Joanisse, Manis, Keating and Seidenberg 2000) have reported modified slopes in logistic curve functions as a result of training or experience. Lower slope-related logistic function coefficients represent steeper slopes, indicating higher degrees of categoriality in perception, whilst the location of the category boundary is a useful measure for assessing the effect of VOT duration on stimuli categorization (Benkí 2005).

The contrasting target vowels were selected on the basis of their relative difficulty for Catalan/Spanish learners of English. On the one hand, both Catalan and Spanish lack a tense-lax distinction for the high vowels /iː/-/ɪ/ and have a vowel (/i/) closer to English /iː/ (Flege and Mackay 2004). On the other hand, a front-back distinction for the open/low vowels /æ/-/ʌ/ is also lacking in the participants' L1, being replaced by a relatively front /a/.

Cross-language research has documented L2 learners' failure to detect subtle spectral differences between contrasting vowels which overlap with a single L1 category. This is due to L1-L2 differences in vowel space, the degree of perceived phonetic distance between L1 and L2 vowels and the nature of the interactions between the learner's L1 and L2 phonetic subsystems. As a consequence, learners tend to produce contrasting L2 vowels without a substantial durational or spectral difference, indicating a merged L1-based category (single-category assimilation) (Best 1995, Best and Tyler 2007, Flege 1995). However, several studies have shown non-native speakers to be overrelying on durational cues which do not exist in their L1 as a strategy to perceive and produce vowel contrasts (Bohn 1995, Cebrian 2006, Escudero and Boersma 2004), along with the production of L2 vowels with formant frequency values somewhere between the values of NSs of the L1 and L2 (Flege, Bohn and Jang 1997).

Based on this evidence, the present study set out to explore the extent to which accuracy in the perception and production of oral stops and vowel contrasts by Catalan/Spanish late learners of English could be improved through a six-week period of phonetic training. On the one hand, the training was expected to (a) improve learners' accuracy in oral stop production through increased VOT durations, (b) shift the learners' perceptual boundary locations (between voiced and voiceless stops) towards longer durations along a VOT continuum, and (c) lead to a higher degree of categoriality in the responses and heightened sensitivity to VOT through a steeper slope in the identification curves. On the other hand, it was hypothesized that the training would also improve learners' accuracy in (a) vowel discrimination through increased English-like cue weighting, and (b) the production of vowels through the formation of new phonetic categories and the appropriate use of spectral/durational cues.

3. Results

3.1 Perception and production of /p/-/b/ and /t/-/d/

3.1.1 Effects of training on perception

The participants' performance on the forced-choice lexical identification task was assessed by computing an identification function for each VOT continuum (C1: b-p and C2: t-d) that plots the percentage of one of the two alternative responses, particularly the voiceless stop response (e.g., *pack* or *tan*), for each of the stimuli making up the 15-step continua. A decrease in the percentage of /p/ and /t/ identification after training, particularly in the stimuli with a VOT duration ranging between 25 and 35 ms, suggests that phonetic training has increased participants' sensitivity to VOT as a voicing cue, because longer VOT durations are necessary to obtain a /p/ or /t/ response from participants at T2.

In order to quantify the identification data more accurately, each participant's categorization curve was fitted to a logistic function using the Logistic Curve Fit function in SPSS (Figure 1-1), yielding mean slope coefficients (values between 0 and 1 per stimulus) for each group and continuum (see Table 1-6) as a measure of degree of categoriality of responses, with lower values representing steeper slopes, indicating a more categorical perception of oral stops. Since training was expected to have a positive effect on the categorical perception of the target contrasts, the participants undergoing the treatment were expected to obtain steeper logistic curves than the untrained listeners.

The effect of training on the participants' identification of oral stops was also assessed by means of another measure: mean category boundaries (the 50% crossover points of the fitted labeling curve) were computed for each subject, group and continuum using the formula -Ln(b0)/Ln(b1), where b0 is the constant value of the logistic curve (Morrison 2005). Because the experimental group was expected to label a greater proportion of stimuli as b/d rather than p/t after training, it was hypothesized that /b/-/p/ and /t/-/d/ category boundaries would be located at higher VOT durations on the continuum at T2. Table 1-6 displays data from the two parameters extracted to characterize each identification function: the slope of the fitted curve, taken as an indication of degree of categoriality in identification, and the phonetic category boundary, showing the effect of VOT duration on categorization.

