INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Does the Importance of Word-initial and Word-final Information Differ in Native Versus Non-Native Spoken-Word Recognition?

Odette Scharenborg [1] *, Juul Coumans* [1]*, Sofoklis Kakouros* [2]*, and Roeland van Hout* [1]

[1] Centre for Language Studies, Radboud University Nijmegen, The Netherlands
[2] Department of Signal Processing and Acoustics, Aalto University, Finland

o.scharenborg@let.ru.nl, sofoklis.kakouros@aalto.fi, r.vanhout@let.ru.nl

Abstract

This paper investigates whether the importance and use of word-initial and word-final information in spoken-word recognition is dependent on whether one is listening in a native or a non-native language and on the presence of background noise. Native English and non-native Dutch and Finnish listeners participated in an English word recognition experiment, where either a word's onset or offset was masked by speech-shaped noise with different signal-to-noise ratios. The results showed that for all listener groups the masking of word onset information was more detrimental to spoken-word recognition than the masking of word offset information. The reliance on word-initial information was larger in harder listening conditions for the English but not so for the Dutch and Finnish listeners. Moreover, no significant differences in the use of word-initial and word-final information were found between the two non-native listener groups. Taken together, these results show that the reliance on word-initial information in deteriorating listening conditions seems to be dependent on whether one is listening in one's native or a non-native language rather than on the listener's native language.

Index Terms: non-native spoken word recognition, listening in noise, language dependency.

1. Introduction

During spoken-word recognition, when listeners hear the word 'shape', partially overlapping candidate words such as 'shade' (onset competitor) and 'cape' (offset competitor) will also be activated and will compete for recognition (e.g., [1]). Although listeners use both word-initial and word-final information for candidate word selection and recognition [1-3], word-initial information seems to be more important in evaluating lexical candidates than word-final information, at least in clean listening conditions [1][4]. In the presence of background noise, however, native listeners have been found to adapt the activations of candidate words during competition [4][5] so that offset competitors are activated relatively more compared to clean listening conditions [4][6]. Since a reduction in intelligibility of word-initial information results in the activation of more words than when word-final information is less intelligible, this suggests that listeners rely more on word-initial information in degraded listening conditions compared to clean listening conditions.

A similar pattern was observed for non-native listeners [7]: Dutch listeners use both word-initial and word-final information for the selection and recognition of English words; while, when listening conditions deteriorate, word-initial information is more important than word-final information, and the number of offset competitors in the English language increases relative to the clean condition [8].

In this paper, we investigate whether the importance and use of word-initial and word-final information in spoken-word recognition in deteriorating listening conditions differs in native and non-native listening. Specifically, we investigate whether the importance and use of word-initial and word-final information in spoken word recognition is dependent on 1) whether one is listening in a native or a non-native language, 2) the native language of the listener, and 3) the presence of background noise. To that end, three listener groups with different native languages, i.e., English, Dutch, and Finnish, were tested on a word recognition experiment in which, crucially, word onsets or word offsets were masked by different levels of noise (see also [7]). The target language was English. First, the use of word-initial and word-final information by native English and non-native Dutch listeners is compared by comparing word recognition accuracies. Next, the role of word-initial and word-final information in non-native listening is investigated for native listeners from a language which deviates in lexical structure from English and Dutch: Finnish. Finnish is a highly inflected language in which many different suffixes can be added to a word stem (e.g., a normally inflected Finnish verb can have up to 12000 different forms [9]). Word suffixes are thus critical in conveying word meaning, and correct lexical parsing requires the processing of both word-initial and word-final information. Moreover, the frequency of compounds is high in Finnish, with inflectional markings appearing as infixes and suffixes [10]. In contrast to English and Dutch, word-final information is, arguably, a more important information source in Finnish words.

2. Methodology

2.1. Participants

Table 1 shows the number of participants for each of the listener groups, their mean age, and the proficiency in English as assessed using LexTale [11]. The English participants were recruited from the University of York, UK. The Dutch participants were recruited from the Radboud University, Nijmegen, Netherlands. Note that the Dutch listeners are a superset of the listeners reported in [7]. The Finnish participants were recruited from Aalto University, Finland. None of the participants reported a history of speech and/or hearing disorders. All participants signed a consent form prior to the experiment, and were paid for their participation.

