Perceptual and Ultrasound Articulatory Training Effects on English L2 Vowels Production by Italian Learners

I. Sonia d'Apolito, Bianca Sisinni, Mirko Grimaldi, Barbara Gili Fivela

***Abstract*—**The American English contrast /ɑ-ʌ/ (*cop-cup*) is difficult to be produced by Italian learners since they realize L2-/ɑ-ʌ/ as L1-/ɔ-a/ respectively, due to differences in phonetic-phonological systems and also in grapheme-to-phoneme conversion rules. In this paper, we try to answer the following research questions: Can a short training improve the production of English /ɑ-ʌ/ by Italian learners? Is a perceptual training better than an articulatory (ultrasound - US) training? Thus, we compare a perceptual training with an US articulatory one to observe: 1) the effects of short trainings on L2-/ɑ-ʌ/ productions; 2) if the US articulatory training improves the pronunciation better than the perceptual training. In this pilot study, 9 Salento-Italian monolingual adults participated: 3 subjects performed a 1-hour perceptual training (ES-P); 3 subjects performed a 1-hour US training (ES-US); and 3 control subjects did not receive any training (CS). Verbal instructions about the phonetic properties of L2-/ɑ-ʌ/ and L1-/ɔ-a/ and their differences (representation on F1-F2 plane) were provided during both trainings. After these instructions, the ES-P group performed an identification training based on the High Variability Phonetic Training procedure, while the ES-US group performed the articulatory training, by means of US video of tongue gestures in L2-/ɑ-ʌ/ production and dynamic view of their own tongue movements and position using a probe under their chin. The acoustic data were analyzed and the first three formants were calculated. Independent t-tests were run to compare: 1) /ɑ-ʌ/ in pre- *vs.* post-test respectively; /ɑ-ʌ/ in pre- and post-test *vs.* L1-/a-ɔ/ respectively. Results show that in the pre-test all speakers realize L2-/ɑ-ʌ/ as L1-/ɔ-a/ respectively. Contrary to CS and ES-P groups, the ES-US group in the post-test differentiates the L2 vowels from those produced in the pre-test as well as from the L1 vowels, although only one ES-US subject produces both L2 vowels accurately. The articulatory training seems more effective than the perceptual one since it favors the production of vowels in the correct direction of L2 vowels and differently from the similar L1 vowels.

***Keywords*—**L2 vowel production, perceptual training, articulatory training, ultrasound.

I. Introduction

LEARNING a second language (L2) in a formal context often implies that the target language is not widely used since teachers focus on lexical and grammatical information and, above all, they very often do not offer a native pronunciation of L2 sounds [1]. An exposure to such pronunciation of L2 impoverishes the correct perception of L2 sounds, and as a consequence their production, which is an important factor to categorize and discriminate the L2 contrasts [1]. However, findings on adults' perception of L2 contrasts suggest that non-native listeners can easily learn to perceive some L2 contrasts, in turn showing great difficulty with other ones. It depends on mother tongue (L1) phonology and phonetics, that is how similar or dissimilar L2 sounds are respect to native language ones [1]. In this respect, L2 pronunciation results widely influenced by L1 as well as by L2 perception. Nevertheless, production quality of non-native sounds by L2 learners can be improved by different training.

Some studies investigated (e.g. [2], [3]) the role of a combined approach using both pronunciation and perceptual training and results show that this approach involves improvements in production. However, few studies investigated (e.g. [4], [5]) the role of L2 pronunciation training only. Generally, a phonetic-pronunciation training is based on a visual feedback, direct or indirect, so that L2 learners can compare their production with the target sound produced by a native speaker [4]. Direct feedback provides dynamic view of the position and movements of articulators involved during the production of a specified sound. Ultrasound imaging [5], [6], electropalatography [7] or electromagnetic articulography [8] can be used to see directly tongue movements during speech. For instance, [5] observed the benefit of ultrasound visualization to improve French L2 high front/back /y-u/ contrast by Japanese learners. Results show an improvement in the production of the L2 /y-u/ contrast for the speakers who received ultrasound training. Indirect feedback provides raw acoustic features of L2 sound (spectrograms which show the first three formants) [9] or an abstract graphic representation of acoustic information which offers relevant phonetic differences between L1 and L2 sounds. For instance, [10] trained Korean leaners of English as L2 to produce the vowels /æ;'ʏə̨ –/ by videos of native speakers and an F1 (vertical dimension corresponding to vowel height)-F2 (horizontal dimension corresponding to tongue advanced/backness) vowel space of the English vowels and the closer Korean vowels /G; o/. Results show that leaners, after a five-hours training, improve the production of one vowel only, i.e. /æ/.

These studies show that feedback is very important for improving the accuracy of L2 sound production. However, it should be easily interpretable [4] and, for this reason, providing only acoustic information (spectrograms or graphic representation) may not be appropriate for learners who may find difficulties in transferring acoustic information onto articulatory movements [6]. On the contrary, approaches which provide articulatory information may be very useful for

I. S. d'Apolito, B. Sisinni, M. Grimaldi, and B. Gili Fivela are with University of Salento, CRIL (Centro di Ricerche Interdisciplinare sul Linguaggio), Italy (e-mail: sonia.dapolito@gmail.com, b.sisinni@gmail.com, mirko.grimaldi@unisalento.it, barbara.gili@unisalento.it).

learners since they face directly with tongue movements that cannot be seen directly otherwise [6]. In fact, movies displaying teacher's or a native speaker's tongue movements during speech as well as visual biofeedback of learner's own tongue moving in real-time on the screen can be very successful [11]. Moreover, ultrasound has been found to help improving pronunciation even after very short (30 min) training [12].

