Effects of audiovisual auditory vs. articulatory training on L2 vowel category learning

Cristina Aliaga-García, University of Barcelona
cristinaaliaga@ub.edu

1. Abstract

Cross-language speech perception research has documented learners' difficulties in using and accessing L2 durational and spectral information. This study investigated the short-term effects of two types of audiovisual high-variability phonetic training -auditory vs. articulatory training- on the perception of the 11 English RP monophthongal vowels (/iː ɪ e ɜː æ ʌ ɑː ɒ ɔː ʊ uː/). Bilingual Catalan/Spanish learners of English (*N*=32 in each training condition) and one control group (*N*=20) were tested before and after training. Perceptual gains were assessed through a 4-choice categorization task with natural CVC words, a forced-choice categorization task based on 4 synthesized /h_d/ continua (/iː/-/ɪ/, /æ/-/ʌ/, /ʌ/-/ɑː/, /ʊ-/uː/), and an AX discrimination task including 16 contrasting pairs. Significant perceptual accuracy gains were obtained in both auditory and audio-visual conditions. Pre- and post- categorization of synthesized vowels suggests that vowel perception improved, and that both training modalities had a similar positive effect on the learners' use of spectral and duration information in vowel categorization. These results are further discussed in terms of input effects on vowel category learning and the re-weighting of acoustic cues through phonetic training.

2. Research question

The primary goal of this study is to investigate and assess the nature of the changes in vowel perception by L2 speakers following each of the treatment procedures, audiovisual auditory vs. articulatory training:

- ✓ Is the learning of new phonetic categories and fine-grained phonetic detail best promoted by exposing the learners to auditory or articulatory training?

3. Experimental design

The general design of this study has three phases: (1) a pre-test phase, (2) a perceptual training phase parallel to an articulatory training phase, and (3) a post-test phase.
A total of 84 participants (75 females, 9 males) between the ages of 20 and 50 (mean age 23.5) took part in the experiment: three groups of bilingual Catalan/Spanish undergraduate students of English Philology at the University of Barcelona.

Participants	Pre-Test (T1)	Audiovisual Phonetic Training		Post-Test (T2)	N
		Identification	Articulatory		
Experimental I	✓	✓		✓	32
Experimental II	✓		✓	✓	32
Control group	×	×		✓	20
Total					84

Table 1. Design of the study

4. Methodology: Pre- and post-tests

4.1. Identification task (ID I) with 264 isolated CVC natural words distributed into Audiovisual (AV) and Auditory (A) conditions (2 orders: A, AV and AV, A).
The task consisted of 66 trained and 66 untrained tokens in varied contexts recorded by 2 native speakers (1 male, 1 female).

The experimental and control participants were presented a total of 132 randomized stimuli (6 tokens x 11 vowels x 2 repetitions) per testing condition. After they had seen a 3-second video-clip of the talker's mouth, they had to choose among 3 or 4 possible responses based on one of the three vowel sets: high front, high back and low vowels (See *Figure 1*).

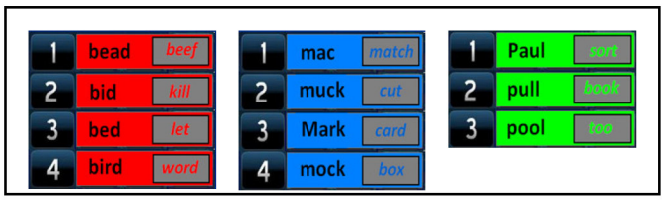


Figure 1. Identification task I with natural CVC words

4.2. Identification task (ID II) with 280 synthetic /h_d/ stimuli based on 8 seven-step continua with manipulated vowel duration.
The vowel continua were made from the vowel pairs /hiːd/-/hɪd/, /hæd/-/hʌd/, /hʌd/-/hɑːd/ and /hʊd/-/huːd/ by modifying vowel duration in 7 steps with Praat: 80, 116.7, 153.3, 190, 226.7, 263.3 and 300 ms. The result was 56 different stimuli (8 vowels x 7 steps) produced by one male talker and presented 5 times for identification. In the identification task the participants heard the stimuli one at a time and had to identify one member of the minimal pair (See *Figure 2*).

