Speech Communication 126 (2021) 61–70

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

The effect of intermittent noise on lexically-guided perceptual learning in native and non-native listening

Polina Drozdova [a,b], Roeland van Hout [a], Sven Mattys [c], Odette Scharenborg [d,*]

[a] Centre for Language Studies, Radboud University Nijmegen, The Netherlands
[b] IMPRS for Language Sciences, Nijmegen, The Netherlands
[c] Department of Psychology, University of York, UK
[d] Multimedia Computing Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Keywords:
Lexically-guided perceptual learning
Non-native speech comprehension
Background noise

ABSTRACT

There is ample evidence that both native and non-native listeners deal with speech variation by quickly tuning into a speaker and adjusting their phonetic categories according to the speaker's ambiguous pronunciation. This process is called lexically-guided perceptual learning. Moreover, the presence of noise in the speech signal has previously been shown to change the word competition process by increasing the number of candidate words competing for recognition and slowing down the recognition process. Given that reliable lexical information should be available quickly to induce lexically-guided perceptual learning and that word recognition is slowed down in the presence of noise, and especially so for non-native listeners, the present study investigated whether noise interferes with lexically-guided perceptual learning in native and non-native listening. Native English and Dutch listeners were exposed to a story in English in clean speech or with stretches of noise. All the /l/ and /ɹ/ sounds in the story were replaced with an ambiguous sound half-way between /l/ and /ɹ/. Although noise altered the pattern of responses for the non-native listeners in a subsequent phonetic categorization task, both native and non-native listeners demonstrated lexically-guided perceptual learning in both clean and noisy listening conditions. We argue that the robustness of perceptual learning in the presence of intermittent noise for both native and non-native listeners is additional evidence for the remarkable flexibility of native and non-native perceptual systems even in adverse listening conditions.

1. Introduction

There are large variations among speakers in how they produce sounds and words. This is due to differences in the speakers' accent, dialect, speaking style, and idiosyncrasies of their vocal tract or, for instance, because the speaker has a speech impediment. There is ample evidence that listeners deal with this variation by quickly tuning into a speaker, even when pronunciations are ambiguous (e.g., Norris et al., 2003; Reinisch and Holt, 2014). In order to do so, listeners use lexical or phonotactic knowledge (for a review, see Samuel and Kraljic, 2009). The mechanism through which adaptation occurs is called lexically-guided perceptual learning or lexical retuning (Norris et al., 2003).

Lexically-guided perceptual learning was first demonstrated by Norris and colleagues (2003). In their study, Dutch listeners were exposed to words in Dutch with an ambiguous sound half-way between /f/ and /s/, denoted as [f/s], in a lexical decision task. One group of listeners heard /f/-final words where the final /f/ sound was replaced by the ambiguous [f/s] sound (e.g., *witlo[f/s]* - chicory). These listeners learned to interpret this ambiguous sound as an /f/, since the word *witlof* is an existing Dutch word while *witlos* is not. The other group of listeners was exposed to /s/-final words where the final /s/ was replaced by the ambiguous [f/s] sound. These listeners learned to interpret the ambiguous [f/s] sound as an /s/, as the /s/-interpretation of the stimulus is an existing Dutch word while the /f/-interpretation is not (e.g., *baa[f/s]*, where *baas* is a Dutch word (boss) and *baaf* is not). Retuning revealed itself in a subsequent phonetic categorization task, where listeners exposed to the ambiguous items in /f/-final words interpreted stimuli on an [ɛf-ɛs] continuum more often as an [ɛf] than the listeners exposed to the ambiguous /s/-final words, and vice versa.

Exposure to an ambiguous sound triggers a temporary change in listeners' phonetic representations (Clarke-Davidson et al., 2008). Lexically-guided perceptual learning generalizes to words that have not been presented earlier (McQueen et al., 2006), so that, e.g., Dutch adults interpret the previously unheard word *lɔ[f/s]* as *lof* (praise)

* Corresponding author.
E-mail addresses: droz.pol@gmail.com (P. Drozdova), r.vanhout@let.ru.nl (R. van Hout), sven.mattys@york.ac.uk (S. Mattys), o.e.scharenborg@tudelft.nl (O. Scharenborg).

https://doi.org/10.1016/j.specom.2020.12.002
Received 23 June 2020; Received in revised form 28 October 2020; Accepted 6 December 2020
Available online 17 December 2020
0167-6393/

or *los* (loose) depending on their previous exposure to *witlo[f/s]* or *baa[f/s]*, respectively (Scharenborg et al., 2015). This generalization of learning to words not present in the exposure phase strongly suggests that the phonetic category adjustment occurs at the prelexical level of processing (McQueen et al., 2006).

Norris et al. (2003) showed that lexically-guided perceptual learning only occurs when the ambiguous sound is included in an existing word, and not when the ambiguous sound is embedded in a non-word. They concluded that listeners adjust their phonetic category boundaries only when their lexical knowledge can be exploited to interpret ambiguous stimuli. Cutler et al. (2008) extended this proposition showing that ambiguous sounds in non-words can also induce phonetic category retuning, but only when they are part of a legal sequence of phonemes in the listener's native language. Jesse and McQueen (2011) showed that no learning occurs in native listening when ambiguous sounds are located at the start of words, and argued that in order for lexically-guided perceptual learning to occur, lexical knowledge should be available quickly and should be reliable enough to guide retuning. Although words containing the ambiguous sounds were recognized as words (80% acceptance rate on the ambiguous items in the lexical decision task which was used in the exposure phase), the disambiguating information was available too late relative to the position of the ambiguous sound at the start of the word for lexically-guided perceptual learning to occur. This again shows the importance of lexical information for lexically-guided perceptual learning, and suggests that lexical competition should be resolved early enough to trigger retuning.

Because of the essential role of lexical information in lexically-guided perceptual learning, non-native listeners, who have arguably less stable and detailed lexical knowledge than native listeners (Garcia Lecumberri et al., 2010), might possibly be hampered in adapting to ambiguous sounds in a non-native language. Moreover, phonetic categories and contrasts present in the non-native language might be absent or realized differently from those in the native language of the listener (Flege, 1995), which could result in failure to recognize the ambiguous sound or not treating it as ambiguous enough to induce retuning. Proficient non-native listeners do however show lexically-guided perceptual learning, and are able to retune their native and non-native phonetic categories at least when their native and non-native languages are phonologically similar (Bruggeman and Cutler, 2020; Drozdova et al., 2016; Reinisch et al., 2013; Schuhmann, 2015) or when the phonological system of the non-native language is simpler than the listeners' native phonological system (Cutler et al., 2018). Both native and non-native phonetic category representations are thus rather flexible.

There are, however, limits to such flexibility. Samuel and Kraljic (2009) showed that retuning is blocked when variation in the signal can be attributed to speaker-external factors. Kraljic et al. (2008a) demonstrated that acoustic deviations due to context-dependent variability, e.g., caused by a certain dialect (e.g., the pronunciation of /s/ as /ʃ/ when followed by /tr/ in Philadelphia English), prohibited adaptation in native listening. Similarly, no retuning emerges when the ambiguity in the signal is caused by a pen in the mouth of the speaker (Kraljic et al., 2008b). Another speaker-external factor blocking lexically-guided perceptual learning was found to be the presence of background noise. Zhang and Samuel (2014) added signal-correlated noise to their stimuli in the exposure phase, masking both the carrier sentences and the critical lexical items, but not the ambiguous sound (a sound between /f/ and /s/). In contrast to listeners who performed the same task in clean speech, no lexically-guided perceptual learning was observed for listeners exposed to the stimuli masked by noise. Zhang and Samuel (2014) hypothesized that when the speech signal is noisy and hence more variable, native listeners do not treat the ambiguous sound as a reliable cue to trigger retuning.

