The influence of an L1 allophonic opposition on the acquisition of an L2 phonological contrast

Thesis for BA Taalwetenschap by Jorina Bodelier

Student ID 5740401

Supervised by prof. dr. P.P.G. Boersma

22 June 2010

Contents

0. Abstract

In this paper, results are presented of a study into the effects of a) two allophones [ʀ, ɹ] in the L1 and b) one sound [ʀ] but no [ɹ] in the L1 on perceptual acquisition of a phoneme contrast between /ʀ/ and /ɹ/ in a fictional L2. A group of seven native speakers of Dutch (with [ʀ, ɹ] allophones) and a group of seven native speakers of Limburgish (with only [ʀ]) were chosen to participate in a word-learning task containing 16 words (8 controls and 8 pseudo-minimal /ʀ, ɹ/ pairs, four of which had /ʀ/ or /ɹ/ in onset position, such as [ʀistɨn] and [ɹiskoʂ] and four had /ʀ/ or /ɹ/ in coda position, such as [bɛʀzela] and [bɛɹzesɔ]). At the end of the word-learning task participants did a test (similar to the learning task, but without the feedback) and eye tracking was used to determine whether participants were able to distinguish pseudo-minimal pairs *before* the onset of the second disambiguation (i.e. the last syllable) during the test. No significant differences were found between groups and within groups for recognition of /ʀ, ɹ/ words over controls. A significant difference (significant at the $\alpha = 0.01$ level) was, however, found for both groups together, as well as for the Dutch group (but not the Limburgish group) for recognition of /ʀ/ in onset position over /ʀ/ in coda position.

1. Introduction

Learning a second language has its difficulties, not least the fact that it seems almost impossible for adults to master a native-like accent in a second language (L2). Infants, with no previous exposure to language, learn their native language accent perfectly. However, as they gain experience in their native language (L1) it is generally thought that their brain becomes 'attuned' to the sounds and the sound system of the L1 and that this influences the way speech sounds (not only of the L1 but in general) are perceived. Influence of the L1 on perception can be found in infants as young as six months of age (Kuhl et al., 1992). According to Flege (1995) this attunement to the sounds of the L1 can lead to a 'foreign accent' in the production of an L2. However, it should be noted that Flege does not claim that a foreign accent is completely due to perception. Even so, a foreign accent is thought for a large part to be on account of perception issues. The L1 attunement can also lead to the related problem of having difficulty (perceptually) understanding speakers of an L2.

What exactly does this attunement to the sounds of the L1 entail? Research shows that some speech sounds will seem perceptually closer to each other than others due to the influence of the categories and category prototypes of the L1, even if these sounds are, objectively speaking, at the same distance from each other. In more concrete terms, this means that a pair of sounds that is contrastive (i.e. different phonemes) in a person's L1 will be more obviously different to them than a pair of sounds that is *not* contrastive in that person's L1, i.e. allophones of the same phoneme (e.g. the phoneme /l/ is pronounced differently in English at the beginning or at the end of a word, and these different sounds are both allophones of /l/) or sounds that are not present in the L1, even if the members of both pairs are equally acoustically different to each other objectively. It also means that if a pair of sounds is near the centre of a phonemic category (which means they are a prototypical or near-prototypical exemplar of that category) of a person's native language, the members of the pair will be harder to distinguish from one another than a pair of sounds exhibiting less 'category goodness' (i.e. further away from the centre of the category) even if the distance between the members of the pair is the same, objectively speaking. This is called the perceptual magnet effect (e.g. Iverson & Kuhl 1995).

Thus, when speakers of different native languages learn a new contrast that occurs in neither of their respective native languages, or any L2 that they might have been exposed to, one would expect, given the above findings, that the relevant properties of the native language (such as whether it contains one or both of the sounds involved, and whether they constitute different phoneme categories or not) would influence the acquisition of the new contrast. And indeed, one can see this, for example, in the well-known phenomenon that occurs in second-language perception of English by native Japanese speakers; they have difficulty distinguishing English /r/ and /l/ phonemes, the sounds for which occur as allophones in Japanese (e.g. Takagi, 2002). The goal of this study is, then, to address the question whether it is easier to split a phonological category into two separate categories for an L2, or whether it is easier to produce two distinct categories for the L2 when one of the sounds is already familiar from the L1 and the other known from another L2. The sounds chosen for this study, [ʀ] (uvular trill, used for example for /r/ in Standard German) and [ɹ] (alveolar approximant, like the sound for /r/ in Standard Southern British English), are – at least objectively speaking – very dissimilar perceptually (they differ in manner of articulation and place of articulation; in place they differ quite drastically), therefore any confusion between the two would have to be caused by perceptual attunement.

