The effect of non-nativeness and background noise on lexical retuning

Polina Drozdova [1], Roeland van Hout [1], Odette Scharenborg [1, 2]

[1] Centre for Language Studies, Radboud University Nijmegen, The Netherlands

[2] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands

P.Drozdova@let.ru.nl, R.vanHout@let.ru.nl, O.Scharenborg@let.ru.nl

ABSTRACT

Previous research revealed remarkable flexibility of native and non-native listeners' perceptual system, i.e., native and non-native phonetic category boundaries can be quickly recalibrated in the face of ambiguous input.

The present study investigates the limitations of the flexibility of the non-native perceptual system. In two lexically-guided perceptual learning experiments, Dutch listeners were exposed to a short story in English, where either all /l/ or all /ɹ/ sounds were replaced by an ambiguous [l/ɹ] sound. In the first experiment, the story was presented in clean, while in the second experiment, intermittent noise was added to the story, although never on the critical words. Lexically-guided perceptual learning was only observed in the clean condition. It is argued that the introduction of intermittent noise reduced the reliability of the evidence of hearing a particular word, which in turn blocked retuning of the phonetic categories.

Keywords: lexically-guided perceptual learning, non-native listening, noise, liquids

INTRODUCTION

A vital characteristic of the native listener's perceptual system is its flexibility [5]. It reveals itself in many ways, among others the ability to quickly adjust phonetic category boundaries to adapt to ambiguous input (see [18] for an overview) using lexical [16] (or phonotactic [6]) knowledge. This mechanism is termed lexically-guided perceptual learning.

This process was first demonstrated by [16]. In their study, Dutch listeners exposed to lexical items where word-final /f/ sounds were ambiguous between /s/ and /f/ (e.g., *witlo[f/s]*, Eng: chicory), interpreted ambiguous items on an /ɛf-ɛs/ continuum more often as /ɛf/ in a subsequent phonetic categorization task than another group of listeners, who was exposed to the same ambiguous sound but replacing the /s/ in /s/-final words (e.g., *radij[f/s]*, Eng*:* radish*)*. This effect generalized to novel, not yet heard, words [13]. Recent experiments with highly-proficient non-native listeners [7,17] revealed that non-natives can retune their first language phonetic category boundaries (L1; [17]) as well as their second language (L2; [7]) phonetic boundaries as a result of ambiguous L2 input. The non-native perceptual system, therefore, seems to be able to draw on similar flexibility regarding tuning into idiosyncratic speech as the native system. As argued by [18], a properly functioning speech perception system should not only be flexible, but also stable and precise, adapting only when there is clear evidence that adaptation is beneficial. But what is clear evidence? Using an eye-tracking study, McQueen and Huettig [14] demonstrated that the presence of noise in an utterance changed the weight listeners assigned to acoustic information during spoken-word recognition. The authors hypothesized that in the condition where noise was added to the carrier sentences, listeners became less confident about having heard the target words, even though target words and the two preceding and following words were not masked by noise.

Since lexical and phonotactic information is necessary to trigger lexically-guided perceptual learning, noise-induced uncertainty about which word was heard might inhibit or impede lexical retuning. Zhang and Samuel [19] found no retuning for native listeners when the stimuli, with the exception of the ambiguous target sounds, were fully masked by noise. They argued that the noise-induced variability in the acoustic signal reduces the reliability of the variability of the ambiguous sound, thus preventing lexical retuning. In the present study, we explicitly investigate the hypothesis that listeners' confidence in having heard a word plays a role in lexical retuning, by masking our stimuli with intermittent noise, similar to [14]. We thus investigate the flexibility of the non-native perceptual system and test whether retuning can still occur when lexical information is made less reliable due to the presence of intermittent background noise.

Two experiments, one in clean and one in noise, were conducted with native Dutch listeners. In both experiments, listeners were exposed to an English text where all words with an /l/ or all words with an

/ɹ/ sound were made ambiguous. The experiments are based on the experiment described in [7]. In the noise condition, noise was added to fragments in the text, with the target words and at least one word preceding and one word following them fully intact. The listening situation was thus far less disadvantageous than that in [19]. After listening to the story participants in both listening conditions performed a phonetic categorization task.

2. METHOD

2.1. Participants

One hundred fourteen native Dutch participants (26 males, M_{age} = 21.6, SD_{age} = 2.0) with no learning or hearing disorders were recruited from the Radboud University Nijmegen subject pool; 54 participants in Experiment 1 (clean) and 60 in Experiment 2 (noise). All participants received a monetary reward. An additional 22 participants (6 Males, M_{age} = 23.1, SD_{age} = 3.3) took part in two pilot studies.