The presence of noise in the speech signal has also been found to change the dynamics of phonological competition in native listeners (Ben-David et al., 2011; Brouwer and Bradlow, 2011, 2016; Hintz and Scharenborg, 2016; McQueen and Huettig, 2012; Scharenborg et al., 2018). In the study by McQueen and Huettig (2012) participants listened to sentences which were occasionally disrupted by bursts of noise. They hypothesized that the presence of intermittent noise increased listeners' expectation of a distortion occurring, which leads to a change in the lexical competition process. This change manifested itself as an increase in the number of looks to the rhyme competitors and a decrease in the number of looks to the onset competitors in comparison to the clean listening condition. Moreover, the presence of noise has been shown to increase the time listeners need to resolve competition in spoken word recognition (e.g., Ben-David et al., 2011; Brouwer and Bradlow, 2011, 2016). This slowing down is due to an increase in the number of candidate words competing for recognition when noise is present (Scharenborg et al., 2018), a longer activation of the candidate words in the memory of the listeners (Brouwer and Bradlow, 2011), and a reduced activation of the candidate words (Hintz and Scharenborg, 2016). Relatedly, an eye-tracking study with cochlear implant (CI) users (Farris-Trimble et al., 2014) showed differences in the degree of peak and late competitor activation between CI users and a CI simulation group of normal hearing participants. The authors hypothesized that, similar to the participants in McQueen and Huettig (2012), CI users keep competitors active in memory longer as they are expecting degraded input, and consequently delay commitment to lexical items. Listeners are thus able to flexibly adjust their interpretation of acoustic information and consequently their spoken-word recognition processes as listening conditions change (see also Brouwer et al., 2012).

Listening in noise is typically found to be more challenging for non-native than for native listeners (e.g., Mayo et al., 1997; Rogers et al., 2006; Scharenborg et al., 2018; see for a review Garcia Lecumberri et al., 2010; Scharenborg and van Os, 2019). Non-native listeners, therefore, may provide an ideal testing ground for establishing the interaction of two potentially crucial factors in lexically-guided perceptual learning: the characteristics of the speech signal and the lexical knowledge available to the listeners. When the speech signal contains background noise, the phonological match between the target word and the activated words decreases (see Garcia Lecumberri et al., 2010). This leads to an increase of the number of candidate words compared to clean listening conditions, and this increase is even larger for non-native listeners compared to native listeners (Scharenborg et al., 2018).

The present study investigates the effect of intermittent noise on lexically-guided perceptual learning in native and non-native listening. Given the effect of noise on interpreting lexical information in the speech signal, lexically-guided perceptual learning might be impeded in noise. Indeed, Zhang and Samuel (2014) showed that noise throughout the stimulus (with the exception of the critical sound) interferes with lexically-guided perceptual learning in native listening. But what happens when the speech signal is only occasionally disrupted with noise? Based on the findings by McQueen and Huettig (2012) that intermittent noise changes the dynamics of the competition process for native listeners, we hypothesize that such a change potentially delays recognition of the word with the ambiguous sound (Ben-David et al., 2011; Brouwer and Bradlow, 2011, 2016), and subsequently disrupts lexically-guided perceptual learning, especially for non-native listeners, for whom the number of competing candidate words is larger than for native listeners (Scharenborg et al., 2018). Therefore, we predict a negative noise effect for native listeners on lexically-guided perceptual learning and an even stronger negative noise effect for non-native listeners, which could possibly be so strong that no lexically-guided perceptual learning will take place. Our study is a test of the robustness of perceptual learning and will help us to understand the flexibility of native and non-native perceptual systems in adverse listening conditions.

To investigate this hypothesis, lexically-guided perceptual learning was examined in four listening conditions. In the first condition, native listeners of English were auditorily presented with a story (no background noise present) in English in which all /l/ and /ɹ/ sounds were replaced with an ambiguous [l/ɹ] sound. In the second condition, another group of native English listeners were presented with the same story, but this time parts of the story were masked with background noise, while, crucially, words containing the target ambiguous sound were left intact. In the third experimental condition, Dutch non-native listeners of English were exposed to the clean version of the same story as the native listeners. Finally, in the fourth condition, Dutch non-native listeners were exposed to the story with intermittent background noise.

Articulation of /l/ is similar in Dutch and English (Collins and Mees, 1999), while British English prevelar bunched approximant /ɹ/ only occurs in Dutch in coda position (Mitterer et al., 2013; Scobbie et al., 2009; Van de Velde and van Hout, 1999), where it never occurs in English. Dutch listeners would thus have to create a language-specific phonetic category for British English /ɹ/ (Drozdova et al., 2016). After listening to the story, all participants performed a phonetic categorization task.

2. Method

Following the standard procedure for lexically-guided perceptual learning studies (e.g., Norris et al., 2003; Scharenborg et al., 2015; Zhang and Samuel, 2014), all experiments included an exposure phase and a test phase. The exposure phase consisted of a story (Drozdova et al., 2016; Eisner and McQueen, 2006) with a between-participant manipulation (see Appendices A and B). Half of the participants listened to the story where all /l/ sounds were replaced by an ambiguous [l/ɹ] sound, while the other half of the participants listened to the same story where all /ɹ/ sounds were replaced by the ambiguous [l/ɹ] sound. Participants were randomly assigned to one of the two versions of the story. During the test phase, all participants had to perform a phonetic categorization task. To obtain a measure of the lexical proficiency in English of the non-native listeners, LexTALE (Lexical Test for Advanced Learners of English: Lemhöfer and Broersma, 2012) was administered to the non-native listeners. LexTALE is an unspeeded lexical decision task in which participants are exposed to 60 items one-by-one shown on a computer screen and have to decide upon the presentation of each item whether it is an existing word in English or not.

2.1. Participants

One hundred and seventy-six native English speakers (33 males, M_{age} = 20.9, SD_{age} = 2.6), recruited from the Psychology Electronic Experiment Booking system of the Department of Psychology of the University of York, participated in the native versions of the experiment. Two hundred and one native Dutch speakers (35 males, M_{age} = 21.6, SD_{age} = 2.1) were recruited from the Radboud University Nijmegen subject pool and participated in the non-native versions of the experiment. The Dutch participants had an average score of 68.5 (SD = 13.7) on the LexTALE test, which corresponds to an upper-intermediate level of proficiency (B2 level according to the Common European Framework of Reference for Languages: Lemhöfer and Broersma, 2012). The groups of native and non-native listeners who were exposed to the story in clean speech are supersets of those reported in Drozdova et al. (2016).

An overview of the number of participants for each condition is shown in Table 1. The sample size of the native + noise condition is smaller than that for the other three conditions due to recruitment constraints during the testing period. Prior to the experiment, all native and non-native participants filled in a questionnaire with questions regarding any hearing or learning disorders and possible difficulties in listening in the presence of background noise. Only participants without self-reported learning or hearing disorders were included in the experiments. Participants were assigned to only one condition.

Table 1
Number of participants in each listening condition assigned to the /l/-ambiguous and the /ɹ/-ambiguous version of the story in clean speech and in the presence of intermittent noise.

Listeners	Clean listening condition		Noisy listening condition	
	/ɹ/-ambiguous	/l/-ambiguous	/ɹ/-ambiguous	/l/-ambiguous
Native	52	48	39	37
Non-native	47	53	50	51

Additionally, 15 native Dutch participants (3 males, M_{age} = 23.1, SD_{age}= 4.7) took part in a pretest of the stimuli, and another eight native Dutch participants (M_{age} = 22, SD_{age} = 2.8) took part in a pilot study to determine the appropriate length of the noise fragments in the noisy condition. None of these participants were included in the main experiments. All participants received a monetary reward for their participation, and signed a consent form prior to the experiment.