For this experiment two different groups of speakers were selected. Firstly, native speakers of Dutch were chosen who use [ʀ] in onset position (at the beginning of a syllable) and [ɹ] in coda position (at the end of a syllable) for the /r/ phoneme. There is considerable dialectal variation in the pronunciation of /r/ in Dutch, but participants of this group were chosen on the condition that they had this particular dialect. Secondly, native speakers of Limburgish were chosen. This language has only [ʀ] for the /r/ phoneme and no [ɹ]. These participants will have had much exposure to Dutch and the [ʀ, ɹ] allophone dialect in particular, as it is a common dialect in Dutch media. Although all speakers of Limburgish also speak Dutch, *their* dialect of Dutch does not contain [ɹ]. For the /r/-phoneme in both their native language and Dutch, they use [ʀ] in all positions, albeit possibly somewhat or completely devoiced at the ends of words. They may do an [ɹ]-like sound in their pronunciation of English /r/, but probably use [ɻ] due to the influence of American English. All Limburgish participants, assumedly, consider both [ɹ] and [ɻ] foreign (L2) implementations of /r/.

Specifically, then, this study addresses the question whether it is easier to split the /r/ category, with [ʀ] and [ɹ] as allophones of /r/, into two separate categories for [ʀ] (namely /ʀ/) and [ɹ] (namely /ɹ/) in a fictional L2, or whether it is easier to produce these two distinct phoneme categories for this L2 when [ʀ] is already familiar from the L1 (as /r/) and [ɹ] is known as a foreign implementation of /r/.

Hall (2007) suggests that there is something called the 'pairwise perceptual magnet effect' which means that whilst two sounds belonging to a different category will appear more distinct than one would expect given the acoustic difference, two sounds belonging to the same category (i.e. allophones) will appear less distinct than one would expect. So this could, for example, lead to speakers of a language which has sound X and Y as allophones and Y and Z as different phonemes, finding it quite easy to distinguish Y from Z, but not so easy to distinguish X and Y and possibly, they would even confuse X and Y with each other. And indeed, this is found in research by Peperkamp et al. (2003), who looked into discrimination between [ʁ] and [χ] by native speakers of French was looked into; in French, these two sounds are allophones. Contrasted with discrimination between [m] and [n], which are separate phonemes in French, discrimination between [ʁ] and [χ] was poorer, even when used in nonsense words that featured a context that was not necessarily the context in which the allophone would normally occur in French, i.e. where the other sound would be used.

One would predict, then, that native speakers of a language which contains two sounds as allophones would perceive them as more similar, as a result of associating them both with the same category, than native speakers of a language which does not have these sounds as allophones (but instead, only has one of the sounds concerned at all) would do. Thus, one could argue that the native speakers of the language that contains these sounds in an allophonic relationship should have more trouble distinguishing them as two separate phonemes than the native speakers of the language which does not contain these sounds in an allophonic relationship, meaning that it is difficult to split a phonological category into two. However, the Limburgish participants in this study, whose native language contains only [ʀ] and not [ɹ], are highly likely to associate [ɹ] with the same phoneme as [ʀ] (the /r/ phoneme) due to L2 influence, which means that they *might* have a similar reaction as the Dutch group. They differ from the participants whose native language contains both sounds as allophones, however, in that they never produce these sounds as allophones of each other. What this will mean in terms of their perception of these two sounds as separate phonemes in a new L2 is not yet clear.

2. Method

2.1. Participants

There were two groups of participants; group 1 consisted of seven native speakers of Dutch. Participants in group 1 varied in age from 18 to 25. Group 2 consisted of seven native speakers of Limburgish and participants in group 2 varied in age from 17 to 44.

Finding participants for this study was not an easy task, even though they were paid. For both groups the fact that the experiment took an hour and a half may have made it unappealing to participate. However, what also plays a role is that for group 1 there was a selection procedure in which participants had to read out a short text from a newspaper article to check for their accent, usually over the phone, before they took part in the experiment. Due to the fact that some participants did not have the right accent, they could not be used in the experiment.

Participants in group 2 had to travel a considerable distance (of approximately two and a half hours) to take part in the experiment, and although travel costs were refunded and this could have been seen as a nice opportunity for a day out, not many people were interested. It was especially hard finding people in the age of 18-25, which is likely due to the fact that in the Netherlands students of this age have a free public transports card.

2.2. Nonsense words

The training trials, discussed below, set out to teach participants to link 16 nonnative nonsense words to 16 nonsense pictures (obtained via and previously used in Escudero et al., 2008) shown in appendix A. The testing phase then tested whether the newly taught /ʀ, ɹ/ phoneme difference that was contained within the words had been acquired. These nonsense words either had a CVCCVC or a CVCCVCV structure. They were paired, making eight pairs. Four of these pairs contained a contrast between /ʀ/ and /ɹ/ and the other four pairs contained a contrast between /t/ and /f/ which already is a phoneme contrast in the native language of every participant. The latter were used as the control condition.

