The perceptual development of a British-English phoneme contrast in Dutch adults

Willemijn Heeren
Utrecht University

1. Introduction

The world's languages differ in their phoneme inventories. While learning a first or second language, a listener must acquire such a set of phonemes. In the continuous flow of speech sounds, however, two instances of one phoneme may show great diversity, due, for example, to speaker characteristics or speech rate. Moreover, two acoustically similar speech sounds may actually be realizations of two different phonemes. Native listeners tend to put sharp boundaries between phoneme categories in their language. Generally, discrimination of utterances taken from different sides of the phoneme boundary is higher than discrimination of within-category tokens. Assuming that this discrimination pattern is a native listener's end state, this would also be what a language learner is trying to achieve. But how does one learn to perceive novel phonemes?

Early studies that tried to change phoneme perception through laboratory training were not very successful (see Strange and Jenkins 1978 for a review). Later studies, however, have shown that nonnative phoneme contrasts can be learnt through relatively short laboratory training (e.g. Jamieson & Morosan 1986; Lively, Logan & Pisoni 1993). But there are also nonnative contrasts for which training is unnecessary. Best et al. (2003) showed that nonnative phoneme contrasts that do not fall within the acoustic range exploited by one's native language can be discriminated quite well.

To study the perceptual development of novel phoneme contrasts, we start from two opposing hypotheses that have been posed to test learning of native phonemes (Liberman, Harris, Kinney & Lane 1961): Acquired Distinctiveness and Acquired Similarity. Both hypotheses deal with the degree to which within-

Linguistics in the Netherlands 2004 21 (2004), **91–101.**
ISSN 0929–7332 / E-ISSN 1569–9919

category and between-category differences can be discriminated by a listener. The first hypothesis, Acquired Distinctiveness, says that listeners learn to perceive differences between those speech sounds that they are trained to categorize differently, although they were unable to hear any differences before learning this new phoneme contrast. After training, discrimination of stimuli on different sides of the phoneme boundary has improved. In support of this hypothesis, a training study by Jamieson and Morosan (1986) reported increased discrimination at the phoneme boundary without within-category improvement.

The second hypothesis, Acquired Similarity, states that both within-category and between-category speech sounds can be distinguished well before training. As a result of training, however, perceptual sensitivity to speech sounds that are categorized together decreases, such that only above-chance discrimination of the stimuli straddling the phoneme boundary remains. This way of learning seems similar to the way infants treat speech sounds during their first year of life (Pisoni 1991). Werker and Tees (1984a), for example, have shown that infants of 6 to 8 months of age can discriminate a natural speech contrast that their language environment does not contain, and that is not discriminated by adults from that language. But, after about 10 to 12 months these infants lose the ability to discriminate these phoneme pairs. It is not probable, however, that the infants' representations lie at a phonemic level.

The question this paper aims to answer is: how do Dutch listeners learn the British English /θ–s/ contrast, as /θ/ is not a phoneme of Dutch. Do Dutch adults learn the /θ–s/ contrast in accordance with Acquired Distinctiveness or Acquired Similarity? We tried to answer this question by means of a laboratory training study, run with Dutch adult listeners. Both before and after training, the perception of the nonnative phoneme continuum was assessed in absolute identification and discrimination tests. A control group that did not participate in training sessions, was also tested.

A subquestion that was addressed in the present study is whether the availability of knowledge of the language one is learning influences perceptual learning. Since most Dutch have learnt some English in primary and secondary school, the /θ–s/ contrast may not be entirely new to the participants. However, knowing that /θ/ is a different sound than /s/ does not imply that a Dutch listener can differentiate the two acoustically. Furthermore, the Dutch often produce /s/ when trying to pronounce /θ/ (Collins & Mees 1999), which shows that they have difficulties with the English phoneme. We want to find out whether listeners who are told which language they are learning, benefit from this knowledge, as opposed to learners who are left in the dark as to what language they are learning.

