Supervision hampers distributional learning of vowel contrasts

Gulian, M.; Escudero, P.; Boersma, P.

Publication date
2007

Document Version
Final published version

Published in
Proceedings of the 16th International Congress of Phonetic Sciences

SUPERVISION HAMPERS DISTRIBUTIONAL LEARNING OF VOWEL CONTRASTS

Margarita Gulian, Paola Escudero and Paul Boersma

University of Amsterdam
margarita_gulian@abv.bg, paola.escudero@uva.nl, paul.boersma@uva.nl

ABSTRACT

We investigate how supervision (in the form of explicit instruction) interacts with distributional learning in the acquisition of the perception of a novel vowel contrast in a second language. An experiment with non-Dutch-speaking Bulgarians reveals that listeners who receive bimodal distributional training *without* explicit instruction can acquire new Dutch vowel contrasts, and that listeners who receive the same training *with* explicit instruction do not acquire the new contrasts nearly as well. We conclude that explicit instruction hampers distributional learning.

1. INTRODUCTION

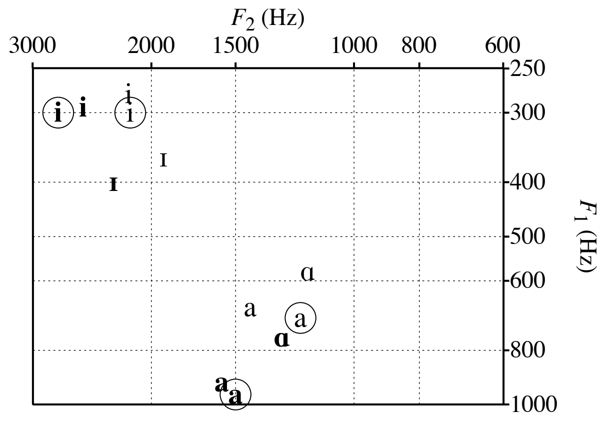
L2 learners have difficulties when perceiving phonological contrasts that are not present in their native language. The question for language pedagogy is what an appropriate method is for helping these learners. Earlier research [5] has shown that simply presenting sounds in a bimodal (two-peaked) distribution on an auditory continuum is enough to enable learners to distinguish between two different sound categories that are not present in their native language. This means that a mere statistical distribution of phonetic tokens can lead to an improvement in L2 sound perception. However, many language teachers are of the opinion that explicit instruction about the number of phonetic categories to be acquired is even more beneficial than training with bimodal distributions.

This study investigates the interaction of distributional learning and explicit instruction in the learning of non-native contrasts. To this end, Bulgarian listeners are trained and tested in their perception of two Dutch vowel contrasts not present in Bulgarian.

2. TWO NEW CONTRASTS

The two Dutch vowel contrasts used in this study are /a/~/ɑ/ (with duration equalized) and /i/~/ɪ/. Figure 1 (uncircled symbols) shows formant values for these four vowels, for both males and females [1]. In these two regions of the vowel space, Bulgarian has only two vowels, /i/ and /a/, whose formant values are shown in Figure 1 as well (circled symbols), for both males and females [7]. A comparison between the two languages indicates that Bulgarian /i/ lies closer to Dutch /i/ than to Dutch /ɪ/, and Bulgarian /a/ lies closer to Dutch /a/ than to Dutch /ɑ/; however, since Bulgarian does not have any vowels closely resembling Dutch /ɪ/ and /ɑ/, Bulgarian listeners are expected to perceive Dutch /ɪ/ and /ɑ/ as (perhaps less typical variants of) Bulgarian /i/ and /a/, respectively. This predicts that Bulgarians will have trouble distinguishing Dutch /i/ from Dutch /ɪ/ and Dutch /a/ from Dutch /ɑ/. The Dutch /a/~/ɑ/ and /i/~/ɪ/ contrasts are therefore suited for training naive Bulgarian listeners.

Figure 1: The first two formants of the relevant Dutch vowels [2] and the relevant Bulgarian vowels [3]. Bold = female, plain = male; circled = Bulgarian, uncircled = Dutch.



3. TRAINING NEW CONTRASTS

In the present study, Bulgarian listeners are trained on the two Dutch contrasts, following a distribution-based approach. This leads to three research questions.

First, previous studies [3, 4] have found that (unsupervised) training with bimodal distributions of auditory forms helps naive listeners discriminate non-native contrasts, whereas training with monomodal distributions does not. The first question is whether this study can replicate this finding for the Dutch contrasts /a/~/ɑ/ and /i/~/ɪ/ applied to naive Bulgarian listeners.