2.2. Exposure phase: clean

The story used in the exposure phase was taken from a previous experiment (Drozdova et al. 2016). It included 19 words containing one /l/ sound and no /ɹ/ sounds, and 19 words containing one /ɹ/ sound and no /l/ sounds. The words were chosen from the CELEX database (Baayen et al., 1995) and had frequencies of at least 100 per million. Since lexically-guided perceptual learning is impeded when listeners hear standard pronunciations of the target sound from the same speaker (Kraljic and Samuel, 2011), no words in the story other than the target words contained /l/ or /ɹ/. As retuning does not transfer to other allophones of the same sound (Mitterer et al., 2013), all /l/ and /ɹ/ occurred in the same position for all target words, i.e., at the onset of the third or fourth syllable (except for one word: *Internet*). The story consisted of 333 words, of which 38 were critical items (see Appendix A for the story). The total duration of the story was 2 min 21 s.

The story was recorded by a male native speaker of British English from South West England in a sound-attenuated booth with a Sennheiser ME 64 microphone at a sampling frequency of 44100 Hz. In order to obtain the ambiguous sound between /l/ and /ɹ/, the story was recorded in three versions. In the first version, all words were pronounced in a natural way. In the second version, all words containing an /l/ sound were pronounced with an /ɹ/ sound (e.g., *accumurated*). In the third version, all /ɹ/ sounds were substituted with /l/ sounds (e.g., *wondeling*). The words were then excised at the positive-going zero crossings from each version of the story and zero-padded with 25 ms silence at the onset and the offset using Praat (Boersma and Weenink, 2009). The pitch contours of the two items from each pair (e.g., *memory-memoly*) were equalized and, following the procedure described by Scharenborg and Janse (2013), morphed with the STRAIGHT algorithm (Kawahara et al., 1999). STRAIGHT first decomposes the input files into source parameters and spectral parameters, and subsequently removes pitch information, while keeping frequency information. In order to keep coarticulatory information of upcoming /l/ and /ɹ/ in the syllable preceding the critical sound available to the listener, whole words were morphed rather than separate sounds. As a result of morphing the item-pairs, an 11-step continuum was created where step 0 was the most /l/-like sound and step 10 the most /ɹ/-like.

To determine the most ambiguous step between /l/ and /ɹ/, a pre-test with 15 Dutch listeners was conducted. We chose the most ambiguous steps on the basis of a pre-test with non-native listeners rather than native listeners to ensure that the chosen steps were indeed ambiguous for the non-native listeners, the group we were primarily interested in.

The pre-test consisted of a phonetic categorization task, where listeners had to decide whether they heard a stimulus with an /l/ or an /ɹ/ sound by pressing the corresponding button on the button box. Participants listened to five different steps of the continuum, i.e., steps

1, 3, 5, 7, 9. The left button of the button box corresponded to the item containing an /l/, whereas the right button corresponded to the item with an /ɹ/. The two possible answers were also presented on a computer screen with the /l/-reading of the stimulus (e.g., wondeling) on the left side of the screen and the /ɹ/-reading (e.g., wondering) on the right side of the screen. In half of the trials, the /l/ answer was an existing word and the /ɹ/ answer a non-word and in the other half of the trials the /ɹ/ answer was a word and the /l/ answer was a non-word. Participants categorized five steps of each critical word (38 words) and test word (4 words: see subsection *Test Phase*). Each step of the continuum was presented twice to the participants. Participants categorized 400 items in total.

The proportions of /l/ and /ɹ/ responses for the test items were calculated. The step on the continuum that received approximately 50% of both responses was chosen as the most ambiguous one. The most ambiguous step was determined individually for each word and then spliced into the corresponding version of the story. Two versions of the story were created: in one version, all words containing an /l/ sound were replaced by the ambiguous [l/ɹ] sound while all /ɹ/ sounds remained natural; in the second version, all words containing an /ɹ/ sound were replaced by the ambiguous [l/ɹ] sound while all /l/ sounds remained natural.

2.3. Exposure phase: noise

For the experiments in the noisy condition, speech-shaped noise was added to the story. For lexically-guided perceptual learning to occur, listeners needed to be able to comprehend the story, hence a signal-to-noise ratio (SNR) was created that challenged listening but did not severely impair recognition accuracy. The SNR was chosen on the basis of a study by Scharenborg and colleagues (2018). In this study, Dutch non-native listeners of English had an average recognition accuracy of 83.8% for English words partially embedded in speech-shaped noise at an SNR of 0 dB. This was deemed an appropriate SNR for our criteria. Following McQueen and Huettig (2012) noise was placed on several fragments of the story, so that at least one word, but typically two words (range 1–4 words), preceding and typically at least one word following the critical word was unmasked.

The noise was automatically added to fragments of the story using a Praat (Boersma and Weenink, 2009) script. First, boundaries were manually placed in the signal on the positive zero crossings in Praat. The fragments that were to be masked were marked with an X on the tier. The Praat script then placed a random chunk of the noise signal on the marked part of the audio file. Before adding noise, the audio file was down-sampled to 16000 Hz to match the sampling frequency of the noise file.

The length of the noise fragments was determined on the basis of a pre-test with eight native Dutch listeners. During the pre-test, participants listened to the partially-masked story, and afterwards had to answer five short questions to check their comprehension of the story. All eight participants answered two to four comprehension questions correctly (M = 3.25), which confirmed that the presence of noise made listening challenging but did not severely harm comprehension. For the noise-added version of the story, see Appendix B.

2.3.1. Test phase

The test phase consisted of a phonetic categorization task. Two minimal pairs, not present in the target story, were used: *collect–correct* and *alive–arrive*. To avoid a bias towards either the /l/ or the /ɹ/ interpretation of the ambiguous stimuli, the two pairs had an opposite pattern of word frequency, with the /l/ word being more frequent for the *alive–arrive* pair (1135 per million for *alive* and 157 per million for *arrive*) and the /ɹ/ word being more frequent for *collect–correct* (117 per million for *collect* and 804 per million for *correct*). The words were recorded by the same speaker who recorded the story. The two members of each word pair were subsequently morphed together using the procedure described in the previous subsection. The test phase in the experiment included five steps from each of the two continua: the most ambiguous step between /l/ and /ɹ/ as determined on the basis of the pre-test, and the two steps preceding and following it. For the *alive–arrive* pair these were steps 3–7, and for *collect–correct* these were steps 2–6.

Table 2
Average percentage of /ɹ/ responses in each experimental condition (with standard deviation in brackets).

Listeners	Clean listening condition		Noise listening condition	
	/ɹ/-ambiguous	/l/-ambiguous	/ɹ/-ambiguous	/l/-ambiguous
Native	65.62% (47.50)	62.62% (48.38)	69.54% (46.03)	62.57% (48.40)
Non-native	62.30% (48.47)	60.19% (48.95)	57.78% (49.40)	53.50% (49.88)

2.4. Procedure

All participants were tested individually in a quiet cubicle or in a sound-attenuated booth. Prior to the experiment, they filled in a consent form and a short questionnaire containing questions about their age, education, and language background. Subsequently, participants were given verbal instructions about the upcoming tasks. Additionally, they saw instructions on the computer screen informing them that they would be hearing a story in English. The story was played to the listeners binaurally through headphones. Once participants finished listening to the story, a message appeared on the screen indicating that they had to press a button on the button box to proceed to the next task. When participants pressed the button, instructions for the test phase of the experiment appeared on the screen.

The test phase was in the form of a phonetic categorization task where participants had to press a button on the button box to indicate which item (*alive* or *arrive*; *collect* or *correct*) they had just heard. The left button on the button box corresponded to the item with the /l/ sound and the right button corresponded to the item with the /ɹ/ sound.