2. Method

2.1 Materials

Eight-stimulus continua were synthesized for the British-English phoneme contrast /θ–s/ by means of linear spectral interpolation (van Hessen 1992). The phonemes occurred in the onset position of a pair of (both in Dutch and in English) nonsense words: *thif – sif*. Nonsense words were used to exclude word frequency effects or lexical bias (cf. Ganong 1980). The phoneme continua were based on speech from six speakers of Standard English, both males and females. The location of the phoneme boundary in each continuum had been determined from a classification study with 31 native British-English listeners.[1]

2.2 Participants

Thirty-four students participated in the test. None of them studied English. All were native speakers of Dutch and reported normal hearing. Half of the subjects received training between pretest and posttest. The other half, the control group, participated only in pretest and posttest, with a time interval of approximately a week.

2.3 Design

The experiment was a pretest-posttest design. In both pretest and posttest four interval AX (4IAX) discrimination and absolute identification with speech from one male speaker were administered. At the end of the posttest a short questionnaire was given. In contrast with earlier training studies, we did not present written labels to our participants, but introduced the categories by means of pictures. This forced the listeners to figure out the acoustic differences by themselves.

On each trial of 4IAX discrimination, two different stimuli A and B were presented in one of eight possible orders: AB–AA, AA–BA, BA–AA, AA–AB, BB–AB, BB–BA, AB–BB or BA–BB. The listener had to indicate whether the first or the second pair consisted of the same stimuli. From the eight stimuli in the continuum, pairs A,B were chosen at either a one-step or a three-step distance. For example, a possible three-step trial is: 14–11. Learning by Acquired Distinctiveness would become most apparent in the one-step test, since the discrimination level in this test was low from the beginning. On the other hand, learning by Acquired Similarity would show in the three-step test, since its initially high discrimination levels could fall as a result of training. The 4IAX discrimination test was expected to reflect both phonemic and auditory perception of the stimulus pairs (Pisoni 1975; Gerrits 2001).

In absolute identification the listener has as many response options as there are stimuli. Since the listener is asked to indicate exactly which stimulus he heard, this test gives us insight into the listeners' control over the continuum.

For training, a classification design with trial-by-trial feedback on the correctness of the participants' responses was used. Classification was preferred over discrimination, since it directs the participants' attention towards the existence of two categories (e.g. Jamieson & Morosan 1986). The training materials consisted of phoneme continua from five speakers, both males and females, other than the test speaker. Speaker variation was included to encourage robust category formation (Lively et al. 1993).

2.4 Procedure

Listeners were tested individually. Half of them were told that they would hear English nonsense words, the other half of the participants were told that they would hear words from a foreign language. All experiments were run in a relatively quiet room at the Utrecht Institute of Linguistics OTS. A laptop computer was used both to present the stimuli at random and to register responses. Stimuli were presented over Beyerdynamic DT 770 headphones at a comfortable listening level.

The pretest was completed on the first day. On subsequent days, training sessions were run until the listener classified the new phoneme contrast correctly in at least 85% of the trials. On each training session, one to four training tests of 480 trials each were run. On the final day the posttest took place. This test was of a similar content as the pretest, apart from a short questionnaire that was given after the listening tests. In this, the listeners were asked about the spelling of the newly learnt words and their experience with foreign languages. The following subsections will discuss the procedures of the listening tests in more detail.

2.4.1 *4IAX discrimination*

The first test administered during pretest and posttest was 4IAX discrimination. This test was given twice, once with one-step stimulus pairs (i.e. 1–2, 2–3, …, 7–8) and once with three-step stimulus pairs (i.e. 1–4, 2–5, …, 5–8). The order of these tests was balanced across subjects.

Listeners received written instructions in which they were asked to indicate whether the first or the second word pair they heard consisted of the same stimuli. It was stressed that differences could be small. Responses were given by striking one of two keys on the computer's keyboard. A short task introduction was given, consisting of eight four-step stimulus pairs from the same continuum. Inter-stimulus intervals were set at 300 ms, inter-pair intervals at 500 ms, and response times were unlimited. The eight different orders per stimulus pair

were each presented four times, resulting in 224 trials for the one-step test (= 4 repetitions × 7 stimulus pairs × 8 possible orders) and 160 trials in the three-step test (= 4 repetitions × 5 stimulus pairs × 8 possible orders). Trial order was randomized and there were three short breaks at regular intervals.