Control condition: /t/-/f/

tɨspiŋ	fɨsnel
toknɔn	foksɨp
z̦ɛtsɔka	z̦ɛfsɔli
ʂutçede	ʂufçeno

These are already separate phonemes in the L1, differing in place and manner.

Test condition: /ʀ/-/ɹ/

ʀistɨn	ɹiskoʂ
ʀaltev	ɹalpɛm
bɛʀzela	bɛɹzesɔ
qɑʀdʉsi	qɑɹdʉka

These are allophones in the L1 of participant group 1. For group 2, [ʀ] is present in the L1 and [ɹ] present as a foreign complementation of the /r/ phoneme ([ʀ] being the native representation of this phoneme). These are now to be learnt as separate phonemes in the same L2 and also differ in place and manner.

The pairs are not minimal pairs as such, because they have multiple differing sounds. The initial part of each pair could, however, be seen as a 'minimally paired part'. The disambiguating part at the end of the word is meant to stop the participants' attention from consciously focusing on the contrast between /ʀ/ and /ɹ/ when learning the words. Eye tracking was used during the test phase to see whether participants reacted to the /ʀ, ɹ/ contrast before the disambiguating end of the word was heard.

As can be seen above, in half of the words the target phoneme occurs in onset position (at the beginning of a syllable) and in the other half of the words the target phoneme occurs in coda position (at the end of a syllable). This was done in order to see if the position would make a difference at all (as it may be possible that the sounds are more salient and therefore more easily recognisable in one of these positions, or other factors might play a role), and if so, whether there would be any differences between the performance of the groups on the onset and coda position trials. The recorded words were spoken by a trained phonetician in a soundproof studio. An individual token was recorded for each training trial, whereas one token was used per word in the test phase (therefore, the 'test token' for each word was used 6 times; this test token had not been used in the training phase).

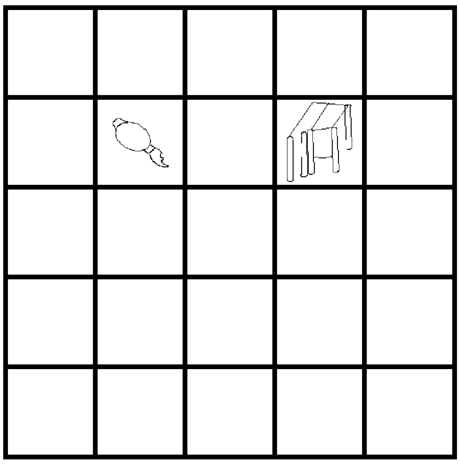


Fig. 1: An example of the screen during the first training phase

2.3 Procedure

Words were presented auditorily as multiple pictures appeared within a grid on a computer screen (Tobii T120). Participants were instructed to click, with a computer mouse, on the picture belonging to the word they heard. Before the auditory stimuli started, there was 1,5 s of silence, so that participants could look at the pictures and their position on the screen before hearing the target word. Since no existing language was used in the experiment, words were not presented with a carrier sentence but on their own. Feedback given to participants in the training phase was non-linguistic: after clicking, a green tick was displayed in case they clicked correctly and a red 'X' was displayed in case they clicked incorrectly. After this, the target picture was displayed on its own and the target word played simultaneously, regardless of whether participants had clicked right or wrong, so as to reinforce the link between the picture and its word. Instruction by the test leader, at the start of the experiment as well as between different experiment phases, took place in the respective participant's native language and participants took part in the experiment individually. The complete experiment session took an hour and twenty minutes on average.

2.3.1. Training

It is not clear after how much exposure lexicalisation of new words takes place, especially when the words have a non-native structure. In Escudero et al. (2008) however, participants learnt to link 20 nonsense words with a non-native structure successfully to pictures after 600 trials. The training phases of this study are thus modelled on that of Escudero et al. (2008).

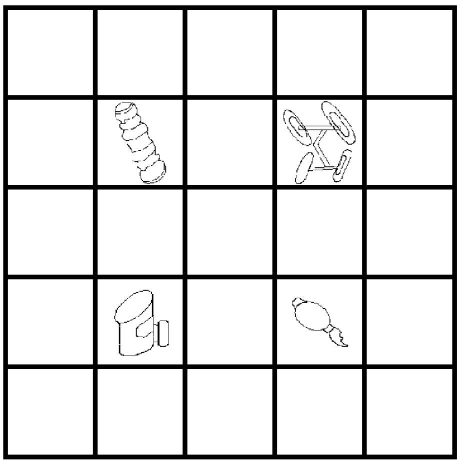


Fig. 2: An example of screen during the second training phase.