2.2. Materials

2.2.1. Exposure phase

The materials used in the exposure phase were the same as in [7]: 19 English words containing one /l/ sound and 19 words containing one /ɹ/ sound were included in a short story in English (333 words), in which no other words contained /l/ or /ɹ/. Since lexically-guided perceptual learning was shown to be allophone dependent [15], target sounds always occurred at the onset of the third or fourth syllable. The story was recorded with a male British speaker in three versions. In the first version of the story all target words were pronounced as normal, in the second version all /ɹ/s in the target words were substituted with /l/s (e.g., *memoly*), in the third version all /l/s were substituted with /ɹ/s (e.g., *accumurated*). Two complementary versions of each target word were then morphed together with the STRAIGHT [8] algorithm in Matlab [11] to create an 11-step continuum. Step 0 of the continuum contained the most /l/-like interpretation of the ambiguous [l/ɹ] sound, and step 10 the most [ɹ]-like interpretation.

The most ambiguous step on the continuum was chosen on the basis of a pre-test with Dutch listeners (see [7]), for each target word individually. Subsequently, the most ambiguous words were spliced back into the story. Two versions were created: one version contained natural /l/ words and ambiguous /ɹ/ words, the second version contained natural /ɹ/ words and ambiguous /l/ words.

2.2.2. Test phase

In the test phase, the same two minimal pairs were used as in [7]: *collect-correct* and *alive-arrive*. The words were recorded with the same speaker, who recorded the short story, and morphed according to the above procedure. Five steps of each continuum were included in the test phase: the most ambiguous step, chosen on the basis of the pre-test described in [7], and its two preceding and following steps.

2.3. Adding noise

Speech-shaped noise (SSN) was added to the story at a signal-to-noise ratio (SNR) of 0 dB (based on [4]) using a PRAAT [2] script. Noise was added to fragments of one to four words. The target words and at least one word preceding and one word following them remained intact, in order to retain the acoustic information necessary for lexically-guided perceptual learning to occur.

The length of the noise fragments was determined on the basis of two pilot studies, which had the same set-up as the actual experiment (see Section 2.4). On the basis of the results of the first pilot the amount of noise in the story was increased by increasing the number and duration of the noise fragments. This new set-up was checked in a second pilot. In order to investigate whether listeners were still able to follow the story, a short comprehension test consisting of five questions was included. All eight participants of the second pilot answered at least two comprehension questions correctly, therefore, this version of the story was used in the main experiment. None of the participants in the pretests participated in more than one pretest or in the main experiment.

2.4. Procedure

Participants in both Experiment 1 and 2 were randomly assigned to one of two exposure groups. One exposure group in each experiment was exposed to the version of the story with ambiguous /l/-items, while the other was exposed to the version with ambiguous /ɹ/-items. In the phonetic categorization task, following the exposure phase, participants heard 120 test stimuli divided over 4 blocks. Each block consisted of the five steps of each minimal pair presented three times. They categorized the stimuli as containing an /l/ (left button on the button box) or /ɹ/ (right button on the button box). After performing both tasks participants in Experiment 2 (noise) had to fill in the short comprehension test.

3. RESULTS

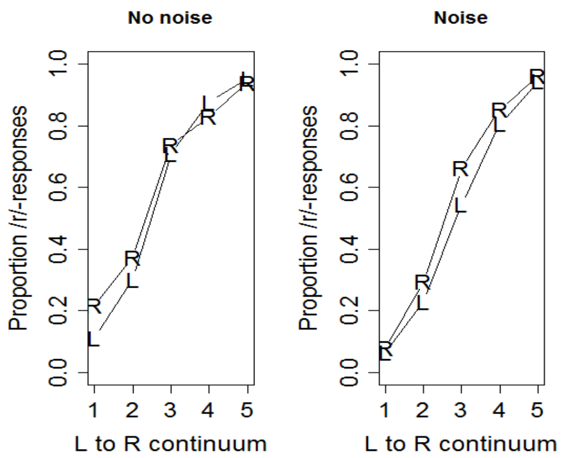
In Experiment 1 (clean), 26 participants listened to the version of the story where all /ɹ/ words were ambiguous, while the rest listened to the story where all /l/ words were ambiguous (amb-r or amb-l version of the story, respectively). This data-set included the 41 participants described in [7] with an additional 13 newly tested participants. In Experiment 2 (noise), 30 participants listened to the version of the story with l-ambiguous words and 30 to the version of the story with /ɹ/-ambiguous words.