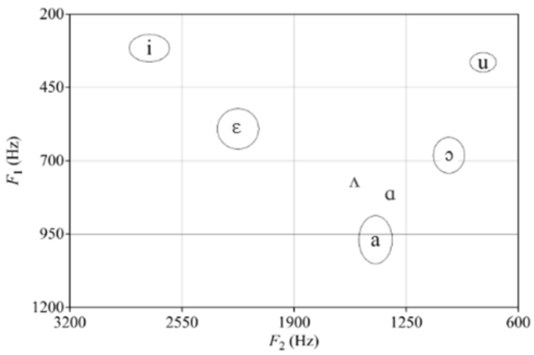


Fig. 1 Salento Italian acoustic vowel space (ellipses) and American English vowels /ɑ-ʌ/ produced by female native speakers (ellipses=±1 SD)

Pronunciation of L2 sounds has been shown to improve also by means of perceptual training, specifically the High Variability Phonetic Training (HVPT) method [2], [3], [13], [14]. This procedure consists in exposing subjects to highly variable natural productions. Subjects perform an identification task, i.e., they listen to a production and have to categorize it by considering two alternative sets/choices. Once they give their answer, feedback is provided. The minimal uncertainty of the two-alternative procedure and the immediate feedback, together with the high stimulus variability, promote the formation of new robust phonetic categories [15] and provide a broad base for generalization to new items and new talkers [16]. HVPT has a positive effect also on L2 production as shown in [3]. The authors found that perceptual learning is closely linked to production learning, although the two processes appear to be distinct within individual subjects [3]. Indeed, one of the most important factors to be taken into account in this domain is the considerable variation across subjects in performances in both perception and production.

The current pilot study focuses on the impact of two different trainings on the improvement of the pronunciation of the English contrast /ɑ-ʌ/ (e.g., cop-cup) by Italian learners of English as L2. This contrast is particularly difficult to be perceived and produced by Italian leaners [17], [18] since, according to the L1phonetic-phonological system, Italian learners produce /ɑ/-L2 as /ɔ/-L1 and /ʌ/-L2, often, as /a/-L1 [1], [18], [19]. The English orthography represents a further difficulty. Indeed, in Italian language the orthography is relatively transparent and native Italians tend to follow the orthographic form and, as a consequence, to mispronounce English written words which are phonologically opaque instead [20]. The aim of this study is to evaluate the effects of two different training, a perceptual and an US articulatory training, on the production of the L2-/ɑ-ʌ/ contrast. Both sounds differ according to tongue height and backness: 1) /ɑ/ has a more retracted tongue root and a lower tongue dorsum position; on the contrary, 2) /ʌ/ has a more anterior tongue root and a higher tongue dorsum position (see Fig. 1). Thus, US imaging can be useful since the target vowels are characterized by both tongue dorsum and root movements which can be easily visible with US (sagittal view).

II. Aim and Hypothesis

A perceptual and an articulatory (using ultrasound) training were performed and compared in order to observe: 1) the effects on L2-/ɑ-ʌ/ productions by Italian learners; and 2) if the US articulatory training can lead to better results than the perceptual training. Both trainings can involve short-terms effects as learners move from the most similar L1 sound toward a new vowel target position. However, we expect the US articulatory training to be more effective than the perceptual one since a direct feedback can help learners to understand how to position the tongue for the production of the L2 vowels, or at least to differentiate them adequately from L1 similar vowels. Moreover, the practice with US is expected to allow subjects to better control tongue position reaching a more stable tongue position. The hypotheses are tested by comparing participants' productions in pre/post-training tests.

III. Method

A. Subjects

Nine female subjects (mean age: 23, s.d. 0,63) from Salento (southern Apulia) participated in the experiment. All speakers are monolingual, they have never been in a foreign country for longer than a month, and they have started studying English as a foreign language at the mean age of 8.5. During their studies, they were exposed mainly to L1-accented English teachers. Participants were divided as follows: i) three experimental subjects performed the articulatory US training (ES-US) i.e., SPK1, SPK2, SPK3; ii) three experimental subjects performed the perceptual training only (ES-P), i.e., SPK4, SPK5, SPK6; and iii) three control subjects (CS) who did not receive any training, i.e., SPK7, SPK8, SPK9. In addition, one female native speaker of American English (AES; 21 years old, Oregon) was recorded to collect native speaker US data, and 5 American English natives were recorded to collect acoustic data for perceptual stimuli to be presented during the training.

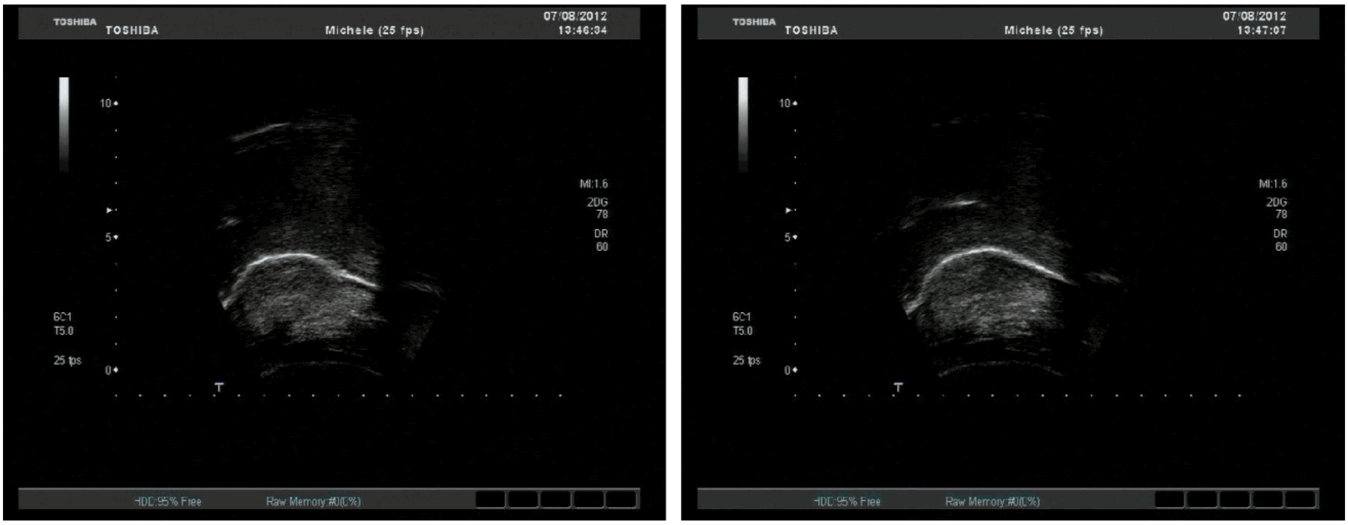
B. Training Procedures

A one-hour training session for each type of training was planned.