During each of the trials in the first training phase, participants saw the picture belonging to the target word, as well as a picture belonging to another word. All possible combinations of pictures featured, with the exception of a target word picture being combined with itself or being combined with the other half of the minimal pair the target word belonged to. So, for example, if the target word was [s̮ufçeno], it was never combined with the picture for [s̮utçede] (see fig. 1). Since this makes 16x14 combinations, the total amount of trials for training phase 1 was set to 224.

Between the first and the second training phase there was a brief break. Before the start of the second training phase instructions about this phase were given; participants were told that they would now have four pictures to choose from instead of two, and that this phase would take slightly longer than the previous one. This phase held 288 trials. During each of these trials, participants saw four pictures, and apart from the target word, the other pictures were randomly chosen out of the other fifteen (see fig. 2).

2.3.2. Test

During the test an eye tracker (Tobii T120, 120 Hz) was used to track the participants' eye movements. There was no break between the second training phase and the testing phase, except for brief instructions and eye tracker calibration. The test consisted of 128 trials in all, meaning that each of the 16 words was the target word in 8 test trials. During each trial, participants saw four pictures. These were the target picture, the picture of the word it was paired with as well as a pair from the other contrast (e.g. [bɛʀzela] as the target, combined with [bɛɹzesɔ], [s̮utçede] and [s̮ufçeno], see fig. 3). There was no feedback during this phase, which participants were made aware of before the test started.

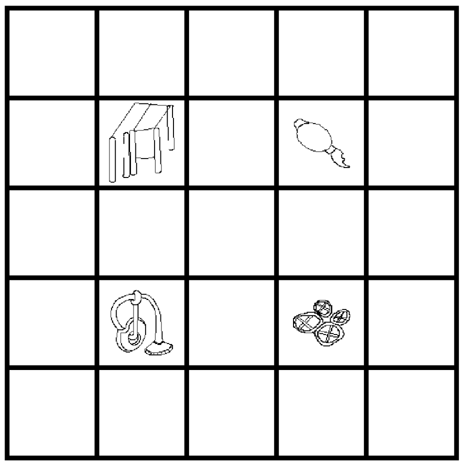


Fig. 3: Example of the screen during the test phase.

2.4. Analysis

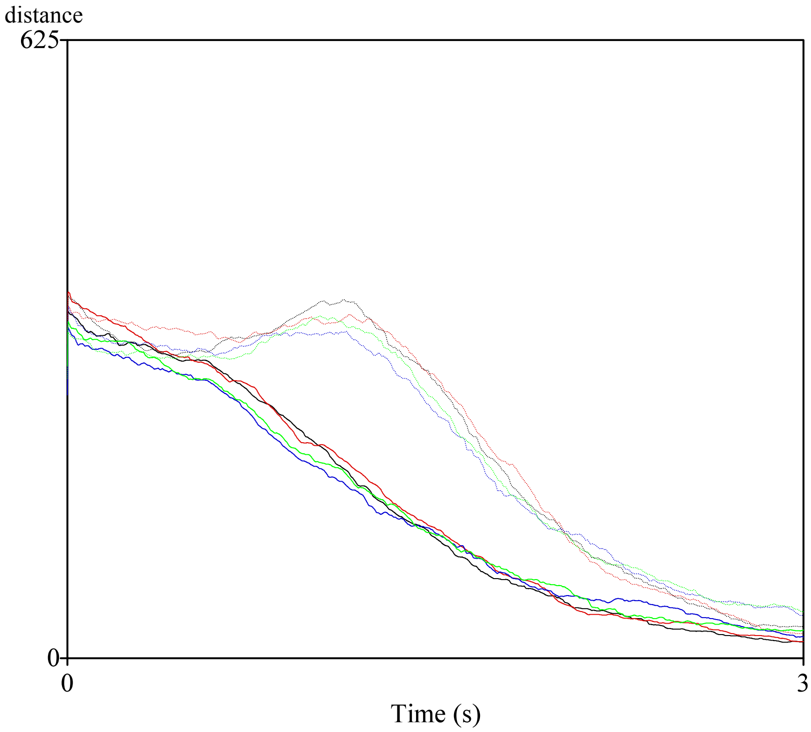
In the testing phase, data regarding eye position were recorded. These eye-tracking data were sampled at 120 Hz. For each trial by each participant, the eye-position distances to the target picture and to the competitor picture (more specifically, to the middle of the grid containing each picture) were then calculated from the onset of the sound to the point where the participant clicked any of the pictures. For missing eye data (e.g. blinks) the last recorded eye position previous to the missing data was used to 'cover' the gap; if this had not been done, the eye-position distance to both target and competitor for the period of the missing data would have appeared at maximum distance (from both target and competitor), which would have made for strange viewing, as the data for multiple target words could not simply be averaged out; the length of the first syllable was not the same for all words, therefore 'coda words' (such as [bɛʀzela] and [bɛɹzesɔ]) could not simply be grouped together. So the time of onset of the first disambiguating phoneme (/t, f, ʀ/ or /ɹ/), or the 'time of first disambiguation', was not be the same for all words. Therefore, the time of first disambiguation *itself* was taken as the starting point for the analysis. So, for example, for [bɛɹzesɔ] the starting point was the onset of [ɹ], and *not* the beginning of the word. If, then, there is any effect of this first disambiguation on recognition of the target word, it can be seen in the averaged results.