Table 1-6. Slope coefficients (between 0 and 1) and boundary location (ms)

Group	/p/-/b/ VOT continua			/t/-/d/ VOT continua		
		T1	T2		T1	T2
Experimental	boundary (ms)	2.8	11.39	boundary (ms)	28.89	21.92
	slope	0.55	0.49	slope	0.33	0.31
Control	boundary (ms)	1.91	29.49	boundary (ms)	7.74	32.62
	slope	0.54	0.53	slope	0.38	0.32
NS	boundary (ms)	5.37		boundary (ms)	33.79	
	slope	0.514		slope	0.278	

When the Catalan/Spanish speakers' identification responses were compared across continua, C1 and C2 were found to yield very different curves. Whereas the /t/-/d/ continuum produced the expected semi-categorical S-shaped curve, the curve obtained for the /b/-/p/ continuum did not conform to the behaviour typically triggered by this type of identification task based on acoustic continua (where stimuli at both ends of the continuum have an identification rate of 0% or 100%) (Figure 1-1). It is worth noting, however, that a similar trend for both C1 and C2 indicated that perceptual accuracy was slightly affected by training in both cases: NNSs' curves at T2 were found to be slightly steeper than at T1 and approximated the NSs' curve in steepness.

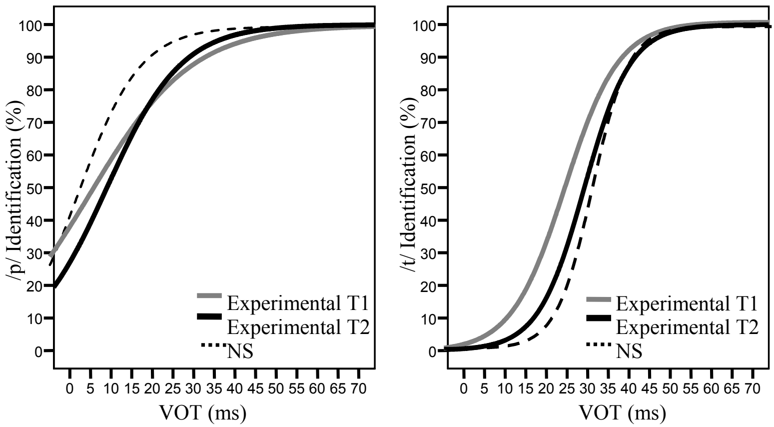

Figure 1-1. Logistic regression functions of /p/ and /t/ categorization curves (% /p/ and /t/ responses) for the NS control group and the experimental group before (T1) and after (T2) training

In order to further explore the effect of training on the identification of the /p/-b/ and /t/-/d/ contrasts, planned comparisons were performed on the mean slope coefficients and mean boundary values between and within groups using non-parametric tests, Mann-Whitney and Wilcoxon. Examination of the average logistic curves of the participants indicated steeper slopes for the /t/-/d/ than for the /p/-/b/ contrast (see Figure 1-2), suggesting that /t/-/d/ was labeled more sharply than /p/-/b/ along the VOT continuum.

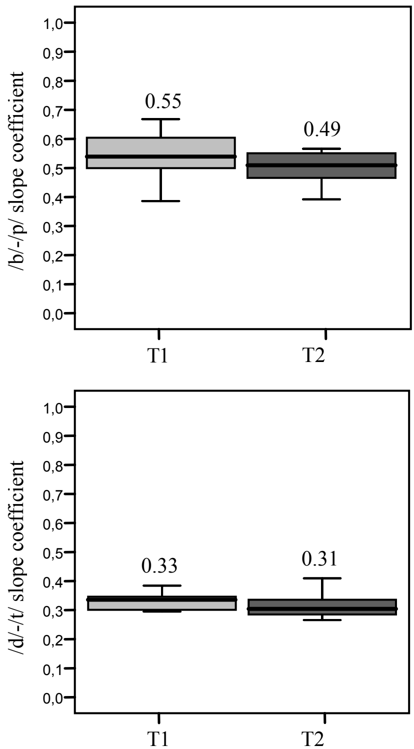

Figure 1-2. Mean slope coefficients of the experimental group for /b/-/p/ and /d/-/t/ identification before (T1) and after (T2) training