Second, it has been found [3, 4, 5] that learners can infer new categories from bimodal distributions without any supervision, i.e. without explicit instruction or feedback about the phonetic categories that are to be learned. However, it has also been found [6] that explicit instruction has a positive effect on the perception of non-native contrasts. The present study is the first to investigate whether the positive effect of explicit instruction generalizes to distributional learning.

Finally, previous studies have shown mixed results with respect to the transfer of distributional learning to new tokens, i.e. tokens not presented during training. For instance, it has been shown that a long and implicit training of Mandarin tones enabled learners to transfer their gained discrimination abilities to new tokens [8]. However, it was also found that the ability gained in the discrimination of one non-native contrast through listening to a bimodal distribution did not transfer to the discrimination of another similar non-native contrast [3]. In the present study, the participants are tested with different productions of the trained contrasts in order to test the transferability of distributional learning.

4. METHOD

4.1. Participants

Forty Bulgarians took part in the study. Their ages varied between 16 and 60. They had no previous knowledge of Dutch and were tested in Bulgaria.

In the experiment, each participant took a *pre-test* for the Dutch /a/~/ɑ/ contrast, then received five minutes of training on the /a/~/ɑ/ contrast, and finally took a *post-test* for the /a/~/ɑ/ contrast. After this, the procedure was repeated for the /i/~/ɪ/ contrast. We now describe the test and training parts in detail.

4.2. Pre-test and post-test

The Bulgarians' perception of each Dutch contrast was assessed twice, namely before and after training. The two tests, which we call the *pre-test* and the *post-test*, respectively, were identical identification tests. Any effect of training should show up as a difference between the participants' performance on the pre-test and their performance on the post-test.

Stimuli: The pre- and post-tests presented naturally produced Dutch vowel tokens that had been selected from the database reported in [2]. We started with the productions of seven female and seven male native speakers of Dutch, who pronounced the vowels in a /sVs/ context, i.e. as /sas/, /sɑs/, /sis/ and /sɪs/. For each of the four vowels we then selected the 5 best male and 5 best female tokens, giving a total of 40 test stimuli. The Dutch vowels /i/ and /ɪ/ contrast only in their formant values, whereas /a/ and /ɑ/ also differ in duration (/a/ is long). In order to make the differences between the members of the two contrasts more similar we removed the duration difference between /a/ and /ɑ/ by adjusting their respective durations with the Praat program.

Procedure: During the pre- and post-tests the listeners were given a forced-choice identification task. To assess a participant's performance on the /a/~/ɑ/ contrast, we presented her all 20 relevant test stimuli once. For each Dutch /a/ or /ɑ/ token we asked her to classify it either as Bulgarian /a/ or as "other", which stood for another unknown or different vowel. If the participant had trouble discriminating Dutch /a/ tokens from Dutch /ɑ/ tokens, we expected her to choose Bulgarian /a/ all of the time; if she discriminated the contrast well, on the other hand, we expected her to choose "other" for the /ɑ/ tokens. An analogous test with 20 test stimuli was performed for the /i/-/ɪ/ contrast.

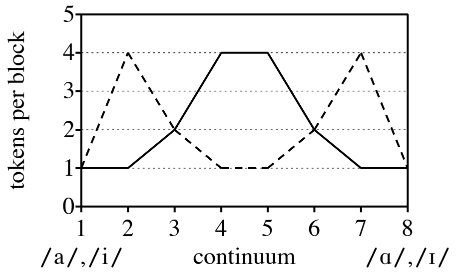
4.3. Training

Between the pre-test and the post-test for each contrast, the participants were trained on the perception of that contrast.

Stimuli: For training, we used synthetic stimuli that formed two continua, one between Dutch /a/ and /ɑ/ and another between Dutch /i/ and /ɪ/. The mean formant values obtained by Adank et al. [5] were used as end points of the continua, after being slightly adjusted according to the perception of a Dutch native speaker; for each contrast, six intermediate positions on the continuum were defined by linear interpolation along a mel scale. We thus created 16 stimuli in total.

Explicit instruction: Each of the 16 training stimuli could optionally be coupled to a simultaneous written instruction about the category it was supposed to belong to. Thus, for stimuli 1 to 4 on the /i/~/ɪ/ continuum the computer screen could optionally tell the participant that she was hearing a foreign vowel that sounds like Bulgarian /i/, whereas for stimuli 5 to 8 she could be told she was hearing a foreign vowel between Bulgarian /i/ and /e/. For the /a/~/ɑ/ continuum, stimuli 1 to 4 were likewise coupled to Bulgarian /a/, and stimuli 5 to 8 to "a foreign vowel between Bulgarian /a/ and /o/".