Since the testing setup used with the native group was not equipped with a button box, participants used the "z" key on the keyboard as the left button and the "m" key as the right button. The two response options were also visually presented on a computer screen. Test stimuli were divided over four blocks, with a self-paced pause after each block. Each block consisted of the five steps of each pair presented three times in a random order. Participants thus listened to 120 test items. Exposure and test phases were followed by the LexTALE task for the non-native listeners.

3. Results

To investigate the effect of the presence of background noise on lexically-guided perceptual learning in native and non-native listeners, the responses in the phonetic categorization task were analyzed. We excluded one participant from the analysis (from the group of native listeners, /ɹ/-exposure group, noise listening condition), because due to a technical error her responses from the final block were missing. Table 2 shows the percentage of /ɹ/ responses in each experimental condition.

All analyses were performed in R (version 3.0.2) using mixed effects logistic regression with glmer (package lme4) with the optimizer set to BOBYQA (Powell, 2009) and the number of iterations set to 100000. The dependent variable was the response to the ambiguous sound, where /l/ responses were coded as 0 and /ɹ/ responses as 1. We started the analysis from an overall model including the native and the non-native listener groups in both listening conditions (clean and noise) containing all predictors: Exposure Condition (/ɹ/-ambiguous or /l/-ambiguous version of the story), Noise (whether the story was presented in clean or in noise), Step on the continuum, Language (whether the participant was a native or a non-native listener), Word

Table 3
The meaning of the category labels in the statistical tables.

Category in statistical tables	Actual category	Other category
Exposure Condition1	/l/-ambiguous	/ɹ/-ambiguous
Language1	Dutch	English
Noise1	clean	noise
Word1	*alive–arrive*	*collect–correct*

pair (*collect–correct* or *alive–arrive*) and all possible five-, four-, three- and two-way interactions between them. Step on the continuum was included as a categorical variable,[1] all other variables were recoded using deviation coding.

A backward selection procedure was applied in which interactions and predictors that were not significant at the 5% level were one-by-one removed from the model, starting with the interactions with the highest, non-significant *p* values. Each change in the fixed effect structure was evaluated by inspecting the likelihood ratio changes with the anova function.

Note, that all tables that display the statistical results contain labels referring to experimental variables. Their meaning is given in Table 3. As we applied deviation coding (the standard coding in analyses of variance), the β value of the category not given can be inferred from the category presented. The β values of the categories add up to zero. The estimates of the parameters from the best fitting model are shown in Table 4.

We expected to find a difference between the two exposure groups, which would manifest itself as a significant effect of Exposure Condition. However, as shown in Table 4 the main effect of Exposure Condition was not significant. However, Exposure Condition contributed to two significant two-way interactions with Word and Step. Moreover, a three-way interaction with Word, Language, and Exposure Condition remained in the final model, as its removal significantly decreased the model fit. Given a significant effect of the word pair in the emergence of lexically-guided perceptual learning (Word and its interactions; also found in Drozdova et al. (2016)), we ran separate analyses for the *collect–correct* and *alive–arrive* word pairs.

3.1. Collect–correct

Responses for the *collect–correct* test continuum were analyzed with the same backward selection procedure explained in the previous section (but excluding the factor Word). Fig. 1 shows proportions of listeners' responses for the *collect–correct* test continuum separately for native and non-native listeners and different listening conditions. Table 5 shows the estimates of the parameters that were included in the final model for this analysis.

Contrasting with the main analysis, the analysis for the *collect–correct* word pair revealed a significant main effect of Exposure Condition (see also the factor Exposure Condition in Table 5), which means that the /ɹ/-exposure group gave significantly more /ɹ/ responses in the phonetic categorization task than the /l/-exposure group for the *collect–correct* test continuum. This effect of Exposure Condition was significantly different for the first two steps on the continuum compared to Step 3 (significant interaction between Step and Exposure Condition; first two continuum steps on Fig. 1) and did not depend on the language background of the listeners nor the presence of background noise during exposure (no significant interactions between Exposure Condition and Noise or Exposure Condition and Language in the final best fitting model).

[1] Figs. 1 and 2 show that the responses do not show a linear pattern, which is confirmed by the fact that the AIC with step as a continuous variable has a lower AIC (AIC categorical = 35326.7, AIC continuous = 35792.3; a quadratic polynomial did not improve our model: AIC quadratic = 35555.1).

Table 4
Fixed-effect estimates of the performance of the listeners in the phonetic categorization task.

Fixed effect	*β*	*SE*	*p*
Intercept	0.919	0.068	<.001
Exposure Condition1	−0.127	0.067	0.057
Step1	−2.890	0.034	<.001
Step2	−1.462	0.027	<.001
Step3	0.371	0.027	<.001
Step4	1.330	0.033	<.001
Noise1	0.085	0.068	0.208
Word1	0.422	0.017	<.001
Exposure Condition1 x Step1	−0.230	0.033	<.001
Exposure Condition1 x Step2	−0.130	0.026	<.001
Exposure Condition1 x Step3	0.025	0.026	0.341
Exposure Condition1 x Step4	0.131	0.030	<.001
Language1 x Noise1	0.132	0.067	0.051
Language1 x Exposure Condition1	0.037	0.067	0.575
Language1 x Word1	−0.100	0.017	<.001
Step1 x Language1	0.010	0.033	0.772
Step2 x Language1	−0.072	0.026	0.006
Step3 x Language1	−0.001	0.026	0.952
Step4 x Language1	0.055	0.033	0.086
Step1 x Noise1	0.156	0.033	<.001
Step2 x Noise1	0.070	0.027	0.009
Step3 x Noise1	0.007	0.027	0.780
Step4 x Noise1	−0.087	0.032	0.007
Exposure Condition1 x Word1	0.156	0.014	<.001
Step1 x Word1	−0.072	0.031	0.020
Step2 x Word1	0.267	0.026	<.001
Step3 x Word1	0.066	0.027	0.014
Step4 x Word1	0.287	0.032	<.001
Noise1 x Word1	−0.016	0.016	0.338
Language1 x Step1 x Noise1	0.165	0.033	<.001
Language1 x Step2 x Noise1	0.035	0.026	0.181
Language1 x Step3 x Noise1	−0.011	0.027	0.666
Language1 x Step4 x Noise1	−0.048	0.031	0.117
Step1 x Noise1 x Word1	−0.013	0.031	0.680
Step2 x Noise1 x Word1	−0.073	0.026	0.005
Step3 x Noise1 x Word1	0.011	0.026	0.676
Step4 x Noise1 x Word1	−0.004	0.031	0.910
Step1 x Language1 x Word1	0.085	0.031	0.005
Step2 x Language1 x Word1	0.030	0.026	0.250
Step3 x Language1 x Word1	0.018	0.027	0.500
Step4 x Language1 x Word1	−0.003	0.032	0.920
Noise1 x Language1 x Word1	−0.129	0.015	<.001
Language1 x Word1 x Exposure Condition1	−0.026	0.014	0.060

Native and non-native listeners differed in the overall number of /ɹ/-responses in the phonetic categorization task in the noise listening condition: native listeners gave more /ɹ/ responses in the noisy listening condition than non-native listeners (see significant interaction between Noise and Language and panels on the right in Fig. 1). Additionally, the difference at the first step of the continuum was larger for the native listeners than for the non-native listeners, especially in the noise condition (significant interaction between Language, Step and Noise).

