Lexical Representation of Non-Native Phonemes

Mirjam Broersma and K. Marieke Kolkman

Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands
mirjam.broersma@mpi.nl

Abstract

This study investigates whether the inaccurate processing of non-native phonemes leads to a not native-like representation of word forms containing these phonemes. Dutch and English listeners' processing of six English phonemes was studied in a phoneme monitoring experiment. Half of the target phonemes were difficult to identify for the Dutch listeners. Lexical mediation was found to play a similar role for the Dutch and the English listeners, and there were no differences in the amount of lexical mediation for 'difficult' and 'easy' phonemes for the Dutch listeners. This suggests that the inaccurate processing of non-native phonemes does not necessarily lead to a not native-like representation of word forms containing these phonemes.

1. Introduction

The comprehension of speech is a hugely complex task, which is often performed without any noticeable difficulty. However, the ease with which listeners understand speech in their native language stands in glaring contrast with the effort it often takes to understand non-native speech. In fact, the ease of the comprehension of native speech and the difficulty of the comprehension of non-native speech are two sides of the same medal.

As infants tune into their native language, they develop a language-specific way of listening, which is helpful for the comprehension of that particular language, but an impediment for the comprehension of other languages.

One way of tuning into the native language is the formation of phoneme categories. Native language phoneme categories enable listeners to restrict their attention to those contrasts that are functionally relevant in the native language. As a result, some non-native phonemes become very hard to distinguish (see [1] for a review). According to the Perceptual Assimilation Model [2], adult listeners assimilate non-native phones to native phoneme categories, assigning them to the category to which they are phonetically most similar. The extent to which non-native contrasts can be distinguished therefore depends on the occurrence and phoneme status of the sounds in the native language.

Broersma [3] investigated the recognition of non-native words containing confusable phonemes. Five pairs of English phonemes were used which were predicted to be hard to distinguish for Dutch listeners. The vowels [æ] and [ɛ], which are expected to be perceptually assimilated to a single Dutch phoneme category, formed one confusable pair. The others were four pairs of voiced and voiceless obstruents which exist as Dutch phonemes as well, but due to phonotactic constraints the distinction between voiced and voiceless obstruents is not functionally relevant at the end of a word in Dutch. In an auditory lexical decision experiment, Dutch and English participants were presented with English words containing a confusable phoneme (e.g. '*cat*'). Besides words in their correct pronunciation ([kæt]), they were also presented with 'near-words', in which the correct phoneme was replaced with its counterpart ([kɛt]). The Dutch listeners gave more 'yes' responses (indicating that they recognized an item as a word) to the real words than to the near-words, which suggests that they noticed a mismatch between the input and the lexical representation for the near-words. However, the Dutch listeners gave significantly more 'yes' responses to the near-words than the English listeners. This could be explained by the Dutch participants' inadequate recognition of the phonemes, but it could also be caused by an inaccurate specification of the phoneme in the representation of the word form.

Another experiment [3] investigated the recognition of pairs of words that only differ in a confusable phoneme (e.g. '*flash*' - '*flesh*'). In this bimodal priming experiment, Dutch and English participants made lexical decisions on visual targets that were preceded by an auditory prime word. The prime was either identical to the target (prime: '*flash*', target: '*flash*') or contained the other phoneme of the pair (prime: '*flesh*', target: '*flash*'). The identical case resulted in priming for both the Dutch and the English participants. For the English participants, the mismatching case caused inhibition. However, for the Dutch participants, the mismatching case did not lead to inhibition, nor to priming. This was interpreted as a result of unresolved competition between the prime and the target word. As the outcome of the identical and mismatching conditions were different for the Dutch participants, they must have had two separate representations for the two words of a pair. The fact that no inhibition was observed can be explained in two ways. One explanation is that the phoneme categories were less distinct for the Dutch than for the English participants, and may even have overlapped. The other possible explanation is that the representation of the confusable phoneme at the lexical level was less accurate for the non-native listeners than for the native listeners.