The responses of the participants on the phonetic categorization task were analysed using generalized-linear mixed effect models [1]. /l/-responses in the phonetic categorization task were coded as 0 and /ɹ/-responses were coded as 1. We started from the most complex model, which included exposure condition (amb-r or amb-l version of the story), continuum step (step 1 is the most /l/-like and step 5 is the most /ɹ/-like), word pair and listening condition (clean vs. noise) as fixed predictors, and all the possible interactions between them. Subject was added as a random factor, and step was included as a nominal variable. Interactions and predictors that were not significant at the 5% level were subsequently removed one-by-one from the model. Each change of the model was evaluated with a likelihood ratio test (AIC). The best-fitting model contained the least number of factors and interactions, and lowest AIC. Further, by-subject and by-word-pair random slopes and intercepts were added to the best-fitting model to ensure that the found effects were not driven by differences across participants and minimal pairs.

The analysis revealed that the presence of noise during exposure significantly influenced the number of /ɹ/ responses: participants exposed to the story containing noise gave less /ɹ/ responses in the phonetic categorization task (average: 65.3, SD: 14.0; β=-0.772, SE=0.372, $p < .05$) than the listeners in the clean condition (average: 72.6, SD: 14.9). Moreover the three-way interaction between step on the /l/-/ɹ/ continuum, exposure condition, and listening condition (clean vs. noise) was significant for all five steps, indicating systematic differences between the clean and noise listening conditions. To estimate the effects properly, separate analyses for each experiment (clean and noise) were conducted.

Figure 1 shows the proportion of /ɹ/ responses for the two exposure conditions when participants listened to the story in clean (left panel) and with added intermittent noise (right panel). The responses of the participants exposed to the story where all words containing an /l/-sound were ambiguous are labelled with L, the responses of the other group are labelled with R. The difference between the L- and R-lines represents the lexically-guided perceptual learning effect.

Figure 1: Proportion of /ɹ/ responses in the two exposure conditions in Experiment 1 (clean, left panel) and Experiment 2 (noise; right panel).



3.1. Experiment 1 (clean)

Parameter estimates for the best fitting model in Experiment 1 are presented in Table 1. In our discussion of the results we will only focus on those effects that are important for testing our hypothesis. The group of listeners who were exposed to the /ɹ/-ambiguous version of the story gave significantly more /ɹ/-responses in the phonetic categorization task than the group exposed to the /l/-ambiguous version of the story (see Table 1: Exposure condition). The non-native listeners in Experiment 1 thus showed a lexically-guided perceptual learning effect, consistent with the finding reported in [7] on a similar, but smaller group of participants.

3.2. Experiment 2 (noise)

Parameter estimates for the best fitting model in Experiment 2 are presented in Table 2. The number of /ɹ/-responses from the group of participants exposed to the story where all /ɹ/s were ambiguous did not differ significantly from the number of /ɹ/-responses from the group exposed to the version of the story where all /l/s were ambiguous. In the presence of intermittent noise, even if the critical word and its immediate context were not masked by noise, no lexically-guided perceptual learning emerged for the non-native listeners tested in this experiment.

Table 1: Experiment 1: Fixed-effect estimates of performance in the phonetic categorization task of the non-native listeners exposed to the story in clean.

Fixed effect	*β*	SE	*p*<
Intercept	-2.428	0.287	.001
Exposure condition	1.188	0.359	.001
Step 2	1.816	0.209	.001
Step 3	4.092	0.229	.001
Step 4	5.493	0.265	.001
Step 5	5.770	0.297	.001
Word pair	-1.003	0.324	.01
Word pair x Step 2	-0.101	0.238	*ns*
Word pair x Step 3	0.336	0.256	*ns*
Word pair x Step 4	0.339	0.285	*ns*
Word pair x Step 5	2.180	0.367	.001
Exposure condition x Step 2	-0.741	0.240	.01
Exposure condition x Step 3	-1.145	0.260	.001
Exposure condition x Step 4	-1.845	0.290	.001
Exposure condition x Step 5	-1.606	0.352	.001

Table 2: Experiment 2: Fixed-effect estimates of performance in the phonetic categorization task of the non-native listeners exposed to the story in noise.