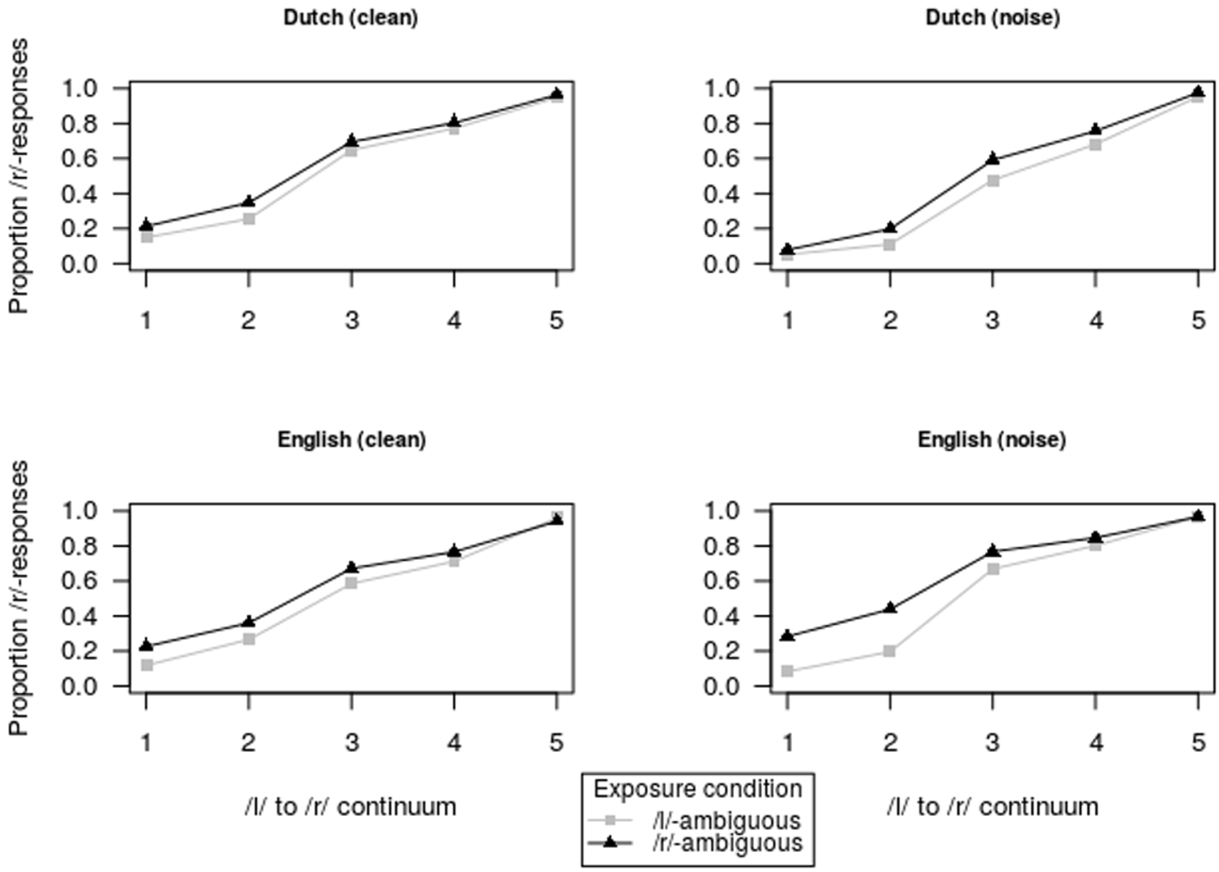

Fig. 1. Proportion of native listeners' (lower panels) and non-native listeners' (upper panels) /ɹ/-responses for the *collect–correct* test continuum in the clean and the noisy listening condition. Responses of participants who were exposed to the /ɹ/-ambiguous version of the story are represented with the black line with triangles. Responses of the participants exposed to the /l/-ambiguous version of the story are shown with the gray line with squares.

3.2. *Alive–arrive*

For the *alive–arrive* test continuum, the estimates of the parameters included in the final model for the native and non-native listeners for both listening conditions together are presented in Table 6.

As can be seen in Table 6, no main effect of Exposure Condition was observed in the analysis for the *alive–arrive* test continuum, although Exposure Condition contributed to a significant interaction with the continuum step. As can be seen in Fig. 2, this significant interaction was not caused by the difference between /ɹ/ and /l/-exposure conditions, but rather by the differences between continuum steps. Although there were no differences between the /ɹ/ and /l/-exposure conditions on the third step of the test continuum, there were slightly more /ɹ/-responses on the first steps of the continuum for the /ɹ/-exposure group than for the /l/-exposure group, while on the last steps of the continuum this difference reversed. In general, irrespective of the native language of the listeners or the listening condition (noise or clean), there was no learning effect for the *alive–arrive* test continuum. However, similar to the *collect–correct* test continuum, native listeners gave more /ɹ/ responses than non-native listeners. This difference, however, was modified by listening condition and continuum step.

4. Discussion and conclusions

The present study investigated the effect of intermittent noise on lexically-guided perceptual learning in native and non-native listening. We hypothesized that intermittent noise has a detrimental effect on lexically-guided perceptual learning, especially for non-native listeners, due to the detrimental effect of background noise on the competition process. However, contrary to our hypothesis, lexically-guided perceptual learning was observed for both native and non-native listeners irrespective of the presence of intermittent noise. Note, however, this effect was only observed for the *collect–correct* word pair while no effect was found for the *alive–arrive* word pair for either listener group. In our discussion, we first focus on the different pattern of responses for the *collect–correct* and *alive–arrive* word pairs, and then discuss the results for the native and non-native listeners in the clean versus the noisy listening condition.

The ambiguous sounds used in the exposure and test phases were chosen on the basis of a pre-test with non-native listeners, as they were our main group of interest. In order to be able to compare the native and non-native listeners' ability to show lexically-guided perceptual learning, the same stimuli were used for both listener groups. Nevertheless, in the present study, lexically-guided perceptual learning was found for both listener groups in both listening conditions for the *collect–correct* word pair, while no lexically-guided perceptual learning was observed for the *alive–arrive* word pair. Since neither listener group showed perceptual learning for the *alive–arrive* continuum in either listening condition, and since perceptual learning has been shown for many different continuums (see for an overview Scharenborg et al. (2019)), including the continuum used in this work Mitterer et al. (2013), it is likely that the lack of a perceptual learning effect for *alive–arrive* was due to idiosyncracies with the steps selected for the *alive–arrive* continuum. Indeed, acoustic analyses in Drozdova et al., 2016 suggest that the steps for *alive–arrive* might not be as well positioned on the continuum as they were for the *collect–correct* pair. In particular, the first step of the *alive–arrive* continuum was found to be more /ɹ/-like than the first step of the *collect–correct* continuum. Moreover, the schwa-initial structure of the *alive–arrive* pair means that