In this paper, we will consider whether the inaccurate processing of particular non-native phonemes leads to a not native-like representation of word forms containing these phonemes. Weber and Cutler [4] argue that listeners whose perception of confusable non-native phonemes is not native-like may nevertheless have native-like lexical representations of words containing these confusable phonemes. We will present the results of an English phoneme monitoring experiment with Dutch and English participants. In this experiment, the processing of six English phonemes was studied. These are the vowels [ɑ:] and [i:], and the plosives [p], [t], [b], and [d].

Of the vowels, [i:] is very similar to the Dutch phoneme [i], which is phonetically short but phonologically long [5]. [ɑ:] on the other hand does not match well with any single Dutch phoneme. It resembles the two Dutch phonemes [ɑ], from which it differs in length, and [a:], which is more central than the English vowel. Because of its resemblance to two Dutch phonemes, the [ɑ:] is expected to be hard to identify for Dutch listeners (see also [6]). All of the four plosives are phonemes in Dutch, but the voiced plosives [b] and [d] cannot occur word-finally. Due to this phonotactic constraint, the distinction between [p] and [b], and between [t] and [d] is not functionally relevant at the end of a word. In Broersma's experiments described above [3], voicing of word-final plosives was compared with phoneme distinctions that do not exist in Dutch. The patterns of word recognition were no different for items that contained phonemes in positions that are phonotactically illegal in the native language and items containing phonemes that are not distinguished in the native language.

The phoneme monitoring paradigm addresses both pre-lexical and lexical levels of processing. Several studies have found shorter reaction times for phonemes that occur later in the word. This is usually interpreted as an indication for lexical mediation (see [7] for a review).

The experiment reported here was based on two phoneme monitoring experiments by Van Ooijen [7]. These experiments were constructed in such a way that lexical involvement was expected. The materials were varied, containing mono-, di-, and trisyllabic words with different stress patterns, and a large number of fillers that did not contain the target sound. The words were presented in lists of varying length, that did not always contain the target sound. All these factors favor the occurrence of lexical mediation [8].

If the representation of a 'difficult' phoneme is not native-like at the word level, lexical involvement will not speed up the recognition of such a phoneme. If word forms are native-like, a lexical effect is expected. For the vowels, there is one 'difficult' phoneme and one which is easy to identify for the Dutch listeners. Not native-like representation of word forms containing an [ɑ:] should therefore show as an interaction between position and phoneme for the Dutch listeners, but not for the English listeners. For the consonants, voicing is not distinctive in word-final position in Dutch. If this is reflected in not native-like lexical representations, reaction times to word-final consonants should not be faster than reaction times for initial and medial position for the Dutch listeners, whereas they should be faster for the English listeners.

2. Method

2.1. Participants

Thirty-six native speakers of Dutch and thirty-six native speakers of British English took part in the experiment. The Dutch participants had English as a second language, whereas the English participants did not know any Dutch. The Dutch participants were recruited through the Max Planck Institute participant pool, and the English through the participant pool of the University of Sussex. None reported any hearing loss. All were volunteers and received a small fee for participation.

2.2. Materials

For each target phoneme, 36 experimental words were used. These are the same items as were used in Van Ooijen's Experiments 1 and 2 [7]. Twelve of these words contained the target phoneme in word-initial position, 12 in medial position, and 12 in word-final position. The mean frequency of the target-bearing words was matched across the three positions and across target phonemes. Target-bearing words were either monosyllabic or disyllabic. For the vowels, 20 were monosyllabic and 16 disyllabic, for the voiceless consonants, 16 were monosyllabic and 20 disyllabic, and for the voiced consonants, 14 were monosyllabic and 22 disyllabic.

For each target phoneme, nine additional monosyllabic or disyllabic target-bearing words were chosen as dummy targets. In each block, between 183 and 189 mono-, di- or trisyllabic words that did not contain the target phoneme served as fillers.

The materials were recorded onto digital audiotape by a male native speaker of British English in a sound-proof booth, and downsampled to 16 kHz during transfer to a computer.

For all experimental words, the onset of the target phoneme was determined on the basis of auditory inspection and visual inspection of the waveform and the spectrogram. For the vowels, phoneme onset was defined as the first positive zero crossing into the vocalic period, and for the plosives as the onset of the burst.