Fixed effect	*β*	SE	*p*<
Intercept	-2.798	0.202	.001
Step 2	2.241	0.169	.001
Step 3	3.978	0.179	.001
Step 4	5.717	0.219	.001
Step 5	6.052	0.234	.001
Word pair	-1.214	0.339	.001
Word pair x Step 2	-0.868	0.297	.01
Word pair x Step 3	0.014	0.301	*ns*
Word pair x Step 4	-0.169	0.335	*ns*
Word pair x Step 5	2.336	0.410	.001

4. GENERAL DISCUSSION AND CONCLUSIONS

The present study investigates the bounds to the flexibility of the non-native perceptual system in the face of ambiguous speech. In particular, the effects of background noise were investigated. The key question addressed in this study was whether the non-native perceptual system is able to retune when noise occurs elsewhere in the input (never on the critical items). In two experiments, one in clean and one with intermittent noise, Dutch listeners were exposed to ambiguous [l/ɹ] sounds in a short story. Analyses of the results from a subsequent phonetic categorization task showed that retuning occurs in clean listening conditions, but, crucially, no retuning occurs when intermittent noise is added to the story in the exposure phase, even when the target words plus at least one word preceding and one word following them remain fully intact.

Even though participants in Experiment 2 answered on average more than half of the five comprehension questions correctly (M=3.2, SD=1.3), indicating that they could follow the story despite the presence of noise, no retuning emerged. Although much less noise was added to the stimuli in the present study compared to the study by [19] on native listeners, the non-native perceptual system remained stable, as was found for the native listeners in [19]. A possible explanation is based on [14]'s suggestion that the presence of intermittent noise makes listeners less confident about the words they hear. The non-native perceptual system then might have remained stable because the input was not perceived as reliable enough [14]. This might then suggest that the critical items, containing the ambiguous sounds, need to be processed deeply enough for lexical retuning to occur.

The final model for the performance in Experiment 1 appears to be an extension of the model in noise with one critical difference: there is an effect of exposure condition in clean listening conditions that did not occur in the noise experiment. Since in the analysis of the two experiments, the interactions between noise and steps on the continuum were not significant, we put forward that the presence of noise does not so much influence sound processing (apart from increasing the number of /ɹ/ responses overall), but rather prevents the retuning of the phonetic category boundaries. This explanation ties in with [14].

According to the Conservative Adjustment/ Restructuring principle [18,19], the perceptual system remains stable when variation in speech can be attributed to speaker-external factors, such as dialectal variation [9], or when the speaker has a pen in his mouth [10]. Our results provide evidence for another external factor preventing lexical retuning: background noise.

In real-life communicative settings, adaptations to ambiguous speech are needed frequently, because of the high variability of speech input and a multitude of speakers. It is therefore necessary to further probe the balance between stability and flexibility of the perceptual system by further investigating the effect of different types of noise and by comparing different listener groups.

5. REFERENCES

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

[2] Boersma, P., Weenink, D. “Praat. Doing phonetics by computer (Version 5.1),” 2005.

[3] Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics*, 70(4), 604-618.

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

[5] Cutler, A. (2012). Native listening: The flexibility dimension. *Dutch Journal of Applied Linguistics*, 1(2), 169-187.

[6] Cutler, A., McQueen, J. M., Butterfield, S., & Norris, D. (2008). Prelexically-driven perceptual retuning of phoneme boundaries. In Proceedings of Interspeech 2008, 2056-2056.

[7] Drozdova, P., van Hout, R., Scharenborg, O. (2014). Phoneme category retuning in a non-native language, *Proceedings of Interspeech 2014,* Singapore, 553-557.

[8] Kawahara, H., Masuda-Katsuse, I., Cheveigne, A (1999) Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds, Speech Communication, 27, 187-207.

[9] Kraljic, T., Brennan, S.E., & Samuel, A.G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition, 107(1),* 51-81.

[10] Kraljic T, Samuel AG, Brennan SE. First impressions and last resorts: How listeners adjust to speaker variability. Psychological Science. 2008, 194(4), 332–338.

[11] Matlab “R 2013a” (Software). The MathWorks Inc.

[12] Mayo, L.H., Florentine, M., Buss, S. (1997). Age of second-language acquisition and perception of speech in noise. J. Speech, Lang. & Hear. Res. 40, 686-693.

[13] McQueen, J. M., Cutler, A., Norris, D. (2006). Phonological abstraction in the mental lexicon, *Cognitive Science*, 30(6), 1113-1126.

[14] McQueen, J.M., Huettig, F. (2012) Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America*, 131(1), 509-17.

[15] Mitterer, H., Scharenborg, O., McQueen, J.M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, 129 (2), 356-361.

[16] Norris, D., McQueen, J. M., Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204–238.

[17] Reinisch, E.**,** Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 75-86.

[18] Samuel, A. G., Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, 71, 1207–1218.

[19] Zhang, X., Samuel, A.G. (2014) Perceptual learning of speech under optimal and adverse conditions. *Journal of Experimental Psychology: Human Perception and Performance.* 40(1), 200-217.