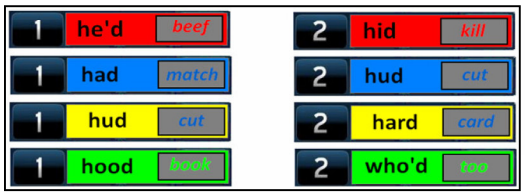


Figure 2. Identification task II with synthesized vowels

4.3. AX Discrimination task (DIS) with 128 natural vowel contrasts in varied contexts.
The experimental and control groups were tested on their discrimination of 16 vowel pairs (See *Table 2*) distributed into two blocks with 64 stimuli each (16 contrasting words produced by 8 talkers x 2 randomizations x 2 repetitions). 25% of the stimuli were same-word pairs (20 same-word pairs per block). The participants heard the vowel pairs and pressed either "same" or "different".

/iː ɪ/	/ɪ e/	/e ɜː/	/ɪ ɜː/	/e æ/	/ɜː æ/	/æ ʌ/	/ʌ ɑː/	/æ ɑː/	/ɜː ʌ/	/ɜː ɑː/	/ɒ ɑː/	/ʌ ɒ/	/ɒ ɔː/	/ʊ uː/	/ɔː uː/

Table 2. 16 vowel contrasts included in the discrimination task

5. Methodology: Audiovisual auditory vs. articulatory vowel training

- 10 one-hour training sessions designed and run using DMDX software and 10 different SBE speakers (5 m, 5f).
- 2 training sessions per week (over 6 weeks) with 2 different NS talkers (1m, 1f) per training session.
- The audiovisual training stimuli were CVC real words containing the 11 SBE monophthongal vowels in high contextual variation. Each training session consisted of 4 blocks and included the three subsets of vowels (high front vowels /iː ɪ e /, high back vowels /ɔː ʊ uː/, and low vowels /æ ʌ ɑː ɒ/) as pronounced by 2 different talkers (i.e. NS1 blocks 1 and 3; NS2 blocks 2 and 4).
- During the training session, the participants were exposed to a minimum of 176 trials (4 tokens x 11 vowels x 4 repetitions or blocks). Immediate feedback was given to

participants through their headphones after each response until they corrected themselves. Participants had a maximum of 4 trials.

5.1. Auditory training
In the identification training task the participants heard the whole set of vowels within a varied CVC context and watched short video-clips showing the articulation of words. On each trial, subjects clicked on 3 or 4 minimal pair alternatives. After errors, they heard the wrong response immediately followed by the correct word so that they could try again.

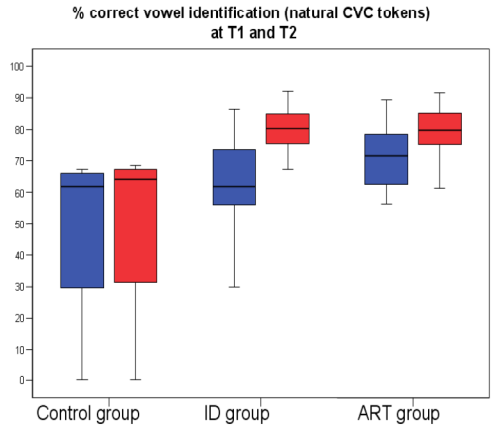
5.2. Articulatory training
In the imitation training task learners were presented words audiovisually so that they could focus on the speakers' mouth position for more accurate vowel articulation. After each section, they heard and compared their own responses with the native speakers' stimuli and then corrected themselves as many times as necessary.

6. Findings and discussion

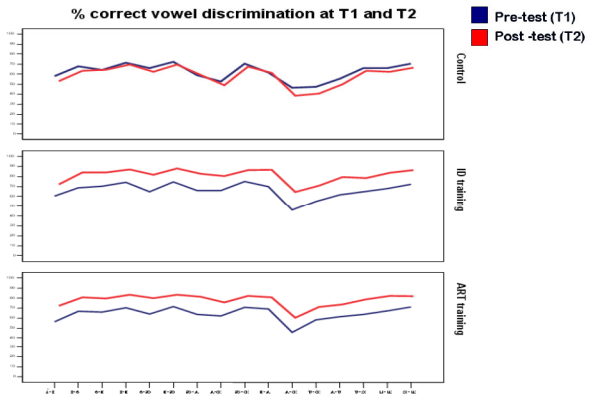
Both Auditory (ID) and Articulatory (ART) training modalities led to significant improvements in the perception of the British English RP vowels by the L2 learners.