Both experimental groups received information about orthography, that is the grapheme-to-sound correspondences of the American English contrast, and phonetics, that is the phonetic differences between the non-native vowels /ɑ-ʌ/ as well as with respect to the closest native vowels /a-ɔ/. Graphic representations of an F1 (tongue height)-F2 (tongue backness) vowel space of the American English contrasts and of the Italian vowels were provided (Fig. 1) followed by verbal instructions about their articulation. Then the two experimental groups performed different trainings. Note that L2 vowels differ according to spectral properties and duration as /ɑ/ is longer than /ʌ/. We focused on vowel spectral properties only and, for this reason, the stimuli have been normalized for duration [17].

The ES-US group individually received an audio-articulatory training by means of the US machine (Toshiba Aplio XV) which offers a real-time biofeedback of learner's tongue position together with a real-time movie of the native speakers' tongue position, used as a visual model [11]. Native speaker production data were previously collected and analyzed and the best frames and videos were chosen to be presented to the ES-US group. Thus, the US training started showing the native speaker's movies (with audio) of tongue contour during the production of /ɑ-ʌ/ in isolation (Fig. 2), then during the production of CVC words (e.g. /bVb/, /pVp/, /mVm/ where V is /ɑ-ʌ/) and finally the videos of entire real sentences (e.g. "*I see /bVb/- /pVp/-/mVm/ inside*"). When subjects asserted to have detected the acoustic and articulatory differences between the target vowels and between them and the native vowels /a-ɔ/ on the basis of videos and audio, they started practicing the production of the L2 sounds with the US probe under their own chin. This procedure allowed them to see their own tongue profile on the screen and to adjust and control their tongue position according to the native speaker's tongue movement, as well as to sound perception. Explicit feedback was given to learners, that is positive feedback if they well matched the targets otherwise they were encouraged to try again if they were uncertain. During the practice, learners preferred to see US native's contour tongue movements as they were more intuitive and easily to replicate.

The ES-P group individually received a perceptual training, that is an identification test according to HVPT procedure. The perceptual training was performed by a web application expressly created by the laboratory staff (CRIL). The training duration was about 20 min. Participants were presented via headphones with one auditory stimulus at the time. The stimuli were English words /bVb/, /pVp/, /mVm/ (V=/ɑ-ʌ/), produced by 5 English speakers and presented 4 times each (filler words were also added; /bVb/, /pVp/, /mVm/ where V=/i-u/). All vowel duration was set at 200 milliseconds. Participants were asked to correctly associate the vowel sound to one of two non-orthographic symbols displayed on the computer screen ("^" was used for /ʌ/, "@" for /ɑ/, "!" for /i/ and "()" for /u/; the alternatives "@-^" were presented for /ɑ/ or /ʌ/ and "!-( )" for /i/ or /u/ -the order of symbols was random). Non-orthographic symbols were used to avoid any confounding effects from potentially inaccurate associations between sounds and English orthography [21]. If the response was correct, a positive feedback was given ("Correct!" appeared on the screen together with the picture of a happy Homer Simpson) and the next trial was presented. If the response was not correct, a negative feedback was provided ("Wrong!" appeared on the screen together with the picture of a doubtful Homer Simpson) and, before the next trial, participants were given the possibility to listen to the auditory stimulus once again; then they could click a button to go on. The next stimulus was always presented after a 500 ms delay. A total of 150 trials were used in the training session (2 target vowels x 5 talkers x 3 contexts x 4 repetitions + 2 control vowels x 3 talkers x 3 contexts).



(a) (b)

Fig. 2 US frames showing the tongue position of vowel /ʌ/ (a) and /ɑ/ (b) produced by the American native speakers

C. Speech Recordings and Analyses

Each recording session consisted in: i) pre-test data collection of L1 and L2 production (randomized across subjects); ii) one-hour training (for ES-US, ES-P; no training for CS), and iii) post-test data collection of L1 and L2 production.

The L1 corpus consisted of /pV_1pV_2/ words and pseudo-words, where V_1 was one of the five Italian native vowels /i, ɛ, a, ɔ, u/ and V_2 was /i/ or /a/, proposed in a carrier sentence (e.g. "*Dicevi pV$_1$pi in su*" or "*Diceva pV$_1$pa a Ken"*). The L2 corpus consisted of /pV_1p/ American English words, where V_1 was /ɑ/ or /ʌ/, inserted in a carrier sentence (e.g. "*I see pV$_1$p inside*"). All subjects read both corpora, displayed on a PC screen, 12 times. L1 and L2 productions of the three groups were collected in pre- and post-test sessions. Namely, the experimental groups' productions were collected before and after the training session, while the CS group's production was collected after a one-hour interval. Acoustic and articulatory data were collected simultaneously at CRIL laboratory in a soundproof room. A convex probe positioned under the subject's chin on the midsagittal plane was used to record US data at 25 Hz. A special stabilization set was used to fix the probe under the chin in order to restrict head movement. Acoustic data (22050 Hz, 16 bit) were analyzed using PRAAT [22] and the first three formants were calculated at the central 40% of the entire vowel duration [23]. Successively, F1 and F2 mean and standard deviation values of the pre- and post-test target vowels were plotted on a Cartesian F1-F2 plan for each subject. Data were statistically analyzed by means of a series of independent test ($p < 0.05$) in order to compare: 1) /ɑ-ʌ/ in pre- *vs* post-test respectively; 2) /ɑ-ʌ/ in pre- and post-test *vs* L1 /a-ɔ/ respectively. The more the target L2 vowels move in the acoustic space in the correct direction, the more the training will be effective. A restricted part of this study has been presented in [24], that is some results about the experimental group with articulatory training and the control group. Here, the experimental group, who performed the perceptual training, has been added and the comparison between articulatory *vs* perceptual training gives an important contribution to understanding the effectiveness of the articulatory training for improving the pronunciation of non-native sounds.