Figure 2: Training continua and distributions.



Stimulus blocks: For each contrast we used the eight training stimuli to create two blocks of 16 tokens, in the same way as [4, 5] did. This is illustrated in Figure 2. For the /a/~/ɑ/ continuum, the *bimodal block* consists of 1 token of stimulus 1 (the most /a/-like stimulus), 4 tokens of stimulus 2 (the /a/-like training peak), 2 tokens of stimulus 3, 1 token of stimulus 4, 1 of stimulus 5, 2 of stimulus 6, 4 of stimulus 7 (the /ɑ/-like training peak), and 1 of stimulus 8 (the most /ɑ/-like stimulus), all in randomized order. The *monomodal block* instead has 4 tokens of stimuli 4 and 5 each and 1 token of stimuli 2 and 7 each. Similar blocks were assembled for the /i/~/ɪ/ continuum.

Training groups: The 40 Bulgarian listeners were randomly divided into four groups, each of which received a different training condition: 10 listeners heard bimodal blocks without explicit instruction, 10 heard bimodal blocks with explicit instruction, 10 heard monomodal blocks without explicit instruction, and 10 heard monomodal blocks with explicit instruction.

Procedure: The entire training session for a participant on each contrast consisted of eight (either bimodal or monomodal) internally randomized training blocks for that contrast, i.e. there were in total 128 stimulus tokens per contrast.

5. RESULTS AND DISCUSSION

We assess the results of the listeners in terms of the number of 'correct' identifications in the pre- and post-tests. The definition of 'correctness' crucially involves the category "other". For instance, a stimulus that has been intended by the speaker as Dutch /a/ or /i/ is considered to have obtained a 'correct' response whenever a listener chose the Bulgarian /a/ or /i/ category, and a stimulus that has been intended by the speaker as Dutch /ɑ/ or /ɪ/ is considered to have obtained a 'correct' response whenever a Bulgarian listener chose the "other" category.

While Dutch listeners would typically classify all 20 /a/~/ɑ/ stimuli and all 20 /i/~/ɪ/ stimuli correctly, the Bulgarian listeners turn out to have trouble with both contrasts. Figure 3 shows the results for all four groups of listeners on both the /a/~/ɑ/ and the /i/~/ɪ/ stimuli, before as well as after training. Every bar in the figure represents an average over 10 listeners. In none of the 16 conditions do the 10 listeners reach an average correct score of 17.

The next thing that meets the eye in Figure 3 is that the Bulgarians score much better on the /a/~/ɑ/ contrast than on the /i/~/ɪ/ contrast. The difference is large: in the pre-test, when all 40 Bulgarian participants are still unspoilt naive listeners of Dutch and we can therefore pool the results of the four groups, an assessment of the 40 differences between the listeners' /a/~/ɑ/ and /i/~/ɪ/ scores reveals that the Bulgarian population classifies about 4.38 more stimuli correctly for the /a/~/ɑ/ contrast than for the /i/~/ɪ/ contrast ($t(40)$ = 6.709; 95% confidence interval = 3.06~5.69). However, this conspicuous fact is irrelevant to the three research questions for this paper (§2), which we discuss next.

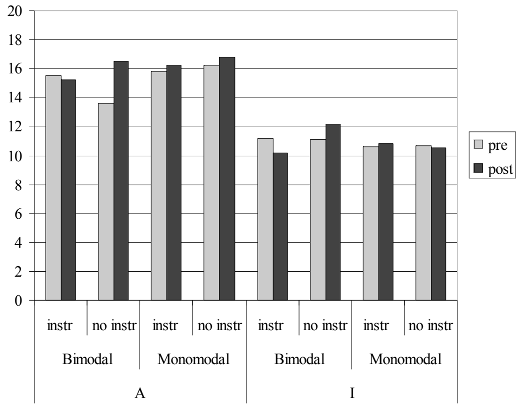
5.1. Distributional learning works

The first research question is whether bimodal distributional training improves the listeners' perception in the present case. To measure the effect of learning, we define a participant's *improvement* on a certain contrast as the difference between her score in the pre-test and her score in the post-test on that contrast. We then note that the improvements in Figure 3 are not reliably different for the /a/~/ɑ/ contrast than for the /i/~/ɪ/ contrast. As a consequence, we can compute the *total improvement* for each participant by adding her improvement on the /a/~/ɑ/ contrast to her improvement on the /i/~/ɪ/ contrast. We thus end up with 10 total improvement values for each of the four training groups.