The difference in LexTale between the native and non-native listeners (Dutch: $t(86.4)=14.2$, $p < .001$; Finnish: $t(70.2)=-7.6$, $p < .001$) and the difference between the Dutch and Finnish ($t(103.6)=5.8$, $p < .001$) listeners was significant.

http://dx.doi.org/10.21437/Interspeech.2016-1095

Table 1. *The number of participants, their mean age, and LexTale scores for each of the three participant groups.*

	N	Age		LexTale	
		Mean	SD	Mean	SD
English natives	49	20.8	2.1	93.5	6.2
Dutch non-natives	60	21.4	2.2	65.7	13.5
Dutch, minimally B2	39	21.8	2.2	73.3	9.5
Finnish non-natives	46	27.2	8.7	79.5	11.0

To reduce the role of proficiency differences on possible language differences, we selected only those Dutch listeners with at least a LexTale score of 60 (= upper intermediate level of proficiency [11]). Mean LexTale scores between the Dutch and Finnish are now much closer, although still significantly higher for the Finnish listeners ($t(83.0) = 2.8$, $p = .007$).

2.2. Materials

The stimuli consisted of 42 English triplets, each consisting of a target word (e.g., *letter*), an onset competitor, which shared word-initial information with the target word (e.g., *lettuce*), and an offset competitor, which shared word-final information with the target word (e.g., *sweater*). Each set thus consisted of 126 words in total: 45 bisyllabic words and 81 monosyllabic words (please see [7] for more details). All words within a triplet had the same stress pattern according to Celex [12]. Word frequencies of the target words ranged from 14 to 13,180 per million.

The stimuli were produced by a male native speaker of Southern British English and recorded in a sound-attenuated booth at 44.1 kHz. Subsequently, the audio files were down-sampled to 16 kHz to make them compatible with the noise file. Next, all stimuli were cut manually using Praat [13] into one-word audio files, and intensity was set at 60 dB SPL.

Stationary speech-shaped noise (SSN) was added to all stimuli at three different signal-to-noise ratios (SNRs). The SNRs used in the experiment were -12 dB, -6 dB, and 0 dB. These values were based on separate pilot tests (12 Dutch non-native listeners of English) to avoid floor and/or ceiling effects. None of the participants tested in the pilot studies participated in the main experiment.

The onset and offset masking was tailored to each target word and onset/offset competitor pair, such that the stretch of speech where the target word and the onset/offset competitor overlap phonemically remained unmasked. For example, the English target word *letter* and its onset competitor *lettuce* share /lɛt/, which is thus left unmasked, while /ə/ of /lɛtə/ and /əs/ of /lɛtəs are masked (this condition is referred to as the offset-masked condition). Following the same logic, the target word *letter* and its offset competitor *sweater* share /ɛtə/, which thus remains unmasked, while [l] and [sw] are masked (the onset-masked condition). The mean overlap between the English target words and their onset/offset competitors was 2.5/2.7 phonemes, respectively.

2.3. Procedure

During the experiment, a listener would only receive one masked version of each target word and would not receive the competitor word with the same masking. So, participants received either a target word with word-initial masking and the onset-competitor with word-final masking from the same triplet, or the target word with word-final masking and the offset-competitor from the same triplet. See [7] for details. Participants were tested individually in a sound-treated booth. The stimuli were presented binaurally over closed headphones at 60 dB SPL. The experiment consisted of three parts. One part only contained words with onset-masking; a second part only contained words with offset-masking. The order of these parts was counterbalanced across participants. Each part was divided into three blocks. The first blocks presented in the two parts got the same SNR (e.g., 0 dB), as did the second and third blocks of both parts. Part three always came last, and consisted of all 84 words in the previous two parts without masking. Words within each block were randomized. After every block there was a self-paced pause. Participants had to type in the word they thought they had heard. The experiment lasted approximately fifteen minutes. Afterwards, participants carried out the LexTale task.