2.3. Design

The experiment contained six blocks, one for each target phoneme. Each block consisted of 55 lists of 2 to 6 words. 10 Lists contained no occurrence of the target phoneme. Of the 45 lists that did, 36 were experimental lists, and 9 contained a dummy target. Target-bearing words always occurred in penultimate position. For the experimental lists this was the third, fourth, or fifth word in the list, whereas dummy targets were the first or second word in the list.

The position of experimental target-bearing words within a block and within a list was matched across blocks with regard to the position of the target phoneme, number of syllables, and stress pattern. For example, the fourth item in the fourth experimental list of each block was a disyllabic word with stress on the second syllable and a target phoneme in word-final position. The target-bearing words for [ɑ:], [i:], [p], [t], [b], and [d] in this positions were '*guitar*', '*degree*', '*escape*', '*support*', '*describe*', and '*ahead*'.

Words immediately preceding a target-bearing word were also matched across blocks for number of syllables and stress pattern. Words containing a phoneme that differed in less than two features from the target phoneme did not immediately precede a target-bearing word.

The six blocks were presented in six different orders. They were arranged such that target phonemes that differed with only one phonetic feature from one another were never presented in subsequent blocks. Within each block, lists were separated by an interval of 3000 ms. The presentation rate within a list was 1500 ms per word.

2.4. Procedure

Participants were tested one at a time in a quiet room. They received written instructions in English to press a response button as soon as they heard a previously specified phoneme. Before each block, participants received auditory instructions about the phoneme they were monitoring for in that block, and four examples of words containing the target phoneme in different positions. Each block started with a short practice part, and lasted approximately 8 minutes.

The experiment started with a practice block of approximately 3 minutes, with the target phoneme [l]. There was a break after the third experimental block.

The experiment was controlled with NESU (Nijmegen Experiment Set-Up) experimental software. Stimuli were presented binaurally over headphones.

3. Results

The dependent variable in this study is reaction time (RT), measured from target phoneme onset. RTs shorter than 100 ms or longer than 1500 ms were discarded from the analysis. A total of 91.2 % of all targets was detected within this time frame.

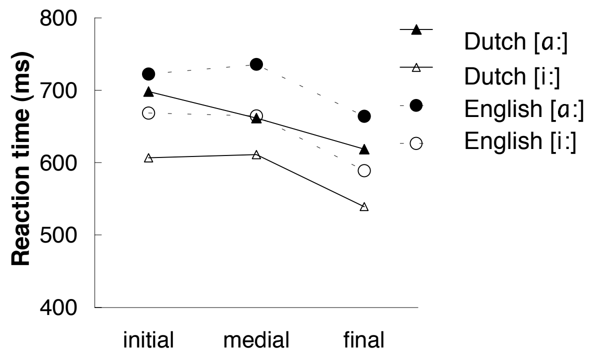

Figure 1: Reaction times to vowel targets.

3.1. Vowel targets

The results for the vowel targets are presented in Figure 1. For the vowels, we were interested in the difference between the two target phonemes, between the three positions, and between the two groups of participants. Therefore, an anova was carried out with RT as dependent variable, phoneme and position as within subjects and between items factors, and language group (Dutch vs. English participants) as between subjects and within items factor.

No interaction was found between phoneme, position and language group (*F1* (2, 140) = 1.67, $p < .193$; *F2* (2, 66) = 1.91, $p < .156$), no interaction between phoneme and language group (*F1* (1, 70) = .07, $p < .798$; *F2* (1, 66) = .52, $p < .474$), no interaction between phoneme and position (*F1* (2, 140) = .86, $p < .426$; *F2* (2, 66) = .16, $p < .854$), and no interaction between position and language group (*F1* (2, 140) = .60, $p < .552$; *F2* (2, 66) = 1.06, $p < .354$). There was, however, a significant main effect of phoneme (*F1* (1, 70) = 61.16, $p < .000$; *F2* (1, 66) = 25.71, $p < .000$), with faster responses to [i:] than to [ɑ:]. There was a significant main effect of position (*F1* (2, 140) = 43.14, $p < .000$; *F2* (2, 66) = 11.26, $p < .000$), the details of which will be described below. The effect of language group was marginally significant by subjects and significant by items (*F1* (1, 70) = 3.55, $p < .064$; *F2* (1, 66) = 77.01, $p < .000$), with faster RTs for the Dutch subjects.