Thus, eye-distance data for both the target and the competitor picture for the 8 trials of each test word each participant completed were then averaged out, so that for any one participant, data for the target eye distance and the competitor eye distance for each *word* (as opposed to each *trial*) could be seen (these, averaged out per group, can be seen in graph 1). Next, the difference between these target and competitor eye-distance data, for each word of each participant, were calculated.

These data were then averaged out (for each participant individually) for the /t, f/ words and the /ʀ, ɹ/ words. They were also averaged out for the /t/ words, the /f/ words, the /ʀ/ words, the /ɹ/ words, words with /t/ in onset position, the same for /f, ʀ/ and /ɹ/, and words with /t, f, ʀ/ or /ɹ/ in coda position.

2.5 Results

Graph 1: eye distance to target and competitor



Each curve is an average over 8 trials of 8 words of 7 participants.

Dutch group (group 1)	*Limburgish group (group 2)*
Red solid line: /ʀ, ɹ/ target	*Green solid line: /ʀ, ɹ/ target*
Red dotted line: /ʀ, ɹ/ competitor	*Green dotted line: /ʀ, ɹ/ competitor*
Black solid line: /t, f/ target	*Blue solid line: /t, f/ target*
Black dotted line: /t, f/ competitor	*Blue dotted line: /t, f/ competitor*

As can be seen in graph 1, participants generally start looking at the target rather than the competitor starting at about 650 ms, after having heard the second disambiguation. The difference made by the second disambiguation (i.e. the last syllable, such as [ka] in [z̡ɛtsɔka] as compared to [li] in [z̡ɛfsɔli]) is therefore in any case larger than that made by the first disambiguation (i.e. the ‘target’ phoneme /t, f, ʀ/ or /ɹ/). The fact that eye distance seems to move closer to the competitor towards the end is caused by the fact that from about 1000 ms or so (at the competitor distance 'peak'), participants generally start clicking. As there is no eye data after clicking, the value is set to zero from the moment of clicking. Some participants have, in some cases, clicked after 3000 ms, hence none of the values reaches zero within the right-hand boundary of 3000 ms set in this graph.

Since seemingly random variation in distance can be seen right from the start of graph 1 (possible reasons for this are given in the discussion section), the eye distance to the competitor picture minus the eye distance to the target picture from 0 to 1 s after the onset of the first disambiguation, conflated for each word for each participant (that is, for 8 trials of each word), was calculated. This can be seen in appendix B. To test for significance, t-tests were performed for different types of conflated results, as outlined in paragraph 2.4.

2.5.1. Differences between groups

The mean general difference (i.e. for all trials of all words, as given in appendix B, together) between all participants of group 1 (Dutch) and all participants of group 2 (Limburgish) for the 0 to 1 s window was -9.01 pixels, meaning that the eye distance between target and competitor was generally larger (within the 0 to 1 s window) for the Limburgish group. Using a two-tailed t-test this was found to be not significant ($t = -0.77$, $df = 12$, $p = 0.45$, 95% confidence interval from -34.44 to 16.41).

2.5.2. /t, f/ versus /ʀ, ɹ/

The mean eye-distance differences between target and competitor pictures were also calculated for the following combinations of words (again, a two-tailed t-test was used to calculate significance levels): firstly the mean difference between the Dutch group and the Limburgish group for the /t, f/ words, which was was -23.53 pixels. This was not significant ($t = -1.76$, $df = 12$, $p = 0.10$, 95% confidence interval from -52.66 to 5.59). Secondly, the mean difference between the groups for the /ʀ, ɹ/ words, which was 5.51 pixels, was not significant either (t = 0.40, df = 12, p = 0.70, confidence interval from -24.54 to 35.56).

The mean eye-distance differences between different *conditions,* as opposed to between the two groups, were calculated for the following combinations of conditions (a two-tailed t-test was used to calculate significance levels);

Between the /t, f/ and the /ʀ, ɹ/ words, for both groups together the mean difference was 8.31 pixels. This was not significant (t = 1.07, df = 13, p = 0.31, 95% confidence interval from -8.53 to 25.15). For the Dutch group the difference was -6.21 pixels, which was not significant (t = -1.01, df = 6, p = 0.35, confidence interval from -21.26 to 8.84): For the Limburgish the difference was 22.83 pixels, which was not significant either (t = 1.83, df = 6, p =0.06, 95% confidence interval from -7.64 to 53.31).