2.4. Statistical analyses

Statistical analyses on the word recognition accuracies were carried out using generalised linear mixed-effect models (e.g., [14]), containing fixed and random effects. To obtain the final, best-fitting model containing only statistically significant effects, we used the standard backward stepwise selection procedure as, e.g., described in [7]. The dependent variable was correct versus incorrect recognition. Fixed factors were the SNR (lowest SNR on the intercept), the Position of Noise (onset-masked (reference category) vs. offset-masked), and the listeners' native Language. Moreover, by-participants and by-stimuli random intercepts were added to the model. Prior to the analyses, obvious typing errors were corrected; moreover, homophones received the same orthographic transcription.

3. Results

Figure 1 shows the proportion of correct responses for the four listening conditions (clean and the three SNR conditions) and the two masking conditions for the three listener groups separately. Onset-masking is plotted with the bulleted line; offset-masking is plotted with the line with squares. The diamond symbols indicate the clean listening condition.

As expected, the proportion of correct responses in the clean condition was significantly higher for the native listeners compared to the non-native listeners (Finnish: $\beta = -2.065$, $SE = .244$, $p < .001$; Dutch: $\beta = -2.663$, $SE = .244$, $p < .001$). Moreover, the Dutch listeners made significantly more errors than the Finnish listeners in the clean condition ($\beta = -.612$, $SE = .184$, $p < .001$).

We investigate whether these language differences could still be due to differences in proficiency. Analyses for the three language groups separately showed an effect of proficiency (LexTale) for the Finnish listeners ($\beta = .033$, $SE = .011$, $p < .01$), however not for the Dutch ($\beta = .013$, $SE = .015$, $p = .37$) and the native listeners ($\beta = .039$, $SE = .033$, $p = .23$). Finnish listeners with higher proficiency gave significantly more correct responses. Note that the analyses below were carried out with and without LexTale as a factor. LexTale either fell out of the analyses or it reduced the effect for the factor Language, while leaving all other effects intact.

Comparing the figures for the English, Dutch, and Finnish listeners shows a high resemblance in the effect of deteriorating listening conditions due to the presence of noise, and the effect of masking on word recognition. Nevertheless, differences in the slopes of the different lines can be observed. Below we will analyse the differences and similarities for each language pair. Table 2 shows the parameter estimates in the best-fitting models of performance for the analyses of the three language pairs.

Table 2. *Fixed effect estimates for the best-fitting models of performance for the English and Dutch non-native listeners, n=7392 observations; the English and Finnish non-native listeners, n=7907 obs.; the Finnish (reference category) and Dutch non-native listeners, n=7067 obs.*

Analysis / Fixed factor	English - Dutch		English - Finnish		Finnish – Dutch	
	B	***SE***	**B**	***SE***	**B**	***SE***
Intercept	1.615***	.297	1.843***	.292	.184	.258
Position of Noise	-1.401***	.238	-1.463***	.241	-.732***	.096
SNR	.959***	.133	.908***	.117	1.081***	.102
Language	-.952**	.225	-1.359***	.151	.215	.197
SNR × Position of Noise	.225*	.114	.255*	.114	*n.s.*	*n.s.*
SNR × Language	-.280**	.100	*n.s.*	*n.s.*	-.204*	.088
Position of Noise × Language	.308*	.155	.347*	.153	*n.s.*	*n.s.*