IV. Results

A. US Training Group

Fig. 3 shows results for native and non-native vowels produced by SPK1. As noted, in the pre-test she realizes L2-/ʌ/ as an L1-/a/, even if a slightly lower one [$t(22)=3{,}333$ $p<0.05$]. In the post test, she raises and advances her tongue to reach the new target position and, indeed, her vowel differs significantly from both L1-/a/ and L2-/ʌ/ as realized in the pre-test [/ʌ/-/a/ F1 $t(23)=-13{,}132$ $p<0.05$; F2 $t(23)=10{,}006$ $p<0.05$; /ʌ/-/ʌ/ pre-test F1 $t(23)=13{,}207$ $p<0.05$; F2 $t(23)=-10{,}305$ $p<0.05$]. In her productions, the non-native vowel /ɑ/ is realized very close to (just slightly lower and more anterior than) the L1-/ɔ/ in pre-test [F1 $t(22) = 6{,}417$ $p<0.05$; F2 $t(22) = 9{,}403$ $p < 0.05$]. On the contrary, in the post-test the produced vowel differs significantly from both L1-/ɔ/ and L2-/ɑ/-pre-test one, being realized as a lower and more anterior vowel [/ɑ/-/ɔ/ F1 $t(23) = 23{,}763$ $p < 0.05$; F2 $t(23) = 15{,}752$ $p = 0.05$; /ɑ/-/ɑ/ pre-test F1 $t(23) = -22{,}540$ $p < 0.05$; F2 $t(23) = -10{,}780$ $p < 0.05$]. Thus, after training, SPK1 seems to be able to realize the main characteristics of the non-native L2 sounds since, in her post-test productions, the tongue position for /ʌ/ is higher and more anterior than that for L1-/a/, while for /ɑ/ her tongue is lower and more anterior than in L1-/ɔ/.

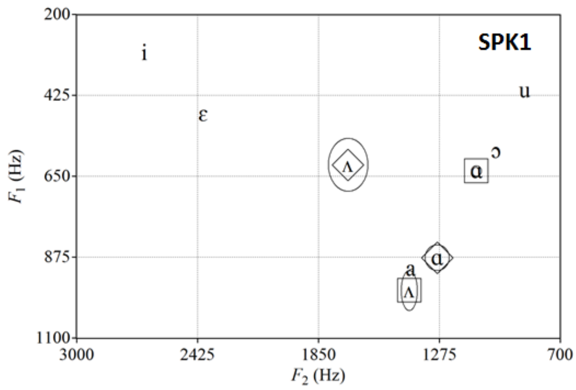


Fig. 3 Native and L2 vowels produced by SPK1 in the pre-test (square) and the post-test (diamond) (ellipses = ±1 SD)

SPK2 in pre-test does not produce any difference between L1-/a/ and L2-/ʌ/, while after the training, it differs significantly for F1 and F2 as she raises and advances the tongue towards the new target position [/ʌ/-/a/ F1 $t(23) = -7{,}592$ $p = 0.05$; F2 $t(23) = 8{,}281$ $p < 0.05$; /ʌ/-/ʌ/ pre-test F1 $t(22) = 7{,}369$ $p < 0.05$; F2 $t(22) = -9{,}393$ $p < 0.05$] – see Fig. 3. Like SPK1, in pre-test she produces L2-/ɑ/ close to L1-/ɔ/, though as lower and more anterior [F1 $t(21) = 4{,}225$ $p < 0.05$; F2 $t(22) = 8.930$ $p < 0.05$] while, after the training, her L2-/ɑ/ vowels are only slightly more anterior than L1-/ɔ/ [/ɑ/-/ɔ/ F2 $t(22) = 14{,}436$ $p < 0.05$; /ɑ/-/ɑ/ pre-test F2 $t(23) = -11{,}938$ $p < 0.05$]. SPK2, therefore, seems to produce correctly L2-/ʌ/ getting a higher and a more anterior tongue position than L1-/a/, but not L2-/ɑ/ which is too close to L1-/ɔ/ even if significantly different as for F2. Results are represented in Fig. 4.

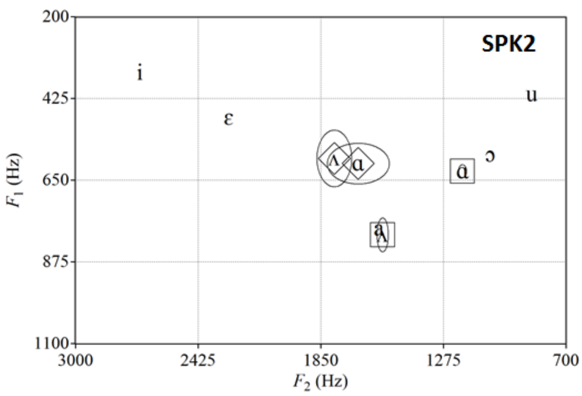


Fig. 4 Native and L2 vowels produced by SPK2 in the pre-test (square) and the post-test (diamond) (ellipses=±1 SD)

As shown in Fig. 5, in pre-test SPK3 realizes both L2-/ʌ/ and /ɑ/ nearly as L1-/a/, but significantly more open [F1: /ʌ/-/a/ $t(19) = 4.582$ $p < 0.05$; /ɑ/-/a/ $t(23) = 2.392$ p 0 <0.05]. In post-test productions, she realizes L2-/ʌ/ differently from L1-/a/, lowering and advancing the tongue position, as well as from L2-/ʌ/ in pre-test, though along the F2 dimension only [/ʌ/-/a/ F1 t(22) = 6,095 p < 0.05; F2 t(22) = 9,264 p < 0.05; F2 /ʌ/-/ʌ/ pre-test t(21) = -7,282 p < 0.05]. As regards L2-/ɑ/, it is realized with a lowering of the tongue respect to L1-/a/ and L2-/ɑ/ pre-test [/ɑ/-/a/ F1 t(23) = 12,147 p < 0.05; /ɑ/-/ɑ/ pre-test F1 t(26) = -8,346 p < 0.05]. Consequently, SPK3 is able to move away from native /a/ for both L2 sounds but they keep realizing in the acoustic space below L1-/a/.