The question whether bimodal distributional learning works better than monomodal distributional learning then reduces to comparing the 10 total improvement values of the bimodal-without-instruction group with those of the monomodal-without-instruction group. The average score of the bimodal group is 3.60 higher than that of the monomodal group, with a 95% confidence interval of 0.40~6.80. This difference is reliably different from zero ($t(10)=2.37$, two-tailed $p=0.029$).[1]

We conclude that unsupervised bimodal distributional training helps the Bulgarian population to improve their perception of the Dutch /a/~/ɑ/ and /i/~/ɪ/ contrasts. We thus successfully replicate Maye's [4, 5] results for a new case.

Figure 3: Mean number of correct responses in the pre- and post-tests for the four groups.



5.2. Instruction hampers bimodal learning

To test whether supervision in the form of explicit instruction benefits distributional learning, we compare the 10 total improvements of the bimodal-without-instruction group with those of the bimodal-with-instruction group. The average score of the group without instruction is 5.30 better than that of the group with instruction, with a 95% confidence interval of 1.39~9.21. This difference is reliably different from zero ($t(10)=2.85$, two-tailed $p=0.005$).[2]

From these results, we conclude that bimodal training helps Bulgarian perception of Dutch vowel contrasts more if it is not accompanied by explicit instruction.

5.3. Learning is transferred to new stimuli

The third research question, namely whether learners are capable of generalizing the results of distributional learning with one kind of stimuli to a new kind of stimuli, can be answered in the affirmative by noting that distributional learning on synthetic training stimuli has been successfully generalized to the natural stimuli of the post-test. This is in line with the results by Wang [8] for Mandarin tones.

6. SUMMARY AND CONCLUSION

The present study compared the perception of two Dutch non-native contrasts by naive Bulgarian listeners before and after perceptual training. Bimodal distributional learning turned out to help the listeners to perceive the Dutch contrasts, but only if no explicit instruction was given.

The consequence from this study for language education is that if distributional training is given it should not contain information about the number of phonemes being taught. This does not imply that explicit instruction cannot work, only that it should not be combined with distributional training. Perhaps one should not mix conscious and subconscious training methods: while bimodal training suggests that the listeners discover the bipolar nature of the stimuli presented at a subconscious level, explicit instruction implies a conscious understanding of the existence of two categories.

7. REFERENCES

[1] Adank, P., van Hout, R., Smits, R. 2004. An acoustic description of the vowels of Northern and Southern Dutch. *J. Acoust. Soc. Am.* 116, 1729–1738.

[2] Escudero, P., Benders, T. in prep. An acoustic description of Northern Dutch vowels produced by university students.

[3] Hayes, R. 2003. *How are second language contrasts learned?* PhD thesis, University of Arizona, Tucson.

[4] Maye, J. 2000. *Learning speech sounds categories on the basis of distributional information*. PhD thesis, University of Arizona, Tucson.

[5] Maye, J., Gerken, L.-A. 2001. Learning phonemes: how far can the input take us? In A.H.-J. Do et al. (eds), *Proc. BUCLD 25*, 480–490. Somerville, MA: Cascadilla Press.

[6] McClaskey, C.L., Pisoni, D.B., Carrell, T.D. 1983. Transfer of training of a new linguistic contrast in voicing. *Perception and Psychophysics* 34, 323–330.

[7] Tilkov, D. 1970. *Le vocalisme bulgare: les mouvements articulatoires et leur effet acoustique dans la formation des voyelles bulgares*. Paris: Klincksieck.

[8] Wang, Y. 1994. Training American listeners to perceive Mandarin tones. *J. Acoust. Soc. Am.* 106, 3649–3658.

---

[1] One can also defend a one-tailed $p=0.015$ on the basis of the expected direction, given that this part of the experiment is a replication.

[2] Readers who challenge the use of two t-tests for partly dependent data may feel reassured to learn that even an exploratory analysis of variance detects that the four groups are not the same ($F(1,36) = 3.59$, which is greater than 1.0 with $p = 0.023$), and that a Tukey post-hoc test subsequently reveals that the group that received bimodal training without explicit instruction has improved more than the group that received bimodal training with explicit instruction (estimated difference in scores = 5.30, which is different from zero with $p = 0.015$; 95% confidence interval = 0.83~9.77).