6.1. Identification task I. Mean percent vowel ID scores were computed for each subject and vowel (See *Graph 1*). Overall significant T1-T2 differences were observed in the identification scores of the ID (T1=59.05, T2=80,16; t(31)=-5.24, p=.000) and the ART (T1=67.28, T2=79.18,16; t(31)=-3.48, p=.002) experimental groups, but not the control group, for both trained (T1=67.28, T2=79.18,16; t(31)=-3.48, p=.002) and untrained (ID: t(31)=-22.42, p=.000; ART: t(31)=-18.76, p=.000) tokens.

6.2. Discrimination task. Statistical analysis of L2 vowel DIS scores at post-test revealed no significant differences between the ID and ART experimental groups but paired comparisons conducted for each of the experimental groups showed significant T1-T2 differences for the 16 vowel contrasts tested (See *Graph 2*).

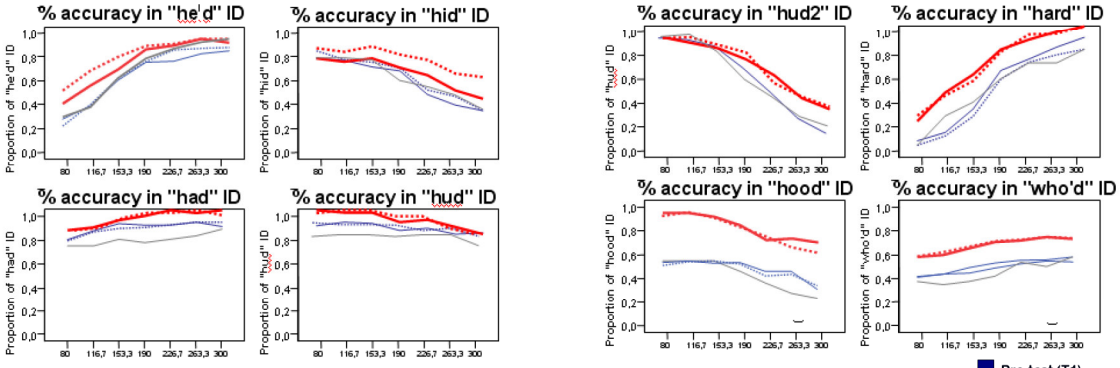


Graph 1. Mean % correct vowel ID



Graph 2. Mean % correct vowel contrasts DIS

6.3. Identification task II. The ID of vowels with manipulated duration at T2 suggests that the ID and ART groups did not rely on durational cues as a vowel differentiation cue as much as they did at T1. Significant T1-T2 differences in the ID accuracy of synthesized vowels for each of the groups and vowels can be used as evidence that laboratory training can change perceptual weightings or at least can help learners develop cue weightings for novel phonetic categories.



Graph 3. Mean % accurate synthesized vowel ID

References

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 171-204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 13-34). Amsterdam: John Benjamins.

Boersma, P., & Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.6.05) [Computer program]. Retrieved June 2, 2007, from http://www.praat.org/.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*, 977-985.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*, 371-387.

Escudero, P. (2000). Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels in Spanish speakers. Unpublished M. Sc. thesis, University of Edinburgh.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*, 1-34.

Flege, J. E. (1995). Second-language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 229-273). Timonium, MD: York Press.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength 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.

García Lecumberri, M. L., & Cenoz, J. (1998). Influencia de la duración en la adquisición de las vocales inglesas. In I. Vázquez, & I. Guillén (Eds.), *Perspectivas pragmáticas en lingüística aplicada* (pp. 201-207). Zaragoza: Textos de Filología 6, Asociación Española de Lingüística Aplicada (AESLA).

Iverson, P., & Evans, B. G. (2007). Auditory training of English vowels for first-language speakers of Spanish and German. *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1625-1628). Saarbrücken, Germany.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America*, *118*, 3267-3278.

Jongman, A., & Wade, T. (2007). Acoustic variability and perceptual learning: The case of non-native accented speech. In O.-S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 135-150). Amsterdam: John Benjamins.

Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 351-378). Timonium, MD: York Press.

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

Mora, J. C., & Fullana, N. (2007). Production and perception of English /iː/-/ɪ/ and /æ/-/ʌ/ in a formal setting: Investigating the effects of experience and starting age. *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1613-1616). Saarbrücken, Germany.