Comparisons between the mean slopes at T1 and T2, separately assessed for each participant and contrast, yielded no significant differences between experimental and control groups ($p > .05$). However, within-group comparisons revealed significant effects of the training on the slope values for the /p/-/b/ contrast ($Z=-2.330$, $p=.020$), but not for /t/-/d/ ($p>.05$). After phonetic training, the experimental group displayed significantly lower slope values for b/p identification, resulting in a less shallow slope indicating a slight increase in degree of categoriality in perception. Although the /t/-/d/ contrast is characterized by a higher degree of categoriality of responses at both T1 and T2 than the /p/-/b/ contrast, the slightly steeper slope obtained after training for /t/-/d/ did not reach statistical significance. Whereas the slope of the /p/-/b/ identification function of the experimental group increased significantly in steepness, no significant increase was observed for the NNS control group from T1 to T2 for any of the contrasts ($p>.05$).

Similarly, between-group comparisons performed on the boundary values of experimental and control group did not reach significance. The within-group analysis showed, however, significant differences in the boundary values of the experimental group between T1 and T2, for both /p/-/b/ ($Z=-2.417$, $p=.016$) and /t/-/d/ ($Z=-2.154$, $p=.031$) identification (Figure 1-3). As expected, the perceptual boundaries of the NNSs– located at lower VOT values when compared to NSs – were shifted to longer VOT values at T2, showing that phonetic training had a slight effect on the modification of non-native patterns of perceptual identification towards more English-like use of VOT as a cue for identifying word-initial stops. Tests conducted for the NNS control group showed no significant differences overall between their mean boundary values at T1 and T2.

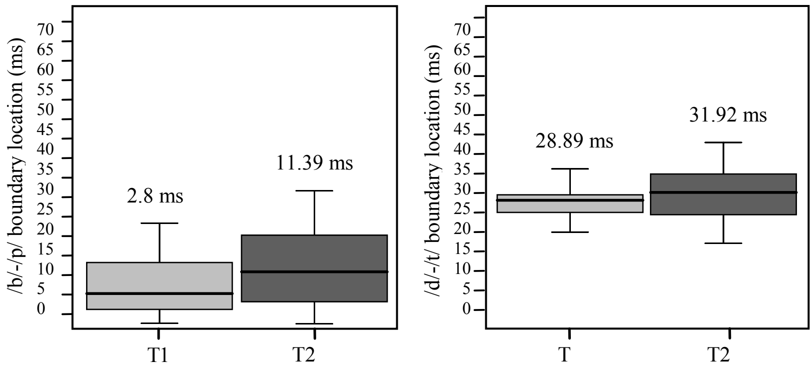

Figure 1-3. Mean boundary location (ms) of the experimental group for /b/-/p/ and /d/-/t/ identification before (T1) and after (T2) training

All in all, the Catalan/Spanish learners in the present study did not exhibit equal sensitivities to VOT in the perception of the contrasts /p/-/b/ and /t/-/d/. Whereas the alveolar stops were categorized far more consistently than the bilabial stops, significant differences between the experimental and NS control groups as regards the slope of the /t/-/d/ function at T1 ($Z=-4.237$, $p=.000$) persisted at T2 ($Z=-3.187$, $p=.001$). Paired comparisons between the experimental and the NS groups revealed that the NSs had mean boundaries located at longer VOT durations and steeper slopes than the NNSs. As opposed to the NSs, the NNSs do not perceive English stops with the same accuracy at different places of articulation, tending to have mean ID boundaries at shorter VOT locations, and with mean boundaries being less sharp and subject to more variability. However, modified boundary and slope values after training suggest that the Catalan/Spanish learners acquired a slightly higher sensitivity to VOT as a cue for identifying /p/-/b/ and /t/-/d/ as a result of phonetic training.