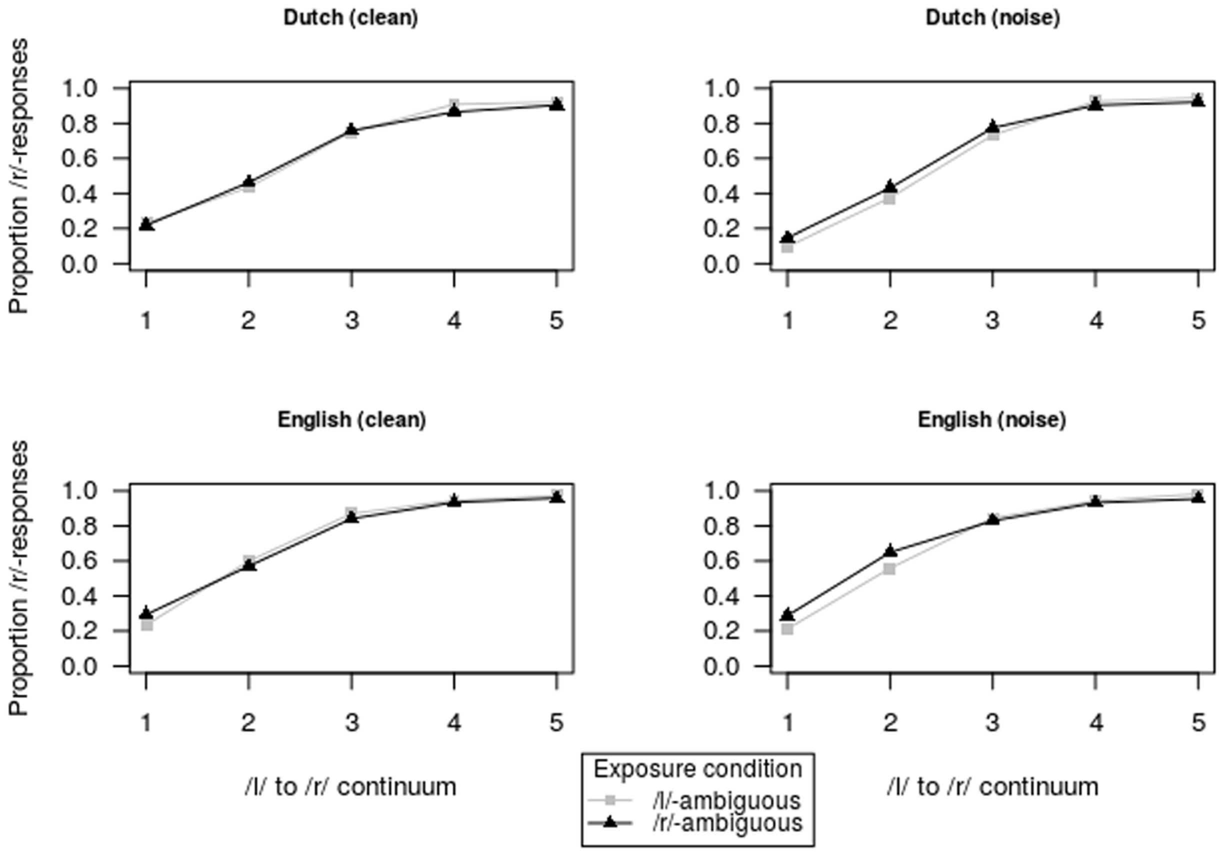

Fig. 2. Proportion of native listeners' (lower panels) and non-native listeners' (upper panels) /ɹ/-responses for the *alive–arrive* test continuum in the clean and noisy listening conditions. Responses of participants who were exposed to the /ɹ/-ambiguous version of the story are represented with the black line with triangles. Responses of the participants exposed to the /l/-ambiguous version of the story are shown with the gray line with squares.

the monosyllabic words *life* and *rife* could have been activated and competed with the carrier words *alive* and *arrive*. This would be consistent with the stress-based segmentation mechanisms assumed in languages with a statistical bias for stress-initial words (Cutler and Norris, 1988; Norris et al., 1995). If the ambiguous [l/ɹ] sound was perceived as word-initial, then it would have been less likely to demonstrate retuning, as we know from Jesse and McQueen (2011). If true, then it might be the case that lexical retuning only occurs or is only revealed in words in which the ambiguous sound is not the start of an existing word embedded in the longer word. Therefore, in our subsequent discussion, we will focus on the *collect–correct* word pair where the [l/ɹ] sound was ambiguous enough to induce lexically-guided perceptual learning in both listener groups.

The results for the clean listening condition are in line with numerous earlier studies (e.g., for native listeners: Eisner and McQueen, 2006; Norris et al., 2003; Scharenborg et al., 2015; and for non-native listeners: Cutler et al., 2018; Drozdova et al., 2016; Reinisch et al., 2013). Both the study by Drozdova et al. (2016) and the present study demonstrate that despite differences in native and non-native listening, relatively proficient non-native listeners are able to retune their non-native phonetic categories as a result of exposure to an ambiguous sound (see Cutler et al., 2018 for a discussion on the role of phonological similarity between the native and non-native language of the listeners on this adaptation process).

With respect to the noise condition, Zhang and Samuel (2014) found that learning was blocked in the presence of noise during native listening, whereas the present study found the opposite. There is, however, an important difference between our study and the study by Zhang and Samuel (2014). During the exposure phase in the Zhang and Samuel study, the entire stimulus was masked by noise with the exception of the critical ambiguous sound. In our study, noise was far less prevalent, since it was never present on the words containing the ambiguous sound and most of the time also not on the words directly preceding and following the critical word. As Zhang and Samuel argued, the wide-spread presence of noise during exposure increased the inherent variability in the pronunciation of the speech sounds. Consequently, the variability of the ambiguous sound, which normally would trigger lexically-guided perceptual learning, would prevent the ambiguous sound from acting as a reliable cue to trigger retuning. In our study, the presence of noise might have increased the variability of the speech signal locally, but it did not reduce the reliability of the variability of the ambiguous sound as a cue to lexically-guided perceptual learning as evidenced by the fact that both the native and the non-native groups of listeners showed retuning in the intermittent noise listening condition. Therefore, our study shows that the presence of background noise does not necessarily disrupt retuning, even when phonological representations and lexical knowledge are non-native as in listening in a non-native language.

McQueen and Huettig (2012) previously demonstrated that the presence of intermittent noise in the speech signal alters the competition process during native spoken word recognition and makes listeners less confident about the words they are hearing. Moreover, Scharenborg et al. (2018) showed that the presence of background noise increases the number of activated words in both native and non-native listening. Keeping multiple word candidates in memory can thus slow down recognition of the target word (Norris et al., 1995) containing the ambiguous sound. However, the present results show that even when intermittent background noise is present in the signal, the crucial lexical information to disambiguate the ambiguous sound is available in time for both native listeners and relatively proficient non-native listeners. Given the role of lexical and phonological knowledge in lexically-guided perceptual learning (Norris et al., 2003; Cutler et al.,

Table 5
Fixed-effect estimates of the performance of the listeners in the phonetic categorization task for the *collect–correct* word pair.

Fixed effect	β	SE	p
Intercept	0.596	0.105	<.001
Exposure Condition1	−0.355	0.104	0.001
Step 1	−3.429	0.062	<.001
Step 2	−2.085	0.047	<.001
Step 3	0.431	0.041	<.001
Step 4	1.320	0.044	<.001
Noise1	0.137	0.105	0.192
Language1	−0.228	0.105	0.031
Exposure Condition1 x Step1	−0.265	0.060	<.001
Exposure Condition1 x Step2	−0.174	0.046	<.001
Exposure Condition1 x Step3	0.061	0.040	0.129
Exposure Condition1 x Step4	0.141	0.043	0.001
Step1 x Noise1	0.150	0.059	0.012
Step2 x Noise1	0.135	0.046	0.003
Step3 x Noise1	−0.012	0.041	0.767
Step4 x Noise1	−0.094	0.043	0.031
Step1 x Language1	−0.081	0.059	0.176
Step2 x Language1	−0.133	0.046	0.004
Step3 x Language	−0.043	0.041	0.281
Step4 x Language	0.058	0.044	0.188
Noise1 x Language1	0.300	0.105	0.004
Language1 x Step1 x Noise1	0.164	0.059	0.006
Language1 x Step2 x Noise1	0.053	0.046	0.248
Language1 x Step3 x Noise1	0.050	0.041	0.218
Language1 x Step4 x Noise1	−0.021	0.044	0.623

Table 6
Fixed-effect estimates of the performance of the listeners in the phonetic categorization task for the *alive–arrive* word pair.