2.4.2 *Absolute identification*

The second test in pre- and posttest was absolute identification. Listeners received written instructions, telling them to indicate exactly which stimulus they had heard from the continuum. The instructions included a picture of a row of eight buttons numbered 1 to 8 from left to right. Over the first and eighth buttons, pictures of a man wearing differently colored headphones were shown. It was explained that each button hid a unique word. The words changed in steps from the name of the first man (behind button 1) to the name of the second man (behind button 8). Next, the stimuli were introduced five times in increasing, decreasing and random sequences. Listeners were instructed to listen very carefully and to try to remember which button corresponded to which word. During testing, stimuli were presented 20 times in random order, which resulted in a total of 160 trials. Listeners responded by mouse-clicking one of eight on-screen buttons. Response times were unlimited and there were two short breaks at regular intervals.

2.4.3 *Classification training with feedback*

During training, a classification design was used. The categories were represented by the pictures of the men and were introduced as the men's names. The listeners had to reach a mean score of at least 85% correct identifications over two subsequent training tests before proceeding to the posttest. The test was introduced by the pronunciation of both endpoint stimuli by each of the five training speakers.

Listeners received immediate feedback on each trial, informing them of the correctness of their choice. After every quarter of the total number of trials a break was given. During each break, and also at the end of a training test, the percentage of correct responses so far was shown. The training tests each contained 12 repetitions per stimulus, resulting in 480 trials (= 5 speakers × 8 stimuli × 12 repetitions).

3. Results

Training results were represented as percentages of *sif*-responses per stimulus for each of the five speakers. Next, the phoneme boundary was determined at the 50%-point for each of the five speakers in the training set and for each listener separately. Also, boundary widths were determined by calculating the

25–75% range. Boundary widths reflect the steepness of the boundary. For 4IAX discrimination, percentages of correct responses were determined per stimulus. From absolute identification results, the mean responses to each of the stimuli and their variances were determined.

3.1 Training

The participants' results from the first and the last training session were compared to the English norm defined by 31 native listeners. The mean boundaries in the first and last training sessions differed significantly from those defined by the English listeners ($F[1,221] = 7.9$, $p = .005$ and $F[1,222] = 7.7$, $p = .006$, respectively). Post-hoc analyses showed, however, that the mean boundary of only one speaker differed from the English norm in both training sessions. The Dutch listeners perceived more stimuli as /s/. In addition, the boundary value of one more speaker in the last training differed from that of the English. Listeners again judged more stimuli as /s/.

The results from the first training session did not lead to boundary width values in 36% of the cases: usually, no 25%-points were found, which means that listeners did not consistently choose the /θ/ category for stimuli at that end of the continuum. In the last training session, this number had decreased to only 6.3%, approximately equal to the number of cases the natives missed (6.5%). The boundary widths of the listeners for whom the ranges were defined differed significantly from the native English ones in the first training session ($F[1,186] = 52.4$, $p < .001$). In the last training session, however, these differences were no longer present.

3.2 4IAX discrimination

For one-step 4IAX discrimination, a repeated measures ANOVA was run with within-subjects factors Test (2) and Stimulus Pair (7), and between-subjects factors Listener Group (2) and Language (2). The results are shown in Figure 1.

First of all, a main effect of Test was found ($F[1,30] = 14.9$, $p = .001$). Listeners gave more correct answers in the posttest than in the pretest. Secondly, a main effect of Stimulus Pair was found ($F[6,180] = 10.2$, $p < .001$). This means that listeners were not equally good at distinguishing all pairs of stimuli.

No main effect of Listener Group or Language was found. The absence of these effects means that the test listeners did not perform considerably better than the controls, and that listeners in the English condition did not benefit from their knowledge of what language they were attending. A post-hoc ANOVA on just the posttest data did reveal an effect of Listener Group ($F[1,224] = 7.5$, $p = .007$): the trained listeners scored better than those in the control group.