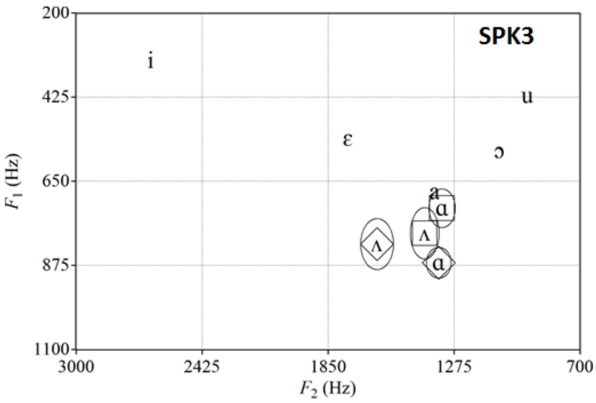


Fig. 5 Native and L2 vowels produced by SPK3 in the pre-test (square) and the post-test (diamond) (ellipses=±1 SD)

These results, above all in the pre-test, show the influence of the orthography for two out of three learners as the grapheme -o- is always associate to L1-[ɔ] sound. Only SPK3 realize, in both tests, the grapheme -o- near to L1-[a] sound due to her acoustic space.

B. Perceptual Training Group

In the pre-test, SPK4 realizes the non-native /ʌ/ close to L1-/a/, even if F2 is significantly higher [F2 t(22)= 4.618, p < 0.05]. In the post-test, on the contrary, no significant difference between L2-/ʌ/ and L1-/a/ is found. As for L2-/ʌ/ realized in the pre- and post-test, only F1 differs [F1 t(20) = 3.774 p < 0.05]. With regards to L2-/ɑ/, in the pre-test it is realized close to L1-/ɔ/ (just showing higher mean F1 values [F1 t(22) = 4.928 p < 0.05]), while in the post-test, no significant difference between L2-/ɑ/ and L1-/ɔ/ is found. Finally, L2-/ɑ/ pre- and post-test differs only along F2 [t(25) = -2.066 p < 0.05]. Except for some slightly differences, SPK4 realizes /ʌ/ close to L1-/a/ and L2-/ɑ/ close to L1-/ɔ/ even after training. Results are reported in Fig. 6.

In the pre-test, SPK5 realizes L2-/ʌ/ quite close to L1-/a/, even if F2 differs significantly [F2 t(22)=-2.539 p<0.05]. After the training, /ʌ/ differs for both dimensions being higher and more anterior than L1-/a/ [F1 t(14) = -4.070 p < 0.05; F2 t(14) = 3.943 p < 0.05]. Indeed, F1 and F2 differ significantly between pre- and post-test [F1 t(23) = 3.230 p < 0.05; F2 t(23) = -6.065 p < 0.05]. In terms of mean and standard deviation values, F1 is 778 Hz (st. dev. 40) and 640 Hz (st. dev. 147) in the pre- and post-test respectively; while F2 is 1389 Hz (st. dev. 100) and 1826 Hz (st. dev. 237). The standard deviation values are too high in the post-test, and this may suggest that L2 vowel production varies a lot across the vowel space. As regards L2-/ɑ/, in the pre-test it is realized quite close to L1-/ɔ/ though F2 is significantly different [F2 t(21) = 2.625 p < 0.05]. In the post-test, /ɑ/ is realized as a lower and more anterior vowel than L1-/ɔ/ and /ɑ/-pre-test and, indeed, F1 and F2 are both significantly different [L1-/ɔ/: F1 t(13) = 4.063 p < 0.05; F2 t(13) = 2.743 p < 0.05; /ɑ/-pre-test: F1 t(21) = -6.973 p < 0.05; F2 t(21) = -4.642 p=,005]. Observing Fig. 7 it is possible to note that in the post-test /ɑ/ is in-between L1/a/ and /ʌ/-pre-test. In any case, /ɑ/ differs significantly only from L1-/a/ and only for F1 being closer [F1 t(13) = -1.764 p < 0.05]. Also for /ɑ/, standard deviation values are too high in the post-test. Indeed, F1 is 587 Hz (st. dev. 31) and 741 Hz (st. dev. 66) in the pre- and post-test respectively; while F2 is 1298 Hz (st. dev. 23) and 1105 Hz (st. dev. 57) and 1435 Hz (st. dev. 231) in the pre- and post-test respectively. Thus, results suggest that SPK5 has been able to produce L2 sounds away from L1 sounds. However, as noted before, the realization of both L2 sounds in the post-test presents a great variability (high standard deviation values) suggesting a high degree of uncertainty in L2 vowel articulation.