***p<.001; **p<.01; *p<.05

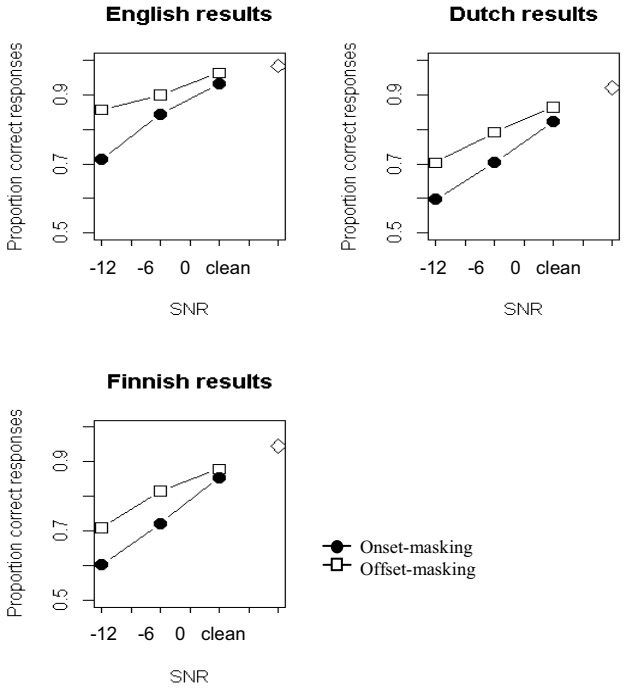

Figure 1. *Proportion of correct responses for the three listener groups, the SNR conditions and the clean condition (diamond symbol), and for the two masking conditions separately.*

3.1. English natives vs. Dutch non-natives

The top panels of Figure 1 show the proportion of correct responses for the English (left) and Dutch (right) listeners. First of all, as expected, an effect of native vs. non-native listening was observed: English listeners gave significantly more correct answers than the Dutch non-native listeners (factor Language), see also the downward shift of the lines in the right top panel compared to the left top panel in Figure 1.

Focussing on our research questions, significantly fewer words were recognised when the onset of a word was masked (bulleted lines in Figure 1; see Position of Noise in Table 2) compared to when the offset was masked (line with squares), and this detrimental effect of onset-masking was larger for the English listeners than for the Dutch listeners (Position of Noise × Language).

With increasingly better listening conditions, significantly more words were recognised by both the native English and the non-native Dutch listeners (factor SNR). Importantly, there is a significant interaction of SNR and Position of Noise, which shows that the difference between the onset-masking and offset-masking conditions reduced significantly for increasingly better listening conditions. In other words, the difference between the onset- and offset-masking conditions increased significantly with deteriorating listening conditions. Moreover, the Dutch non-native listeners suffered more from deteriorating listening conditions than the native listeners, which is shown by the former's larger decrease in proportion correct responses when the SNR decreases from 0 to -12 (for Dutch: 19.3%, for English: 16.1%; SNR × Language). None of the three-way interactions were significant. The maximal random slope structure of the model (and also the models in the other analyses) included a stimulus random slope for SNR (AIC –Akaike Information Criterion– of the model without a stimulus random slope for SNR: 5082.8; AIC of the model with stimulus random slope for SNR: 5012.8), indicating that the proportion of correct responses decreases faster for some stimuli than for others when listening conditions deteriorate.

Analyses of the English and Dutch data separately showed a significant interaction between Position of Noise and SNR for the English ($\beta = .368$, $SE = .183$, $p = .045$) but not for the Dutch non-native listeners ($p = .16$). This suggests that when non-native listeners reach a minimum level of proficiency, the importance of word-initial and word-final information for word recognition does not seem to become more similar to that of native listeners.

A separate analysis in which LexTale was included in the model showed no effect of LexTale, while Language is significant. Language thus seems to capture differences between native and non-native listeners beyond proficiency, possibly related to L1 listening.

3.2. English natives vs. Finnish non-natives

The results of the Finnish group are plotted in the bottom left panel of Figure 1. The results of the English-Finnish analysis (see columns 'English-Finnish' in Table 2) are highly similar to the results of the English-Dutch analysis. Focussing on the, for our research question, crucial variables of Position of Noise and Language, we observe a main effect of both and an interaction between the two variables. Similar to what was found in the English-Dutch analysis, the detrimental effect of onset-masking was significantly larger for the native English listeners than for the non-native Finnish listeners (Position of Noise × Language). However, no interaction was observed between Language and SNR (nor of any of the three-way

interactions). The Finnish non-native listeners did not suffer significantly more from deteriorating listening conditions than the native English listeners. Again, an interaction between Position of Noise and SNR was observed. Interestingly, a separate analysis of the Finnish data alone showed no interaction between Position of Noise and SNR (p = .19), similar to Dutch non-native listeners.