Fixed effect	β	SE	p
Intercept	1.430	0.080	<.001
Exposure Condition1	0.044	0.079	0.578
Step1	−3.191	0.049	<.001
Step2	−1.293	0.038	<.001
Step3	0.485	0.042	<.001
Step4	1.747	0.055	<.001
Noise1	0.086	0.081	0.283
Language1	−0.423	0.080	<.001
Exposure Condition1 x Step1	−0.258	0.046	<.001
Exposure Condition1 x Step2	−0.142	0.037	<.001
Exposure Condition1 x Step3	−0.019	0.040	0.633
Exposure Condition1 x Step4	0.187	0.052	<.001
Step1 x Noise1	0.109	0.047	0.021
Step2 x Noise1	−0.027	0.038	0.471
Step3 x Noise1	0.040	0.041	0.339
Step4 x Noise1	−0.054	0.054	0.322
Step1 x Language1	0.109	0.047	0.020
Step2 x Language1	−0.047	0.037	0.213
Step3 x Language1	0.019	0.041	0.652
Step4 x Language1	0.059	0.054	0.276
Noise1 x Language1	0.007	0.080	0.921
Step1 x Noise1 x Language1	0.173	0.047	<.001
Step2 x Noise1 x Language1	0.086	0.038	0.023
Step3 x Noise1 x Language1	−0.041	0.041	0.324
Step4 x Noise1 x Language1	−0.123	0.054	0.023

2008), and the influence of the presence of background noise on the interpretation of this information (Ben-David et al., 2011; Brouwer and Bradlow, 2011, 2016; McQueen and Huettig, 2012; Scharenborg et al., 2018), future studies should increase the length of the noise fragments and/or reduce the SNR to determine the conditions under which lexically-guided perceptual learning is fully disrupted as was found in the study by Zhang and Samuel (2014).

Native and non-native listeners in the present study were surprisingly similar in how they dealt with the ambiguous sound in the *collect–correct* test continuum. The only observed difference between the two groups was the number of /ɹ/-responses in the noisy listening condition: native listeners gave significantly more /ɹ/-responses than non-native listeners. This result suggests that the presence of noise changed the perception of the target sound (despite not being masked by noise) for the non-native listeners, but since this change occurred for both the /l/-exposure and the /ɹ/-exposure group, no difference in lexically-guided perceptual learning between clean and noisy listening conditions was observed. Broersma and Scharenborg (2010) previously demonstrated that the presence of noise affects Dutch listeners' perception of English /ɹ/ to a greater extent than native listeners' perception. Apparently, this difference is present even when noise occurs intermittently, and can be observed even in the subsequent perception of /l/ and /ɹ/ when listeners hear the items in clean.

Previous studies underlined a number of factors impeding lexically-guided perceptual learning in native listening, such as variability attributed to a certain dialect (Kraljic et al., 2008a) or a pen in the mouth of the speaker (Kraljic et al., 2008b), initial position of the ambiguous sound in the word (Jesse and McQueen, 2011), or the presence of background noise, which covered all the stimuli except the target sounds (Zhang and Samuel, 2014). The present study demonstrates that the presence of intermittent noise does not fall into the group of these impeding factors, as lexically-guided perceptual learning remains robust in native and non-native listening irrespective of the listening conditions. This is an important finding showing that the perceptual system of non-native listeners can remain as flexible as that of native listeners even in harder and challenging listening conditions.

CRediT authorship contribution statement

Polina Drozdova: Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Project administration. **Roeland van Hout:** Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Supervision. **Sven Mattys:** Resources, Writing - review & editing. **Odette Scharenborg:** Conceptualization, Methodology, Resources, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by a Vidi-grant from the Netherlands Organization for Scientific Research (NWO; grant number 276-89-003) awarded to Odette Scharenborg. The here presented experiment was carried out while P. D. and O. S. were at the Centre for Language Studies, Radboud University, Nijmegen, the Netherlands. We would like to thank Alastair Smith for being the voice of our experiment, and Joop Kerkhoff for creating the Praat script.

Appendix A. Clean version of the story used in the exposure phase, with target words indicated in boldface.

He opened the magazine, **immediately**[2] saw his own name, and began **wondering** how many fans had commented on his team's web page since Monday. He had been **ignoring** his phone, TV and the **Internet** since Monday evening, and wished the event to **quietly** fade out of his **memory**. His team had **happily** gone to an away game on Monday, but met an unexpected and **humiliating** defeat. It ended in a one-to-seven defeat against the **neighboring** city's team, **undoubtedly** thought to be the weakest of the two. The bookies gains on this one must have seemed **apparent** to anyone.

Nobody could **adequately** imagine that outcome: the team had **accumulated** wins and defeated opponents, attacking and defending with the **acquired** ease. It had **operated** as a machine does: it was fast and **accurate**. Magazines had been **admiring** him, speaking about his **inherent** gift as a coach, his **coherent** tactics, and his **ability** to change any team into one of the best **category**. But on Monday those outstanding **capabilities** vanished as if they had not once existed. The team showed a sudden **inability** to attack, **cooperate** and defend. He knew: he had to quit his post **immediately**. No **moderate** steps can be expected in this situation. It was so sad: his job had given him money, fame, and **mobility**.

Upon **entering** into the top-division competition, he hadn't expected to achieve anything. In an off-**camera dialogue** with a talk-show host, he even **openly** admitted it. But now the thought of having to join that **catalogue** of coaches, each one queuing up to find a new coaching position, intimidated him. He expected no **equality** of chances: no famous team was going to invite him now as a coach. No one. That's enough, he thought. He had to face the situation and this **inequality** and pay no attention to **ignorant** fans. Act **independently** of what they might say. The exact moment he decided that mind-**wandering**, sitting and thinking about his devastating situation had no **utility**, somebody knocked at his window.

Appendix B. Noisified version of the story used in the exposure phase

He opened[3] the magazine, **immediately** saw his own name, and began **wondering** how many fans had commented on his team's web page since Monday. He had been **ignoring** his phone, TV and the **Internet** since Monday evening, and wished the event to **quietly** fade out of his **memory**. His team had **happily** gone to an away game on Monday, but met an unexpected and **humiliating** defeat. (pause). It ended in a one-to-seven defeat against the **neighboring** city's team, **undoubtedly** thought to be the weakest of the two. The bookies gains on this one must have seemed **apparent** to anyone.

Nobody could **adequately** imagine that outcome: the team had **accumulated** wins and defeated opponents, attacking and defending with the **acquired** ease. It had **operated** as a machine does: it was fast and **accurate**. Magazines had been **admiring** him, speaking about his **inherent** gift as a coach, his **coherent** tactics, and his **ability** to change any team into one of the best **category**. But on Monday those outstanding **capabilities** vanished as if they had not once existed. The team showed a sudden **inability** to attack, **cooperate** and defend. He knew: he had to quit his post **immediately**. (pause). No **moderate** steps can be expected in this situation. It was so sad: his job had given him money, fame, and **mobility**.

Upon **entering** into the top-division competition, he hadn't expected to achieve anything (pause). In an off-**camera dialogue** with a talk-show host, he even **openly** admitted it. But now the thought of having to join that **catalogue** of coaches, each one queuing up to find a new coaching position, intimidated him. He expected no **equality** of chances: no famous team was going to invite him now as a coach. No one. That's enough, he thought. He had to face the situation and this **inequality** and pay no attention to **ignorant** fans. (pause). Act **independently** of what they might say. The exact moment he decided that mind **wandering**, sitting and thinking about his devastating situation had no **utility**, somebody knocked at his window.

[2] Target words are in bold.
[3] Underlined fragments are masked by noise.