The effect of position was further investigated with three anovas similar to the one just described, comparing two positions at a time. There was no difference between RTs in initial and medial position (*F1* (1, 70) = .19, $p < .662$; *F2* (1, 44) = .12, $p < .731$). RTs in final position were shorter than those in initial position (*F1* (1, 70) = 76.76, $p < .000$; *F2* (1, 44) = 18.88, $p < .000$) and shorter than RTs in medial position (*F1* (1, 70) = 74.02, $p < .000$; *F2* (1, 44) = 18.75, $p < .000$).

These results show that a lexical effect was present and equally strong for the two targets and for the two language groups.

3.2. Consonant targets

As a preliminary step for the analysis of the consonants, it was determined whether the two phonemes of each type (voiceless and voiced) could be collapsed into one factor each. Anovas were carried out for voiceless and voiced consonants separately, with RT as dependent variable, phoneme and position as within subjects and between items factors, and language group as between subjects and within items factor. For both the voiceless and the voiced consonants, no interaction was found between phoneme, position and language group, between phoneme and language group, or between phoneme and position, and no main effect of phoneme. Therefore, the two voiceless consonants are collapsed into one single category in all further analyses, and the two voiced consonants into another.

The results of the consonant targets are presented in Figure 2. A mixed model anova was carried out, with RT as dependent variable, voice type and position as within subjects and between items factors with two and three levels respectively, and language group as a between subjects and within items factor with two levels. A significant interaction between voice type, position, and language was found (*F1* (2, 140) = 8.71, $p < .000$; *F2* (2, 138) = 5.64, $p < .004$).

Separate anovas were carried out for both language groups and both voice types, comparing two positions at a time. The results are presented in Table 1. For both language groups and both voice types, all positions differ significantly, with RTs being longer for initial than for medial position, and longer for medial than for final position.

Dutch	Position	F1(1,35)	p	F2(1,46)	p
Voiceless	in.-med.	19.60	.000	10.89	.002
	med.-fin.	108.30	.000	58.13	.000
Voiced	in.-med.	22.94	.000	10.35	.002
	med.-fin.	175.98	.000	73.18	.000
English					
Voiceless	in.-med.	9.68	.004	5.81	.020
	med.-fin.	369.32	.000	85.87	.000
Voiced	in.-med.	145.21	.000	55.96	.000
	med.-fin.	223.48	.000	59.66	.000

Table 1: Results anova for Dutch and English participants and voiced and voiceless targets separately, comparing positions pairwise ('in.' = initial, 'med.' = medial, 'fin.' = final).

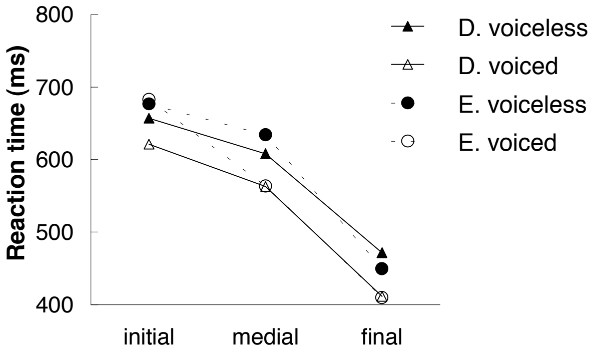

Figure 2: Reaction times to consonant targets ('D.' = Dutch, 'E.' = English).

Next, interactions between voice type and language were investigated for each position separately.

For initial position, an interaction was found between voice type and language (*F1* (1, 70) = 8.79, $p < .004$; *F2* (1, 46) = 9.71, $p < .003$). This interaction was further investigated with four analyses. For the Dutch listeners, RTs were significantly shorter for voiced than for voiceless consonants (*F1* (1, 35) = 9.25, $p < .004$; *F2* (1, 46) = 8.61, $p < .005$). For the English listeners, there was no difference between voiceless and voiced targets in initial position (*F1* (1, 35) = .91, $p < .346$; *F2* (1, 46) = .27, $p < .603$). For the voiceless consonants in initial position, there was no significant difference between the Dutch and the English listeners (*F1* (1, 70) = .40, $p < .528$; *F2* (1, 23) = 5.86, $p < .024$). For the voiced consonants in initial position, RTs of the Dutch listeners were shorter than those of the English listeners (*F1* (1, 70) = 5.00, $p < .029$; *F2* (1, 23) = 35.17, $p < .000$).