3.1.2 Effects of training on production

Accuracy in learners' production of /p/-/b/ and /t/-/d/ was assessed by means of VOT duration measures (2 words x 4 contexts x 3 repetitions x 2 data collection times x 3 subject groups = 864 VOT measurements per oral stop). 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 values obtained at T2 were higher than those at T1 (see Table 1-7 and Figure 1-4). 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/ ($t(23)=-6.24$, $p=.000$) and /t/ ($t(23)=-3.38$, $p=.003$) had disappeared after training according to the t-tests performed on the T2 data ($p> .05$), which confirms the positive impact of phonetic training upon mean VOT duration. Production of alveolar /t/−which is dentialveolar in the learners' L1 and therefore more dissimilar to English /t/ than Catalan/Spanish /p/ is to English /p/−showed less improvement at T2, as it was already pronounced more accurately before training. The fact that English and Catalan/Spanish /t/ do not share the same place of articulation probably prevents learners from directly assimilating English /t/ to their L1 /t/, thus promoting the formation of a new phonetic category for English /t/ and allowing them to produce it with more native-like VOT values. However, the mean VOT values obtained at T2 are still far from the NSs' average. The mean VOT for /p/ and /t/ in the present study was almost midway between the means observed for monolingual Spanish adults in Caramazza's (1973) study (26 ms) and the NS values from the present study.

Table 1-7. Mean VOT duration (ms) at T1 and T2 (*SD* in parenthesis) and significant differences (*) at α=.05

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

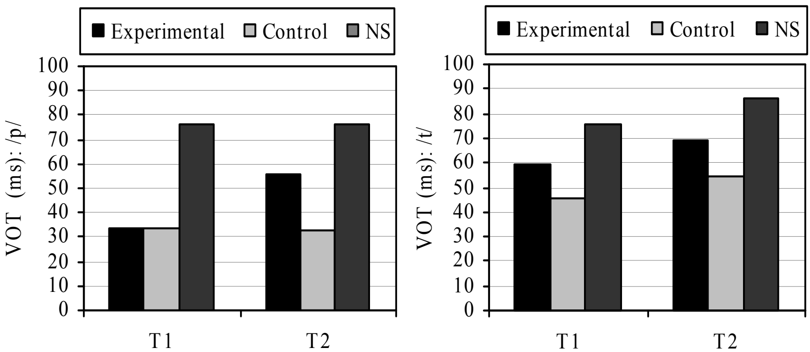

Figure 1-4. Mean VOT (ms) of /p/ and /t/ at T1 and T2

A paired-samples t-test confirmed that the production of /b d/ did not benefit from training (cf. Mackay, Flege, Piske and Schirru 2001). The experimental group showed a higher amount of prevoicing in /b/ and /d/ than NSs at T2, and the mean VOT durations of /b/ and /d/ at T1 and T2 (see Table 1-7) were not significantly different. Similarly, independent-samples t-tests showed no significant advantage for the experimental group over the control group in this respect (see Figure 1-5).