3.3. Finnish non-natives vs. Dutch non-natives

The Finnish-Dutch data comparison (see columns 'Finnish - Dutch' in Table 2) showed that the Dutch non-native listeners suffered more from deteriorating listening conditions than the Finnish listeners (SNR × Language). Importantly, no significant interaction between Language and Position of Noise was observed (nor of any of the three-way interactions). Thus, the size of the detrimental effect of the masking of word-initial information compared to that of word-final information did not differ between the two non-native listener groups, although it did between the non-native listener groups and the native listeners.

No interaction between Position of Noise and SNR was observed. In agreement with the analyses of the language groups separately (see Sections 3.1 and 3.2), the detrimental effect of onset-masking did not increase with deteriorating listening conditions for both non-native listener groups.

4. General Discussion and Conclusions

This paper investigates whether the importance and use of word-initial and word-final information is dependent on whether someone is listening in a native or non-native language, whether these differences are dependent on the listener's native language, and whether there is an influence of deteriorating listening conditions. The results of the three language conditions of the word recognition experiment showed foremost that the role of word-initial and word-final information and a reduced availability of these information sources due to the presence of noise is highly similar in native and high-proficient non-native listening. In both native and non-native listening, the masking of word-initial information is more detrimental to spoken-word recognition than the masking of word-final information. So, in line with previous studies [1-7], word-initial information is more important for successful word recognition than word-final information. Moreover, as expected, fewer words are recognized when listening conditions deteriorate (in line with, e.g., [15]).

Despite the similarities in the role of word-initial and word-final information between the language groups, however, differences were also observed. The size of the detrimental effect of the masking of word-initial information was found to be larger for the native English listeners than for the non-native Dutch and Finnish listeners, while no difference was observed between the Dutch and Finnish non-native listeners. Non-native Dutch and Finnish listeners of English seem to attach less differential weights to word-initial and word-final information than native English listeners.

Although the masking of word onset is more detrimental to spoken-word recognition than the masking of word offset for all tested listener groups, the difference between the two masking conditions increased significantly with deteriorating listening conditions for the English native listeners. This interaction, however, was not significant for the high-proficient Dutch and Finnish non-native listener groups. Thus, a high proficiency in the non-native language does not seem to result in a change in listening strategy towards the strategies used in the non-native language. Possibly, word-initial information is more important for successful word recognition in English than in Dutch or Finnish. If this is true, this would point at a role of listeners' L1 on the use of word-initial and word-final information. An alternative, more likely explanation is, we believe, that non-native listeners rely less on one information source, but rather keep candidate words that match either with the word's onset or its offset alive. Consequently, the effect of noise on word-initial information, although worse than the effect of noise on word-final information, does not further deteriorate. This would be in line with evidence that listeners activate more spurious word candidates during non-native listening than during native listening (e.g., [16]). So, although non-native listeners also rely more on word-initial than on word-final information, they do so less than native listeners. Native listeners, on the other hand, have a stronger tendency to rely more on word-initial information and this is hard to suppress, even when listening conditions are such that a change of strategy would be beneficial. More research is needed to clarify the cause of these different listening strategies in different languages, and whether indeed the native language does not play a role in the use and reliance on word-initial and word-final information in successful spoken-word recognition in different languages.

We only focussed on high-proficient non-native Dutch and Finnish listeners. If language differences would be observed, these would be due to differences in the listeners' native language rather than due to differences in proficiency in English. While language effects were observed between the native and the non-native listeners, no language effects were observed between the Finnish and the Dutch listeners, apart from a larger detrimental effect of harder listening conditions for the Dutch listeners – despite the differences in lexical structure between the two languages. These results seem to hint at differences in the use of word-initial and word-final information between native versus non-native listening rather than differences between native languages.