For medial position, no interaction was found between voice type and language (*F1* (1, 70) = 3.57, $p < .063$; *F2* (1, 46) = 2.96, $p < .092$). There was a significant effect of voice type (*F1* (1, 70) = 68.42, $p < .000$; *F2* (1, 46) = 11.18, $p < .002$), with longer RTs for voiceless than for voiced consonants. There was no effect of language (*F1* (1, 70) = .21, $p < .645$; *F2* (1, 46) = 2.39, $p < .129$).

Similarly, for final position, no interaction was found between voice type and language (*F1* (1, 70) = 3.32, $p < .073$; *F2* (1, 46) = 2.45, $p < .124$). There was a significant effect of voice type (*F1* (1, 70) = 65.05, $p < .000$; *F2* (1, 46) = 8.89, $p < .005$), with longer RTs for voiceless than for voiced consonants. There was no effect of language (*F1* (1, 70) = .25, $p < .616$; *F2* (1, 46) = 1.83, $p < .183$).

To summarize, the results show a lexical effect for both voiceless and voiced plosives and for both language groups. The interaction between type, position, and language group results from the pattern in initial position. Here, the Dutch participants were faster to respond to voiced targets than to voiceless targets, whereas for the English participants RTs for voiceless targets were equally fast as for voiced targets. In all other positions, RTs were shorter for voiced than for voiceless targets for both language groups. The English listerners' relative advantage for initial voiceless plosives may be explained by their use of aspiration as a cue for voiceless plosives. Aspiration is most salient word-initially, which explains the pattern found for the English listeners. The Dutch listeneres did not benefit from aspiration, as Dutch voiceless plosives are produced with little or no aspiration [9].

4. Conclusions

For the vowel targets, both the Dutch and the English participants showed an effect of lexical mediation on phoneme processing. RTs to targets in word-final position were shorter than RTs to targets in word-initial and medial position. This was the case for both phonemes. There was no difference in the amount of lexical mediation for the two language groups and for the two vowels.

Similarly, for the consonants, lexical effects were found for both language groups. RTs were shorter in medial than in initial position, and shorter in final than in medial position. Again, the amount of lexical mediation was similar for both language groups, and for voiceless and voiced plosives.

The results suggest that the Dutch listeners had native-like word form representations of words that contain a 'difficult' English phoneme. The word form contributed as much to the recognition of the 'difficult' phoneme as it did for the English listeners. Therefore, in line with [4], we conclude that the inaccurate processing of non-native phonemes does not necessarily lead to a not native-like representation of word forms containing these phonemes.

5. References

[1] Werker, J. F., "Cross-language speech perception: Developmental change does not involve loss", in J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words*, pp. 93-120, MIT, Cambridge, MA, 1994.

[2] Best, C. T., "The emergence of native-language phonological influences in infants: A perceptual assimilation model", in J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words*, pp. 167-224, MIT, Cambridge, MA, 1994.

[3] Broersma, M., "Comprehension of non-native speech: Inaccurate phoneme processing and activation of lexical competitors", in *Proceedings of the 7th ICSLP*, pp. 261-264, U. of Colorado Boulder, 2002 (CD-ROM).

[4] Weber, A., and Cutler, A. "Lexical competition in non-native spoken-word recognition", *J. Memory and Language*, 50:1-25.

[5] Booij, G., *The phonology of Dutch*, Oxford University Press, Oxford, 1995.

[6] Escudero, P., and Boersma, P., "The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish", paper presented at the 26th Annual Boston University Conference on Language Development, 2002.

[7] Van Ooijen, B., *The processing of vowels and consonants*, Doctoral dissertation, University of Leiden, 1994.

[8] Cutler, A., Mehler, J., Norris, D.G., and Segui, J., "Phoneme identification and the lexicon", *Cognitive Psychology*, 19:141-177, 1987.

[9] Van Alphen, P.M., *Perceptual relevance of prevoicing in Dutch*, Doctoral dissertation, Nijmegen University.