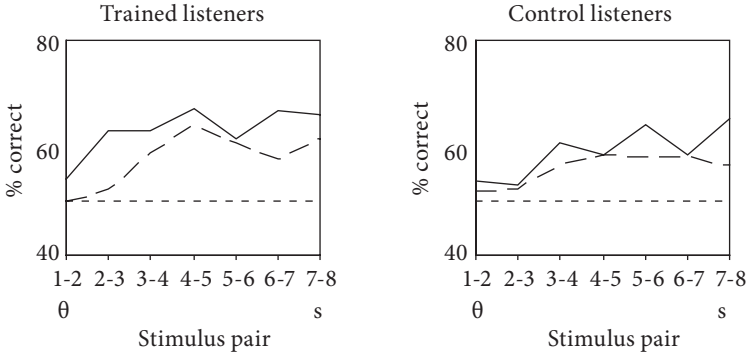

Figure 1. Results of one-step 4IAX discrimination of pretest (dashed) and posttest (solid) with a reference line at chance level.

Three-step 4IAX discrimination showed similar results. A repeated measures ANOVA resulted in main effects of Test ($F[1,30] = 13.4$, $p = .001$) and Stimulus Pair ($F[3.4,101.4] = 9.3$, $p < .001$). Again, no main effect of Listener Group of Language was found. A post-hoc ANOVA on the posttest data showed that trained listeners performed better than the controls ($F[1,155] = 10.8$, $p = .001$).

3.3 Absolute identification

A repeated measures ANOVA was run on the mean responses with within-subjects factors Test (2) and Stimulus (8) and between-subjects factors Listener Group (2) and Language (2).

Firstly, a Test × Stimulus interaction ($F[4.5,136] = 3.9$, $p = .003$) was found. At the /θ/-end of the continuum, participants became better at identifying the stimuli, while this was not the case at the /s/-end. This can be seen in Figure 2 by the trend towards the ideal case of absolute identification at the /θ/-end only. Furthermore, main effects of Test ($F[1,30] = 28.2$, $p < .001$) and Stimulus ($F[2.6,78.3] = 794.1$, $p < .001$) were found. But again, no effects of the between-subjects factors Listener Group and Language were present.

The repeated measures ANOVA on the mean variances showed main effects of Test ($F[1,30] = 20.6$, $p < .001$) and of Stimulus ($F[4.3,129.7] = 5.5$, $p < .001$). Posttest variances were smaller than those in the pretest, meaning that participants had become better at identifying the stimuli.

In sum, differences between pretest and posttest were clearly present in the three tests. Trained listeners' progress, however, was hard to discriminate from that of control listeners. The fact that some test listeners needed only a few training sessions to reach criterion may partly account for this. In the following subsection this explanation will be investigated.

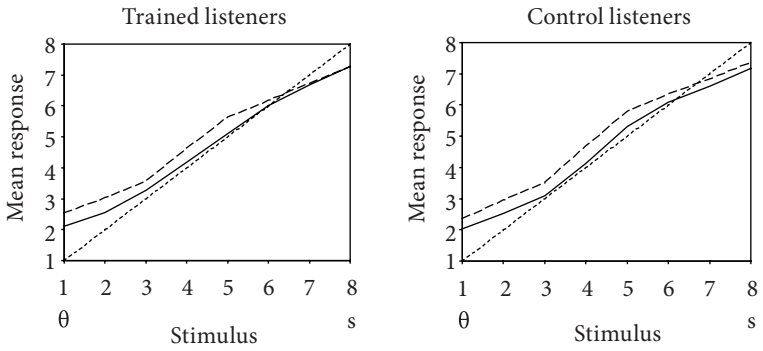

Figure 2. Mean responses to each of the stimuli in pretest (dashed) and posttest (solid). The line of short dashes shows the ideal case of error-free perception.