To conclude, the results for the three listener groups are highly similar: both word-initial and word-final information are important for successful word recognition, where word-initial information is the most important. This is even the case for Finnish, which is a highly inflective language where word-final information is highly relevant for speech understanding, suggesting a language-independent importance of word-initial information for successful word recognition. The reliance on word-initial information was relatively larger in harder listening conditions for the English, but not so for the Dutch and Finnish non-native listeners. The reliance on word-initial information in deteriorating listening conditions thus seems to be dependent on whether one is listening in one's native or a non-native language rather than on the listener's native language. Possibly, because non-native listeners have more difficulty eliminating possible word candidates early during the recognition process.

5. Acknowledgements

This research is sponsored by a Vidi-grant from the Netherlands Organisation for Scientific Research (grant number: 276-89-003) to O.S. We thank Polina Drozdova and Esther Kroese for their help in co-running this experiment, Alastair Smith for support in recording the stimuli, and Sven Mattys for the use of his lab. J.C. is now at Brain Center Rudolf Magnus, University Medical Center Utrecht.

6. References

[1] P. D. Allopenna, J. S. Magnuson, and M. K. Tanenhaus, "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models", *Journal of Memory and Language*, 38, pp. 419-439, 1998.

[2] L. M. Slowiaczek, H. C. Nusbaum, and D. B. Pisoni, "Phonological priming in auditory word recognition", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 13, no. 1, pp. 64-75, 1987.

[3] M. J. Van der Vlugt, and S. G. Nooteboom, "Auditory word recognition is not more sensitive to word-initial than to word-final stimulus information", *Journal of the Acoustical Society of America*, 81, pp. 41-49, 1986.

[4] J. M. McQueen, and F. Huettig, F., "Changing only the probability that spoken words will be distorted changes how they are recognised", *Journal of the Acoustical Society of America*, vol. 131, no. 1, pp. 509-517, 2012.

[5] B. M. Ben-David, C. G. Chambers, M. Daneman, K. M., Pichora-Fuller, E. M., Reingold, and B. Schneider, "Effects of aging and noise on real-time spoken word recognition: Evidence from eye movements", *Journal of Speech, Language and Hearing Research,* vol. 54, pp. 243-262, 2011.

[6] S. Brouwer, and A. R. Bradlow, "The temporal dynamics of spoken word recognition in adverse listening conditions", *Journal of Psycholinguistic Research*, in press.

[7] J. Coumans, R. van Hout, and O. Scharenborg, "Non-native word recognition in noise: The role of word-initial and word-final information," *Proceedings of Interspeech*, Singapore, pp. 519-523, 2014.

[8] O. Scharenborg, J. Coumans, and R. Van Hout, "The effect of background noise on native and non-native spoken-word recognition," *in preparation.*

[9] K. Suomi, J. Toivanen, and R. Ylitalo, "Finnish sound structure. Phonetics, phonology, phonotactics and prosody", Oulu, Finland: Oulu University Press, 2008.

[10] J. Niemi, "Compounds in Finnish," *Lingue e linguaggio*, vol. 2, pp. 237-256, 2009.

[11] K. Lemhöfer and M. Broersma, "Introducing LexTALE: A quick and valid lexical test for advanced learners of English," *Behavior Research Methods*, vol. 44, pp. 325-343, 2012.

[12] R. H. Baayen, R. Piepenbrock, R., and H. van Rijn, "The CELEX Lexical Database (CD-ROM)", Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA, 1993.

[13] P. Boersma, D. Weenink, D. "Praat: doing phonetics by computer [Computer program]", 2013. Retrieved from http://www.praat.org/

[14] R. H. Baayen, D. J. Davidson, and D. M. Bates, D.M. "Mixed-effects modeling with crossed random effects for subjects and items", *Journal of Memory and Language*, vol. 59, pp. 390-412, 2008.

[15] M. Cooke, M. L. Garcia Lecumberri, and J. P. Barker, "The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception", *Journal of the Acoustical Society of America*, vol.123,pp. 414–427, 2008.

[16] A. Cutler, A. Weber, and T. Otake, "Asymmetric mapping from phonetic to lexical representations in second-language listening", *Journal of Phonetics*, vol. 34, pp. 269–284, 2006.