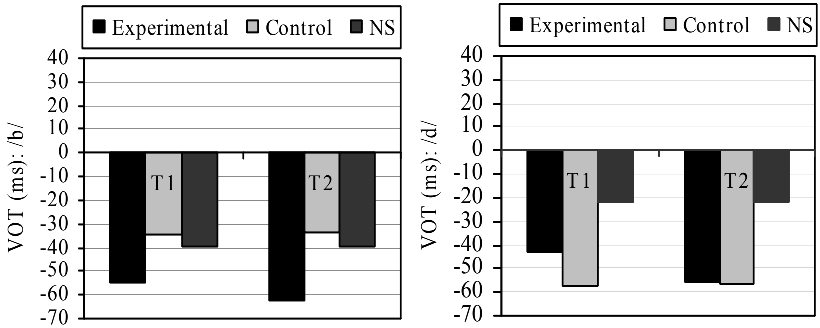

Figure 1-5. Mean VOT (ms) of /b/ and /d/ at T1 and T2

3.2 Perception and production of /iː/-/ɪ/ and /æ/-/ʌ/

3.2.1 Effects of training on perception

Mean percent correct vowel discrimination scores were computed for each subject and vowel contrast (see Table 1-8). As expected, overall significant T1-T2 differences were observed in the discrimination scores of the experimental group ($t(17)=-3.64$, $p=.002$), but not the control group, for both /iː/-/ɪ/ ($t(17)=-2.66$, $p=.017$) and /æ/-/ʌ/ ($t(17)=-3.32$, $p=.005$), which suggests that phonetic training had significantly improved learners' discrimination ability. A more detailed statistical analysis conducted for each of the vowel contrasts independently indicated a similar amount of significant increase for the /iː/-/ɪ/ contrast (8.75%) and the /æ/-/ʌ/ contrast (9.65%) from T1 to T2, although discrimination rates were higher for /æ/-/ʌ/ than for /iː/-/ɪ/.

Despite the fact that the experimental group improved significantly in their ability to discern the vowel contrasts, the significant differences between NSs' and NNSs' performance observed at T1 ($t(23)=-6.27$, $p=.000$) persisted at T2 ($t(23)=-4.22$, $p=.000$), suggesting that the Catalan/Spanish learners of English continued to perceive /iː ɪ æ ʌ/ in a nonnative-like manner after training (see Figure 1-6).

Table 1-8. Mean percent correct vowel discrimination (*SD* in parentheses) and significant differences (*) at α=.05

Perception: % correct vowel discrimination							
Group		Mean		/iː/-/ɪ/		/æ/-/ʌ/	
		T1	T2	T1	T2	T1	T2
NNS	Exp	70.45 (9.85)	*79.71 (9.07)	64.82 (12.26)	*73.57 (9.63)	75.21 (10.74)	*84.90 (11.19)
	Cont	69.32 (9.87)	75.52 (7.89)	62.26 (14.04)	69.15 (9.56)	75.29 (11.60)	80.89 (8.06)
NS		95.04 (4.66)		95.67 (4.90)		94.51 (5.00)	

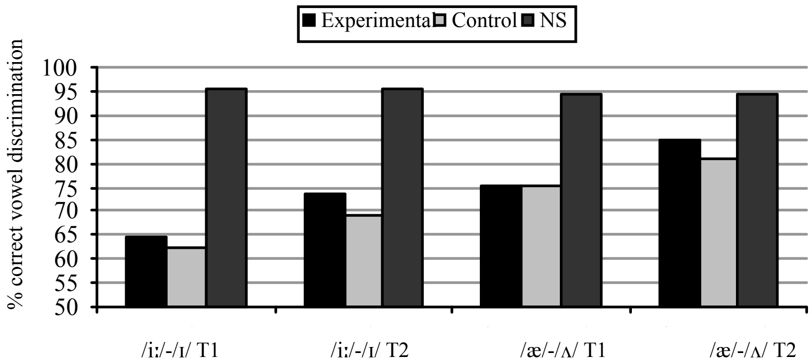

Figure 1-6. Mean percent correct vowel discrimination at T1 and T2.

3.2.2 Effects of training on production

Accuracy in L2 vowel production was assessed by means of F1 and F2 frequency (Hz) and duration measures (ms) (8 words x 3 repetitions x 2 data collection times x 3 subject groups = 1708 F1-F2 frequency and duration measurements). Formant frequencies were measured by obtaining the mean F1 and F2 frequencies of a 20 ms window placed at the midpoint of the steady state of the vowel. Spectral contrasts were then calculated by measuring the distance between the F1 and F2 values of the contrasting vowels, independently. Vowel duration was measured (ms) from the first positive peak in the periodic portion of the signal to the onset of the following consonant, for each vowel and environment. Formant frequency analysis was expected to reveal significant differences between groups, since training was likely to affect the degree of tongue height (F1) and the degree of tongue frontness/backness (F2), making learners produce the target vowels with more English-like frequency values (lower F1 values). The analysis of F1 and F2 differences between the two members of each vowel pair was expected to yield a larger distance between the contrasting vowels /iː/ and /ɪ/, and between /æ/ and /ʌ/, for the experimental group than for the control group at T2.