References

Baayen, R.H., Piepenbrock, R., Gulikers, L., 1995. The CELEX lexical database (release 2). Linguistic data consortium.
Ben-David, B.M., Chambers, C.G., Daneman, M., Pichora-Fuller, M.K., Reingold, E.M., Schneider, B.A., 2011. Effects of aging and noise on real-time spoken word recognition: evidence from eye movements. J. Speech Lang. Hear. Res. 54 (1), 243–262.
Boersma, P., Weenink, D., 2009. Praat: doing phonetics by computer (Version 5.1. 05)[Computer program]. Retrieved May 1, 2009.
Broersma, M., Scharenborg, O., 2010. Native and non-native listeners' perception of English consonants in different types of noise. Speech Commun. 52 (11–12), 980–995.
Brouwer, S., Bradlow, A.R., 2011. The influence of noise on phonological competition during spoken word recognition. In: Proceedings of the International Congress of Phonetic Sciences, Vol. 2011. NIH Public Access, pp. 364–367.
Brouwer, S., Bradlow, A.R., 2016. The temporal dynamics of spoken word recognition in adverse listening conditions. J. Psycholinguist. Res. 45 (5), 1151–1160.
Brouwer, S., Mitterer, H., Huettig, F., 2012. Speech reductions change the dynamics of competition during spoken word recognition. Lang. Cogn. Process. 27 (4), 539–571.
Bruggeman, L., Cutler, A., 2020. No L1 privilege in talker adaptation. Bilingualism: Lang. Cogn. 23 (3), 681–693.
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? Attention, Percep. Psychophys. 70 (4), 604–618.
Collins, B., Mees, I.M., 1999. The phonetics of English and Dutch. Brill.
Cutler, A., Burchfield, L., Antoniou, M., 2018. Factors affecting talker adaptation in a second language. In: The 17th Australasian International Conference on Speech Science and Technology. SST 2018.
Cutler, A., McQueen, J.M., Butterfield, S., Norris, D., 2008. Prelexically-driven perceptual retuning of phoneme boudaries. In: The 9th Annual Conference of the International Speech Communication Association. pp. 2056–2056.
Cutler, A., Norris, D., 1988. The role of strong syllables in segmentation for lexical access. J. Exp. Psychol.: Hum. Percep. Perform. 14 (2), 113–121.
Drozdova, P., van Hout, R., Scharenborg, O., 2016. Lexically-guided perceptual learning in non-native listening. Bilingualism: Lang. Cogn. 19 (5), 914–920.
Eisner, F., McQueen, J.M., 2006. Perceptual learning in speech: Stability over time. J. Acoust. Soc. Am. 119 (4), 1950–1953.
Farris-Trimble, A., McMurray, B., Cigrand, N., Tomblin, J.B., 2014. The process of spoken word recognition in the face of signal degradation.. J. Exp. Psychol.: Hum. Percep. Perform. 40 (1), 308–327.
Flege, J.E., 1995. Second language speech learning: Theory, findings, and problems. Speech Percep. Linguist. Exper.: Issues Cross-Lang. Res. 92, 233–277.
Garcia Lecumberri, M.L., Cooke, M., Cutler, A., 2010. Non-native speech perception in adverse conditions: A review. Speech Commun. 52 (11), 864–886.
Hintz, F., Scharenborg, O., 2016. The Effect of Background Noise on the Activation of Phonological and Semantic Information During Spoken-Word Recognition. ISCA, San Francisco, CA, USA.
Jesse, A., McQueen, J.M., 2011. Positional effects in the lexical retuning of speech perception. Psychonomic Bull. Rev. 18 (5), 943–950.
Kawahara, H., Masuda-Katsuse, I., De 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 Commun. 27 (3), 187–207.
Kraljic, T., Brennan, S.E., Samuel, A.G., 2008a. Accommodating variation: Dialects, idiolects, and speech processing. Cognition 107 (1), 54–81.
Kraljic, T., Samuel, A.G., 2011. Perceptual learning evidence for contextually-specific representations. Cognition 121 (3), 459–465.
Kraljic, T., Samuel, A.G., Brennan, S.E., 2008b. First impressions and last resorts: How listeners adjust to speaker variability. Psychol. Sci. 19 (4), 332–338.
Lemhöfer, K., Broersma, M., 2012. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. Behav. Res. Methods 44 (2), 325–343.
Mayo, L.H., Florentine, M., Buus, S., 1997. Age of second-language acquisition and perception of speech in noise. J. Speech Lang. Hear. Res. 40 (3), 686–693.
McQueen, J.M., Cutler, A., Norris, D., 2006. Phonological abstraction in the mental lexicon. Cogn. Sci. 30 (6), 1113–1126.
McQueen, J.M., Huettig, F., 2012. Changing only the probability that spoken words will be distorted changes how they are recognized. J. Acoust. Soc. Am. 131 (1), 509–517.

Mitterer, H., Scharenborg, O., McQueen, J.M., 2013. Phonological abstraction without phonemes in speech perception. Cognition 129 (2), 356–361.
Norris, D., McQueen, J.M., Cutler, A., 1995. Competition and segmentation in spoken-word recognition.. J. Exp. Psychol. Learn. Mem. Cogn. 21 (5), 1209–1228.
Norris, D., McQueen, J.M., Cutler, A., 2003. Perceptual learning in speech. Cogn. Psychol. 47 (2), 204–238.
Powell, M.J., 2009. The BOBYQA Algorithm for Bound Constrained Optimization Without Derivatives. Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, pp. 26–46.
Reinisch, E., Holt, L.L., 2014. Lexically guided phonetic retuning of foreign-accented speech and its generalization.. J. Exp. Psychol.: Hum. Percep. Perform. 40 (2), 539–555.
Reinisch, E., Weber, A., Mitterer, H., 2013. Listeners retune phoneme categories across languages.. J. Exp. Psychol.: Hum. Percep. Perform. 39 (1), 75–86.
Rogers, C.L., Lister, J.J., Febo, D.M., Besing, J.M., Abrams, H.B., 2006. Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. Appl. Psycholinguist. 27 (3), 465–485.
Samuel, A.G., Kraljic, T., 2009. Perceptual learning for speech. Attention Percep. Psychophys. 71 (6), 1207–1218.
Scharenborg, O., Coumans, J.M., van Hout, R., 2018. The effect of background noise on the word activation process in nonnative spoken-word recognition. J. Exp. Psychol. Learn. Mem. Cogn. 44 (2), 233.
Scharenborg, O., Janse, E., 2013. Comparing lexically guided perceptual learning in younger and older listeners. Attention Percep. Psychophys. 75 (3), 525–536.
Scharenborg, O., Koemans, J., Smith, C., Hasegawa-Johnson, M., Federmeier, K.D., 2019. The neural correlates underlying lexically-guided perceptual learning. In: Proceedings of Interspeech.
Scharenborg, O., van Os, M., 2019. Why listening in background noise is harder in a non-native language than in a native language: A review. Speech Commun..
Scharenborg, O., Weber, A., Janse, E., 2015. The role of attentional abilities in lexically guided perceptual learning by older listeners. Attention Percep. Psychophys. 77 (2), 493–507.
Schuhmann, K.S., 2015. Perceptual Learning in Second Language Learners (Ph.D. thesis). The Graduate School, Stony Brook University: Stony Brook, NY..
Scobbie, J.M., Sebregts, K., Stuart-Smith, J., 2009. Dutch rhotic allophony, coda weakening, and the phonetics-phonology interface. QMU Speech Science Research Centre. Working Papers. WP-18.
Van de Velde, H., van Hout, R., 1999. The pronunciation of (r) in Standard Dutch. Linguist. Netherlands 16 (1), 177–188.
Zhang, X., Samuel, A.G., 2014. Perceptual learning of speech under optimal and adverse conditions.. J. Exp. Psychol.: Hum. Percep. Perform. 40 (1), 200–217.