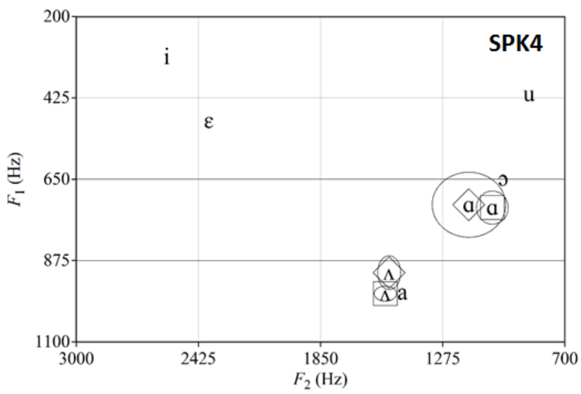


Fig. 6 Native and L2 vowels produced by SPK4 in the pre-test (square) and the post-test (diamond) (ellipses=±1 SD)

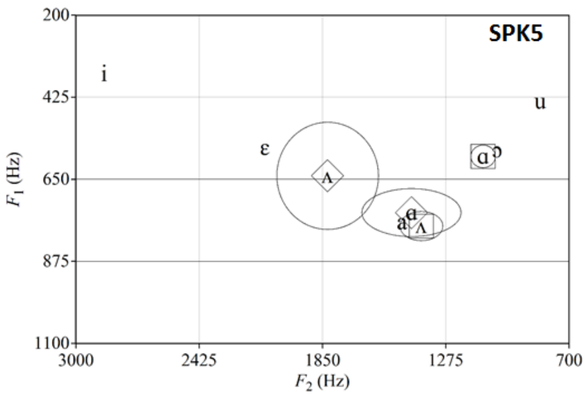


Fig. 7 Native and L2 vowels produced by SPK5 in the pre-test (square) and the post-test (diamond) (ellipses=±1 SD)

SPK6 realizes /ʌ/ in the pre-test close to L1-/a/, with closer being F1 significantly lower [F1 t(21) = -6.468 p < 0.05]. After the training, it differs along both dimensions because it is realized as a closer and more posterior vowel [F1 t(12) = -4.450 p < 0.05; F2 t(12) = -2.429 p < 0.05]. L2-/ʌ/ between pre- and post-test differs only for F2 being more posterior in the post-test [F2 t(20) = -5.977 p < 0.05]. As for /ɑ/, in the pre-test it is close to L1-/ɔ/, though F2 is significantly higher [F2 t(23) = 4.443 p < 0.05]. In the post-test, /ɑ/ is realized as a lower and more anterior vowel L1-/ɔ/ with both F1 and F2 showing statistically different values [F1 t(14) = 7.429 p < 0.05; F2 t(14) = 4.216 p < 0.05]. L2-/ɑ/, from pre-test to post-test has been realized modifying significantly F1 since the L2 vowel shows a greater opening [F2 t(24) = -3.926 p < 0.05]. Thus, the learner realizes some changes for the production of both non-native sounds, but the main phonetic differences between /ʌ/ and /ɑ/ and between them and their L1 counterparts have not been detected. This is true above all for F1, because in the post-test /ʌ/ and /ɑ/ are realized too close to each other and in-between L1-/a/ and /ɔ/.

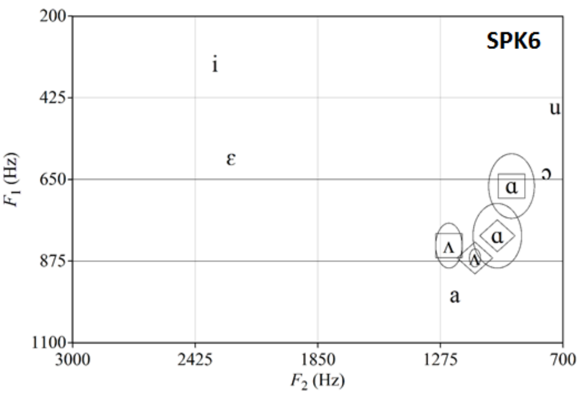


Fig. 8 Native and L2 vowels produced by SPK6 in the pre-test (square) and the post-test (diamond) (ellipses=±1 SD)

As regards the influence of the orthography, all speakers read the grapheme -o- as L1-[ɔ] in the pre-test, and for SPK4 even in the post-test.

C. Control Group

Fig. 9 shows the F1-F2 plans for SPK7 (Fig. 9 (a)) and SPK8 (Fig. 9 (b)) of the control group. SPK7 and SPK8 produce the non-native /ʌ/ significantly different from both L1-/a/ [SPK7: F1 t(22) = -4.127 p < 0.05; F2 t(22) = -7,162 p < 0.05; SPK8: F1 t(19) = -6.437 p < 0.05; F2 t(19) = -8,555 p < 0.05] and L1-/ɔ/ [SPK7: F1 t(21) = 10,144 p < 0.05; F2 t(21) = 9.344 p < 0.05; SPK8: F1 t(20) = 8.543 p < 0.05; F2 t(20) = 11,855 p < 0.05]. The same vowel is realized in the post-test [/ʌ/-/a/ SPK7: F1 t(21) = -9.660 p < 0.05; F2 t(21) = -17.480 p < 0.05; /ʌ/-/ɔ/ F1 t(20) = 7.970 p < 0.05; F2 t(20) = 13.311 p < 0.05; /ʌ/-/a/ SPK8: F1 t(21) = 11.743 p < 0.05; F2 t(21) = 6.006 p < 0.05; /ʌ/-/ɔ/ F1 t(22) = 6.968 p < 0.05; F2 t(22) = 11.020 p < 0.05]. As regards /ʌ/ realized in pre- and post-test, only F1 differs significantly, as the vowel is realized with a higher tongue position [SPK7: F1 t(21) = 4.257 p < 0.05; SPK8 F1 t(20) = 3.492 p < 0.05]. On the contrary, they produce L2-/ɑ/ close to L1-/ɔ/, even if SPK7 realizes it as slightly more anterior and lower in the pre-test [SPK7: F1 t(21) = 2.178 p < 0.05; F2 t(21) = 4.058 p < 0.05]. In the post-test, both speakers realize L2-/ɑ/ as a lower and more advanced vowel [SPK7: F1 t(23) = 3.182 p < 0.05; F2 t(23) = 4.627 p < 0.05; SPK8 F1 t(22) = 5.787 p < 0.05; F2 t(22) = 5.495 p < 0.05]. Even though there are some differences between pre- and post-test, the non-native sounds are not realized differently enough to suggest that a new target position has been reached.