3.4 The effect of training amount on pretest-posttest differences

The amount of training needed to reach criterion varied among test listeners: some needed only 960 training trials before reaching criterion, while others needed as many as 7680 trials. It is conceivable that those listeners who needed more training also show larger differences between pre- and posttest results. Therefore, the trained listeners' data were tested in repeated measures ANOVA's, this time with Training Amount as a covariate. Only the findings that differ from the effects described in Sections 3.2 and 3.3 will be reported here.

For three-step 4IAX discrimination, a Text × Training Amount interaction was found ($F[1,15] = 4.8$, $p = .044$) as well as a main effect of Training Amount ($F[1,15] = 16.4$, $p = .001$). The rise in correct responses between pretest and posttest tended to be larger as the amount of training increased. The mean variances of the stimuli in absolute identification also showed a main effect of Training Amount ($F[1,15] = 5.5$, $p = .033$).

3.5 Posttest questionnaire

After the perception tests, listeners wrote down the words they had heard as the two extremes during absolute identification. The reported sets varied greatly. From the test group, six listeners reported the correct pair of words, but none of the control listeners succeeded in doing this. Remember that acoustically only the first consonant varied between stimuli *thif* and *sif*. Three types of errors were identified. Listeners (a) replaced target phonemes with other phonemes, (b) added phonemes that were not present in the signal or (c) reported hearing differences between the vowels or between the coda consonants.

In total, participants in the control group made more errors: 55 as opposed to 37 for the trained listeners. The controls made the majority of these, 49%, in their responses to the first consonant, i.e. the target phoneme. They used almost

twice as many substitutions, and also additions that resulted in complex onsets such as *stif* instead of *thif*. They also more often replaced the coda consonant with another one as in *striss* instead of *thif*.

4. Discussion

Listeners participated in training sessions until they learnt the contrast to the predetermined level of 85% correct. The training results showed that the phoneme boundaries were already close to the English norm during the first training session. But the widths of the phoneme boundaries became much smaller and even indistinguishable from the English norm as a result of training.

The pre- and posttest revealed main effects of Test for both discrimination and absolute identification tests. However, the absence of Test × Listener Group interactions showed that listeners — whether in the test or control group — all performed better when they did the tests for the second time. Differences between the test and control groups were only present within the posttest data of three-step discrimination and of the variance results in absolute identification. As for three-step discrimination, test listeners performed better than controls, especially at the /θ/-end of the continuum in the posttest. Response variances in absolute identification were smaller for the test group.

By taking into account the amount of training participants received, we found that listeners who completed many training sessions showed a larger difference between their pretest and posttest results for three-step discrimination and for the variances in absolute identification. For these results, our assumption that the more training a listener needs, the more improvement will be found, was borne out.

The main research question of this paper was: do Dutch adults learn the British-English /θ–s/ contrast in a way compatible with Acquired Distinctiveness or with Acquired Similarity? We found no support for learning by Acquired Similarity. In that case, perceptual sensitivity to speech sounds that belong to the same category should decrease, but we found no such effects at all. On the contrary, the improvement we found for the discrimination tests mainly occurred within instead of between categories. So these findings do not strongly support our other hypothesis, Acquired Distinctiveness, either.

Despite the progress during training, little evidence of an increase in discrimination levels at the phoneme boundary was found, contrary to earlier findings (Jamieson & Morosan 1986). This may have been caused either by the nature of the tasks used in pre- and posttest, or by the participants' high pretest levels. Firstly, the tasks used may have directed the listeners' attention too much towards the acoustic differences between the stimuli by testing auditory instead of phonemic perception. However, we expected to find a combination of these

listening levels in our results (Gerrits 2001; Pisoni 1975). Secondly, most participants performed already quite well in the pretest (see, for example, Figure 2). The space left for improvement as a result of training was thereby restricted and possibly difficult to distinguish from improvement by task repetition. We also think that the pretest was a training in itself due to its length, which helped control listeners to improve their scores in the posttest. Most earlier studies however, did not test a control group (e.g. Strange & Dittmann 1984; Logan, Lively & Pisoni 1991) and could therefore only report the test group's progress.