Between the /t/ and the /f/ words for the Dutch group the difference was -21.01, which was not significant (t = -1.70, df = 6, p = 0.07, 95% confidence interval from -51.24 to 9.23). For the Limburgish group the difference was 5.73, which was not significant (t = 0.39, df = 6, p = 0.36, 95% confidence interval from -30.36 to 41.82). For both groups combined the difference was -7.64, which was not significant (t = -0.77, df = 13, p = 0.46, 95% confidence interval from -29.15 to 13.88). For the Dutch group, between the /ʀ/ and the /ɹ/ words the difference was 19.01, which was not significant (t = 1.29, df = 6, p = 0.12, 95% confidence interval from -17.08 to 55.10). For the Limburgish group the difference between the /ʀ/ and the /ɹ/ words was -3.28, which was not significant (t = -0.13, df = 6, p = 0.45, 95% confidence interval from -65.32 to 58.77). For both groups together, the difference was 7.86, which was not significant (t= 0.54, df = 13, p = 0.59, 95% confidence interval from -23.30 to 39.03). It can be concluded, then, that no significant results have been found to answer the research question.

2.5.3. Onset versus coda position of /ʀ, ɹ/

/ʀ/ or /ɹ/ occurred in onset position in some words and in coda position in other words (for example in [ʀistɨn] /ʀ/ occurs in the onset, whilst it occurs in coda position in [qɑʀdʉsi]), which was done in order to see whether the position of the /ʀ/ or /ɹ/ in the syllable had any effect on discrimination between /ʀ/ and /ɹ/. Therefore, the following tests were also performed (two-tailed t-tests were used to calculate significance levels).

For the Dutch group, the mean difference between /ʀ/ in onset and in coda position was 48.53 pixels, which was significant even at the $\alpha = 0.01$ level ($t = 4.29$, df = 6, **p = 0.005**, 95% confidence interval from 20.85 to 76.20). For the Limburgish group, the mean difference between /ʀ/ in onset and in coda position was 37.74 pixels, which was not significant ($t = 1.46$, df = 6, $p = 0.19$, 95% confidence interval from -25.54 to 101.03). For both groups together, the mean difference between /ʀ/ in onset and in coda position was 43.13 pixels, which was significant at the $\alpha = 0.01$ level as well ($t = 3.16$, df = 13, **p = 0.007**, 95% confidence interval from 13.66 to 72.61). This means that particularly the Dutch participants are better at recognising /ʀ/ in onset than in coda.

The mean difference for the Dutch group between /ɹ/ in onset and in coda position was 18.43 pixels, which was not significant ($t = 1.12$, df = 6, $p = 0.31$, 95% confidence interval from -22.01 to 58.87). The mean difference for the Limburgish group between /ɹ/ in onset and in coda position was 22.37 pixels, which was not significant ($t = 0.67$, df = 6, $p = 0.53$, 95% confidence interval from -59.67 to 104.42). For both groups together, the difference between /ɹ/ in onset and in coda was 20.40, which was not significant ($t = 1.14$, df = 13, $p = 0.28$, 95% confidence interval from -18.41 to 59.21). Therefore, the asymmetry of recognising /ʀ/ better in onset than in coda is not seen for /ɹ/.

3. Discussion

This study attempted to gain knowledge about whether it is easier to split a phonological category into two separate categories for an L2, or whether it is easier to produce two distinct categories for the L2 when one of the sounds is already familiar from the L1 and the other known from another L2. In order to do this, two groups of participants were selected, one group of seven native speakers of Dutch was chosen, all members of which use [ʀ] in onset position (at the beginning of a syllable) and [ɹ] in coda position (at the end of a syllable) for the /r/ phoneme and another group, consisting of seven native speakers of Limburgish, which is a language that has only [ʀ] for the /r/ phoneme and no [ɹ], was chosen. Specifically, the research question was whether it is easier to split the /r/ category, with [ʀ] and [ɹ] as allophones of /r/, into two separate categories for [ʀ] (namely /ʀ/) and [ɹ] (namely /ɹ/) in a fictional L2, or whether it is easier to produce these two distinct phoneme categories for this L2 when [ʀ] is already familiar from the L1 (as /r/) and [ɹ] is known as a foreign implementation of /r/.