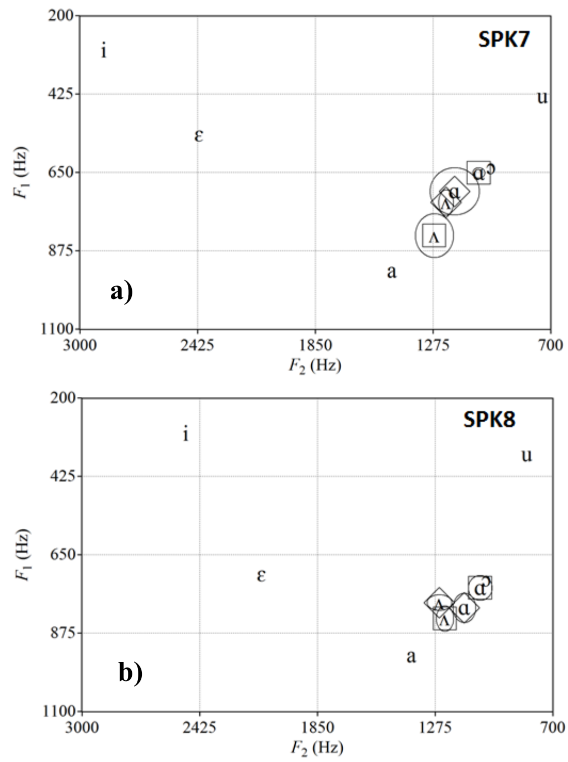


Fig. 9 Native and L2 vowels produced by SPK7 (a) and SPK8 (b) in the pre-test (square) and the post-test (diamond) (ellipses=±1 SD)

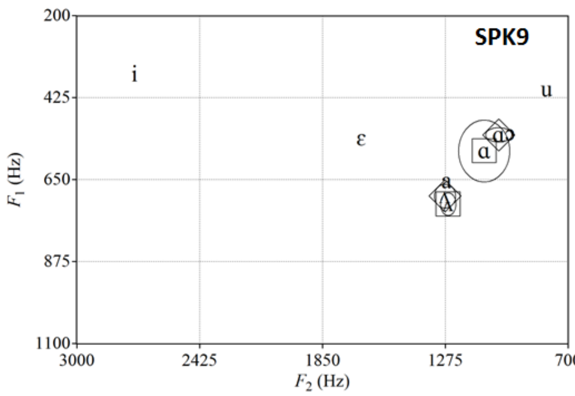


Fig. 10 Native and L2 vowels produced by SPK9 in the pre-test (square) and the post-test (diamond) (ellipses=±1 SD)

Finally, SPK9 realizes L2-/ʌ/ close to L1-/a/ even if significantly lower in both pre-test [t(19) = 3.154 p < 0.05] and post-test [t(22) = 2.292 p < 0.05]. L2-/ɑ/ is realized significantly lower and more anterior than L1-/ɔ/ in the pre-test [F1 t(26) = 2.227 p < 0.05; F2 t(26) = 3.763 p < 0.05], and only slightly more anterior in the post-test [F2 $t(22) = 2.470$ $p < 0.05$]. Thus, SPK9 do not differentiate significantly the non-native vowels from the native ones. Results are represented in Fig. 10.

The influence of the orthography is clear for control subjects who read the grapheme -o- as L1-[ɔ] in both pre- and post-test.

V. Discussion and Conclusion

The American English contrast /ɑ-ʌ/ is particularly difficult to be produced by Italian learners [8] due to Italian phonetic-phonological system and also to orthography opacity [20].

Our research focuses on identifying which training, a perceptual *vs* an articulatory one, can lead to a better pronunciation of the L2-/ɑ-ʌ/ vowels. For this reason, two short training (a one-hour training session each) were performed: 1) a perceptual training: an identification test according to HVPT procedure; and 2) an articulatory training: an audio-articulatory training by means of both US videos of the native's production in different contexts and real-time feedback of learners' tongue movement. Both training can have an effect on the improvement of the pronunciation but we expect the articulatory training to be more effective.

In the pre-test, regardless of the group, all speakers realize the L2-/ʌ/ as similar to L1-/a/ and the L2-/ɑ/ as L1-/ɔ/. In the post-test, the subjects of the control group (CS), who did not receive any training, keep realizing the contrast as they do in the pre-test. On the contrary, some learners of the other two groups systematically and significantly vary their production strategies. In particular, two out of three learners - SPK1 and SPK2 - after the articulatory training are be able to move away from the L1 counterparts consistently with the L2 targets. Both learners move the tongue in the right direction for the L2-/ʌ/ since they rise and advance their tongue position. As for L2-/ɑ/, only SPK1 correctly changes the tongue position lowering it. SPK3, on the contrary, realizes the contrast in the pre- and post-test in the acoustic space below L1-/a/, perhaps because the phonetic space above it, that is among L1-/ɛ, a, ɔ,/, is too restricted to move the tongue in order to differentiate L2 vowels from native ones. As regards the learners who received the perceptual training, only SPK5 is able to produce L2 vowels differently from the native ones, "moving" in the correct direction for both L2 sounds. Indeed, she produces L2-/ʌ/ advancing and rising the tongue, and L2-/ɑ/ lowering and advancing the tongue even if it results to be too closer to L1-/a/. However, her changes show a great variability in terms of standard deviation values which indicates that the subject approaches the L2 sounds as a new goal each time rather than as a systematic attempt to realize different vowel targets as the subjects of the US training group seem to do. SPK6 realizes L2-/ʌ/ moving the tongue too back, close to L2-/ɑ/, and thus neutralizing the differences between them. SPK4's productions reflect those of control subjects as, after the training, L2 sounds are realized nearly as the L1 counterparts.

Finally, results also show a strong influence of orthography as the grapheme <o> is read as [o].

As expected, the articulatory training seems to be very useful to improve learners' productions of the AE contrast and its effects are more effective and more stable than those related to the perceptual one. Indeed, learners are able to get the same tongue position throughout the repetitions as showed by low standard deviation values. To some extent also the perceptual training leads to an improvement of the L2 sounds, at least for one speaker who shows a great variability, revealing different attempts at producing L2 sounds. However, more subjects are necessary to validate our results as well as longer trainings and/or multiple sessions in order to improve the fine control of tongue movement to get more precise tongue position. Further researches are needed to observe long-term effects, the motor control reorganization and its effects on the knowledge at the phonological level.