An explanation that may account for a portion of the errors made by the participants in reporting which words they had heard, is 'verbal transformation' (Warren 1961a). When listeners are repeatedly confronted with the same syllable, word or sequence of words, they start hearing differences that are not present in the signal. The types of changes found in this study are consistent with the types of misperceptions reported by Warren (1961a). But, control listeners made more errors than test listeners. We assume that the speaker variation that was available to the trained participants helped them to form the correct representations more often in comparison with the control group.

A sub-question that was addressed in the present study was whether the availability of knowledge of the language you are learning influences perceptual learning. We found that participants who knew they were listening to English did not benefit from this knowledge. So either listeners in both the English and the Foreign Language conditions used their knowledge of English equally, or neither of the groups accessed this knowledge.

5. Conclusion

Dutch listeners improved their perception of British-English /θ–s/ during training. Trained listeners performed better in the posttest than in the pretest and in several respects they also did better than the control group. Their improved performance excluded Acquired Similarity, but did not strongly support Acquired Distinctiveness either. This effect was thought to be due to both the high pretest performances of our participants and the nature of the tests used in pretest and posttest. Furthermore, control listeners, who received no training, also improved by simply performing the tests in pretest and posttest twice. These results show that it is important to include a control group into the design of a phoneme training study, which has often not been the case. Finally, listeners who knew that they were listening to British-English did not benefit from this knowledge opposed to listeners who were told they were listening to a foreign language.

Note

1. This work was supported by the Netherlands Organisation for Scientific Research (NWO).

References

Best, C. T., Traill, A., Carter, A., Harrison, K. D. & Faber, A. (2003) !Xóõ click perception by English, Isizulu, and Sesotho listeners. *Proceedings of the 15th ICPhS, Barcelona*, 853–856.

Collins, B. S. & Mees, I. M. (1999) *The phonetics of English and Dutch*. Brill, Leiden.

Ganong, W. F. (1980) Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6(1), 110–125.

Gerrits, E. (2001) *The categorisation of speech sounds by adults and children*. Doctoral Dissertation, Utrecht University.

Hessen, A. J. van (1992) *Discrimination of familiar and unfamiliar speech sounds*. Doctoral Dissertation, Utrecht University.

Jamieson, D. G. & Morosan, D. E. (1986) Training non-native speech contrasts in adults: acquisition of the English /ð/–/θ/ contrast by francophones. *Perception & Psychophysics* 40, 205–215.

Liberman, A. M., Harris, K. S., Kinney, J. A. & Lane, H. (1961) The discrimination of relative onset-time of the components of certain speech and nonspeech patterns, *Journal of Experimental Psychology* 61(5), 379–388.

Lively, S. E., Logan, J. S. & Pisoni, D. B. (1993) Training Japanese listeners to identify English /r/ and /l/. II The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94, 1242–1255.

Logan, J. S., Lively, S. E. & Pisoni, D. B. (1991) Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89, 874–886.

Pisoni, D. B. (1975) Auditory short-term memory and vowel perception. *Memory & Cognition* 3, 7–18.

Pisoni, D. B. (1991) Modes of processing speech and nonspeech signals. In I. G. Mattingly & M. Studdert-Kennedy (eds.): *Modularity and the motor theory of speech perception*. Lawrence Erlbaum Associates, Hillsdale NJ, 225–238.

Strange, W. & Dittmann, S. (1984) Effects of discrimination training on the perception of /r–l/ by Japanese adults learning English. *Perception & Psychophysics* 36, 131–145.

Strange, W. & Jenkins, J. J. (1978) Role of linguistic experience in the perception of speech. In R. D. Walk & H. L. Pick (eds.): *Perception and experience*. Plenum Press, New York, 125–169.

Warren, R. M. (1961a) Illusory changes of distinct speech upon repetition-the verbal transformation effect. *British Journal of Psychology* 52, 249–258.

Werker, J. F. & Tees, R. C. (1984a) Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7, 49–63.