Formant frequency analysis showed no significant differences overall in L2 vowel production as a result of phonetic training (see Table 1-9), and revealed significant differences between experimental and NS control groups at T1 and T2, especially for the tense-lax contrast. In general, learners produced the target vowels with significantly higher F1 values than those of NSs, indicating that tongue position for L2 vowels was too low in the mouth. However, a closer inspection of F1 and F2 values revealed modest gains as regards the production of the tense-lax contrast on the part of the experimental group, but not the control group. Paired-samples t-tests showed that, in all cases, NNSs produced /iː/ ($t(16)=2.74$, $p=.015$) and /ɪ/ ($t(16)=2.38$, $p=.030$) with significantly lower F1 values, indicating that vowels were produced with greater height after training. On the other hand, no significant T1-T2 differences were found for the F1 and F2 values of /æ/ and /ʌ/ ($p>.05$).

Table 1-9. Mean F1-F2 vowel frequencies (Hz) and significant differences (*) at α=.05

Mean F1-F2 frequencies (Hz) of vowels					
Group		**/iː/**		**/ɪ/**	
		F1	**F2**	**F1**	**F2**
Experimental	T1	541.34	1590.17	553.86	1617.22
	T2	*515.41	1623.22	*525.03	1555.67
Control	T1	547.52	1647.18	546.99	1691.54
	T2	*495.79	1745.44	503.03	1787.44
NS		376.36	2260.67	476.07	1648.54
Group		**/æ/**		**/ʌ/**	
		F1	**F2**	**F1**	**F2**
Experimental	T1	928.66	1401.03	803.21	1350.85
	T2	908.33	1353.83	787.96	1326.73
Control	T1	890.47	1420.49	816.47	1337.45
	T2	879.85	1394.31	*773.97	1350.02
NS		814.83	1404.33	657.93	1223.74

Statistical analysis of the quality differences between the target vowel categories revealed significant differences between the experimental and NS groups for the F1 ($t(23)=-4.03$, $p=.001$) and F2 ($t(7.07)=-2.47$, $p=.043$) values of /iː/-/ɪ/, suggesting that, 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/ (see Table 1-10 and Figure 1-8).

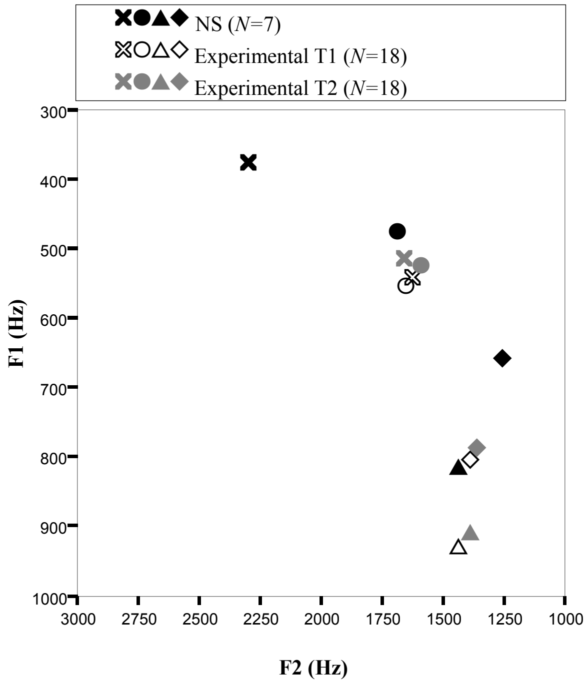

Figure 1-7. F1-F2 vowel formant plot for the English vowels /iː/ (✖), /ɪ/ (●), /æ/ (▲) and /ʌ/ (◆), pronounced by NSs (✖●▲◆) and NNSs at T1 (⌧○△◇) and T2 (✖●▲◆)