Between the two groups no significant difference was found for /t, f/ words, as would be expected, since both languages have /t/ and /f/ as separate phonemes. However, no significant difference was found for the /ʀ, ɹ/ words either. Overall, no significant results were found, except for the discrimination of /ʀ/ in onset versus coda position; both groups together did have a significant difference here, and for the Dutch group separately this was also significant (whilst for the Limburgish group it was not). The Dutch group, and both groups together, were quicker to recognise the /ʀ/ words in onset position than they were in coda position. Neither group had a significant difference between /ɹ/ in onset and in coda, however. The fact that participants are better at discriminating /ʀ/ in onset position than in coda position might be explained by the fact that both groups are familiar with [ʀ] and [ɹ] as being the same phoneme in Dutch, but only in coda position (where some speakers of Dutch, such as the native Dutch participants in this study, would produce [ɹ] and other speakers of Dutch such as the native Limburgish participants (when speaking Dutch) would produce [ʀ]). In onset position, [ɹ] is not used. This might mean that speakers of Dutch, and especially native speakers of Dutch who have the [ʀ]-onset [ɹ]-coda dialect, can see /ʀ/ in onset as a separate phoneme to /ɹ/ in onset, but have a harder time distinguishing /ʀ/ from /ɹ/ in coda as these are, perhaps, seen as interchangeable, due to the variation that occurs in coda position in Dutch.

We would like to suggest a few amendments to the experiment which could possibly have led to more clear-cut results. To minimise the effect of participants looking at a picture before the sound starts only to keep looking at it whilst hearing the sound (which means it is unclear whether they were looking at it due to the effects of the first disambiguation or not and which might have caused the seemingly random variation that can be seen at the beginning of graph 1) we would like to suggest a geometrical form appearing in the middle of the screen before the sound starts, so that participants would look at this form, rather than at any of the pictures. Activating the sound only by participants looking at this geometrical form in the middle of the screen would also probably be preferable, to ensure that they have had enough time to look at the individual pictures and their positions on the screen before the sound starts. In this experiment participants had a fixed amount of time (1.5 s) to look at the pictures between the pictures appearing on screen and the start of the auditory stimuli instead, however it is unclear whether this was enough time for all participants to look at each individual picture.

Some participants also reported feelings of fatigue and loss of concentration due to the duration and repetitiveness of the overall experiment, which may have had an impact on the results. An option here would be to include different filler tasks to introduce variety, although a problem with this would be that it would lengthen the overall experiment, which was already very long (depending on the participant between an hour and ten minutes and an hour and twenty minutes). It would, of course, also be desirable to have more participants in each group, since the samples in this experiment were very small. However, finding many participants was complicated by the duration of the experiment, the specific requirements and the travel distance for some.

Many of the participants also seemed to clearly remember the use of [ʀ] and [ɹ] or ‘two different types of R’ in the experiment directly after the experiment was completed. Some said they had not been aware of this during the experiment but did realise upon being made aware of it afterwards (which is interesting in that, if this is indeed so, they must have stored the two sounds as different somehow, without being aware of it – this would then show that the use of pseudo-minimal pairs had, for these participants at least, its intended effect of taking the conscious emphasis away from the /ʀ, ɹ/ contrast), but others indicated that they had been aware. However, as participants were not explicitly asked about this but rather informally remarked upon it during an explanation of the goal of the experiment they had been participating in, not all participants reported whether they had been aware or not.

4. References

Escudero, P., Hayes-Harb, R. & Mitterer, H.
2008 'Novel second-language words and asymmetric lexical access'. In: *Journal of Phonetics*, vol. 36, pp. 345-360.

Fischer, B.
1992 'Saccadic reaction time: Implications for reading, dyslexia and visual cognition'. In: K. Rayner (Ed.), *Eye Movements and Visual Cognition: Scene Perception and Reading,* New York: Springer Verlag, pp. 31-45.

Flege, J. E.
1995 'Second Language Speech Learning: Theory, Findings, and Problems'. In: Strange, W. (ed.), *Speech Perception and Linguistic Experience*, Baltimore: York Press, pp. 3-45.

Hall, K.C.
2007 'Pairwise perceptual magnet effects' In: *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 669-672

Iverson, P. and Kuhl, P.K.
1995 'Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling'. In: *Journal of the Acoustical Society of America*, vol. 91, pp. 553-562.

Kuhl, P. K.; Williams, K.A; Lacerda, F.; Stevens, K.N. and Lindblom, B.
1992 'Linguistic experience alters phonetic perception in infants by 6 months of age.' In: *Science*, vol. 255, pp. 606-608.

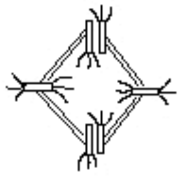
Peperkamp, S., Pattinato M., Dupoux, E.
2003 Allophonic variation and the acquisition ofphoneme categories. In: B. Beachley, A. Brown, F. Conlin (eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development,* vol. 2, Sommerville, MA: Cascadilla Press, pp. 650-661.