References

[1] C. T. Best, M. D. Tyler, "*Nonnative and Second Language Speech Perception: Commonalities and Complementarities*", in Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production, Amsterdam, John Benjamins, 2007, pp. 13-34.

[2] J. S. Logan, S. E. Lively, D. B. Pisoni, "Training Japanese listeners to identify English /r/ and /l/: A first report", in *The Journal of the Acoustical Society of America*, vol. 89(2), 1991, pp. 874-886.

[3] A. R. Bradlow, D. B. Pisoni, R. Akahane-Yamada, Y. I. Tohkura, "Training Japanese listeners to identify English /r/ and /l/. IV: some effects of perceptual learning on speech production", in *The Journal of the Acoustical Society of America*, vol. 101(4), 1997, pp. 2299-2310.

[4] N. Kartushina, A. Hervais-Adelman, U. H. Frauenfelder, & N. Golestani, "The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds", *Journal of Acoustical Society of America*, 138(2), 2015, pp. 817-832.

[5] C. Pillot-Loiseau, T. K.Antolík, T. Kamiyama, "Contribution Of Ultrasound Visualization To Improving The Production Of The French /y/-/u/ Contrast By Four Japanese Learners", in *Phonetics, Phonology, Languages In Contact. Contact: Varieties, Multilingualism, Second Language Learning*, 2013, pp. 86-89.

[6] S. M. Wilson, B. Gick, "Ultrasound Technology and Second Language Acquisition Research", in *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference*, 2006, pp. 148–152.

[7] A. M. Schmidt, & J. Beamer, "Electropalatography treatment for training Thai speakers of English", in *Journal of Clinical Linguistics & Phonetics*, vol. 12 (5), 1998, pp.389-403.

[8] A. Suemitsu, & J. Dang, T. Ito, M. Tiede, "A real-time articulatory visual feedback approach with target presentation for second language pronunciation learning", in *Journal of Acoustical Society of America*, vol. 138 (4), 2015, pp. 382-387.

[9] R. Akahane-Yamada, E. McDermott, T. Adachi, H. Kawahara, J. S. Pruitt, "Computer-based second language production training by using spectrographic representation and HMM-based speech recognition scores", in *Proceedings of ICSLP*, 1998, Sydney, Australia.

[10] M. Carey, "CALL visual feedback for pronunciation of vowels: Kay Sona-Match", in *CALICO Journal*, vol. 21 (3), pp. 571-601.

[11] I. Wilson, "Using Ultrasound For Teaching And Researching Articulation", in *Acoustical Science & Technology*, vol. 35(6), 2014, pp. 285-289.

[12] B. Gick, B. M. Bernhardt, P. Bacsfalvi, I. Wilson, "Ultrasound imaging applications in second language acquisition", in *Phonology and Second Language Acquisition*, 2008, pp. 309-322.

[13] P. Iverson, & B. G. Evans, "Learning English vowels with different first language vowel system II: Auditory training for native Spanish and German speakers" in *The Journal of the Acoustical Society of America*, vol. 126, 2009, pp. 866-877.

[14] S. Ylinen, M. Uther, A., Latvala, S. Vepsäläinen, P. Iverson, R. Akahane-Yamada, R. Näätänen, "Training the brain to weight speech cues differently: A study of Finnish second-language users of English, in *Journal of Cognitive Neuroscience*, vol. 22(6), 2010, pp. 1319-32.

[15] S. E. Lively, D. B. Pisoni, R. A. Yamada, Y. Tokhura, and T. Yamada, "Training Japanese listeners to identify English /r/ and /l/ III. Long-term retention of new phonetic categories", in *The Journal of the Acoustical Society of America*, vol. 96, 1994, pp. 2076-2087.
[16] M. Posner, and S. Keele, "On the genesis of abstract ideas", in *Journal Experimental Psychology*, vol. 77, 1968, 353-363.
[17] P. Escudero, T. Benders, K. Wanrooij, "Enhanced Vowel Distributions Facilitate The Learning Of Second Language Vowels", in *The Journal of the Acoustical Society of America,* vol. 130(4), 2011, pp. El206-El212.
[18] J. E. Flege, I. Mackay, & D. Meador, "Native Italian Speakers' Production And Perception Of English Vowels", in *The Journal of the Acoustical Society of America*, Vol. 106, 1999, pp. 2973-2987.
[19] P. Escudero, B. Sisinni, M. Grimaldi, "The Effect Of Vowel Inventory And Acoustic Properties In Salento Italian Learners Of Southern British English Vowels", in *The Journal of the Acoustical Society of America*, vol.135(3), 2014, pp. 1577-1584.
[20] B. Bassetti & N. Atkinson, "Effects Of Orthographic Forms On Pronunciation In Experienced Instructed Second Language Learners", in *Orthographic Effects In Second Language Phonology. Special Issue. Applied Psycholinguistics*, vol. 36(1), 2015, pp. 67-91.
[21] R.I. Thomson, "Computer Assisted Pronunciation Training: Targeting second language vowel perception improves pronunciation" *in CALICO Journal*, vol. 28(3), 2011, pp. 744-765.
[22] P. Boersma, and D. Weenink, "Praat: Doing phonetics by computer" (computer program), Version 6.0, 2016. http://www.praat.org/
[23] K. Chládková, P. Escudero, P. Boersma, "Context-Specific Acoustic Differences Between Peruvian And Iberian Spanish Vowels", *in The Journal of the Acoustical Society of America*, Vol. 130, 2011, pp. 416–428.
[24] B. Sisinni, I. S. d'Apolito, B. Gili Fivela, M. Grimaldi, Ultrasound articulatory training for teaching pronunciation of L2 vowels, in Proceedings of the 9th International Conference ICT for Language Learning, 17-18 November, Florence, Italy, 2016, pp. 265-270.