Takagi, N.
2002 The limits of training Japanese listeners to identify English /r/ and /l/: Eight case studies. In: *Journal of the Acoustical Society of America,* vol. 111, pp. 2887-2896.

5. Appendix A

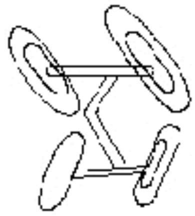
Nonsense words and accompanying pictures

tɨspiŋ



fɨsnel

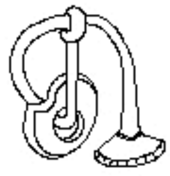
toknɔn



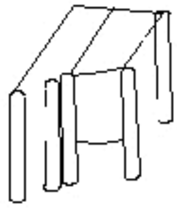
foksɨp

ʐɛtsɔka

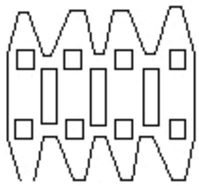
ʐɛfsɔli



ʂutçede

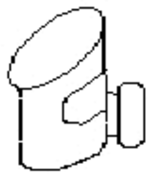


ʂufçeno



ʀistɨn

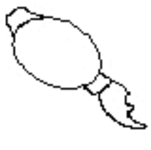
ɹiskoʂ



ʀaltev

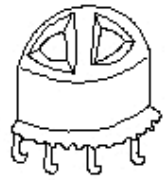
ɹalpɛm

bɛʀzela



bɛɹzesɔ

qɑʀdʉsi



qɑɹdʉka

6. Appendix B

Participant	Language	tɨspiŋ	fɨsnel	toknɔn	foksɨp	z̧ɛtsɔka	z̧ɛfsɔli	ʂutçede	ʂufçeno	ʀistɨn	ɹiskoʂ	ʀaltev	ɹalpɛm	bɛʀzela	bɛɹzesɔ	qɑʀdʉsi	qɑɹdʉka
1	Dutch	20.920	72.289	-79.328	0.954	65.615	-30.752	24.985	104.359	-27.574	131.914	152.409	60.432	81.510	-58.018	-35.207	88.269
2	Dutch	-0.309	-29.726	-39.225	25.877	109.716	-69.422	44.652	37.474	75.675	-22.239	50.916	67.118	91.698	5.599	-15.666	66.546
3	Dutch	-9.038	99.558	-65.274	-24.485	49.802	166.967	-29.647	-61.068	121.300	31.935	-11.204	-14.562	-37.377	85.727	-24.871	-33.194
4	Dutch	9.518	-5.972	82.139	69.542	28.852	-3.420	-57.230	2.226	2.583	-9.726	61.709	64.887	-44.530	-48.087	38.989	93.754
5	Dutch	87.050	-25.802	-61.547	56.049	-111.458	-38.971	57.087	129.294	113.080	-38.460	67.727	-1.831	18.793	17.983	-15.028	-168.820
6	Dutch	48.154	122.051	55.042	18.160	-133.361	26.820	31.977	24.392	-32.006	18.107	97.451	3.002	72.908	14.563	-27.239	71.340
7	Dutch	134.735	82.451	-27.402	-16.520	-21.437	74.125	-0.763	-4.078	66.550	15.736	75.246	30.745	10.145	12.381	20.348	-69.005
8	Limburgish	20.977	-4.146	29.685	166.594	184.590	142.524	81.475	-95.557	10.801	104.209	165.844	-6.778	54.704	-38.317	69.577	81.907
9	Limburgish	75.582	73.595	-29.610	142.229	-4.760	-54.311	-21.292	39.831	35.899	106.989	9.768	6.966	31.509	-104.788	103.038	-2.708
10	Limburgish	16.789	13.377	5.355	26.404	-108.299	18.293	17.435	-109.214	-76.529	63.400	31.092	14.543	8.417	35.269	-0.011	21.209
11	Limburgish	123.248	26.575	-73.742	147.969	95.400	-24.757	9.594	69.085	78.767	-146.886	53.626	-90.259	-100.754	103.988	22.077	-78.043
12	Limburgish	-38.083	153.611	26.512	-14.250	124.885	-22.612	34.065	-80.540	73.279	-51.074	-118.108	-10.982	27.344	-34.458	-95.952	-11.718
13	Limburgish	154.889	-72.082	-55.117	31.137	123.072	123.058	116.256	-29.492	32.953	164.449	64.195	192.544	-36.239	139.879	-158.346	-67.314
14	Limburgish	109.390	160.177	-9.690	110.372	53.882	108.281	185.030	40.842	134.145	109.182	9.468	-12.770	29.298	110.109	22.156	-24.701

Table showing the mean difference, in pixels, between the eye distance to the target and competitor picture for the 0 to 1 s window for each word, averaged over the 8 test trials of this word