Phoneme Category Retuning in a Non-native Language

Polina Drozdova [1,2], *Roeland van Hout* [1], *Odette Scharenborg* [1,3]

[1] Centre for Language Studies, Radboud University Nijmegen, The Netherlands
[2] IMPRS for Language Sciences, Nijmegen, The Netherlands
[3] 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 studies have demonstrated that native listeners modify their interpretation of a speech sound when a talker produces an ambiguous sound in order to quickly tune into a speaker, but there is hardly any evidence that non-native listeners employ a similar mechanism when encountering ambiguous pronunciations. So far, one study demonstrated this lexically-guided perceptual learning effect for non-natives, using phoneme categories similar in the native language of the listeners and the non-native language of the stimulus materials. The present study investigates the question whether phoneme category retuning is possible in a non-native language for a contrast, /l/-/r/, which is phonetically differently embedded in the native (Dutch) and nonnative (English) languages involved. Listening experiments indeed showed a lexically-guided perceptual learning effect. Assuming that Dutch listeners have different phoneme categories for the native Dutch and non-native English /r/, as marked differences between the languages exist for /r/, these results, for the first time, seem to suggest that listeners are not only able to retune their native phoneme categories but also their non-native phoneme categories to include ambiguous pronunciations.

Index terms: perceptual learning; non-native listening; human word recognition; liquids

1. Introduction

There is now ample evidence that when listening to ambiguous pronunciations of certain sounds, listeners apply their lexical knowledge to interpret what they hear (see [1] for an overview). For instance, when hearing the word *cactu[f/s]*, where the final sound is ambiguous between the sounds /f/ and /s/, native English listeners will interpret the word as *cactus,* since this word is an existing word in English, while *cactuf* is not. This process, termed lexically-guided perceptual learning [2], causes a temporary change in phonemic category representations [3], such that the ambiguous sound is included in the phoneme category. Lexically-guided perceptual learning generalizes to other ambiguous words that have not been heard before [4]. Coming back to our example, listeners who have heard a certain number of words where all instances of /s/ are ambiguous will perceive the ambiguous sound *[f/s]* in the word *li[f/s]e* as an /s/, while listeners exposed to the ambiguous sound *[f/s]* in /f/ contexts would interpret this same input as *life*. This phoneme category retuning is important in everyday listening situations as it allows listeners to quickly adapt to idiosyncratic pronunciations.

Lexically-guided perceptual learning has been extensively demonstrated in native listening for different sound contrasts, such as stops [2,5], liquids [6,7], and vowels [8]. The effect has typically been shown by means of an exposure-test paradigm, in which the listeners are first exposed to an ambiguous input in, e.g., a short story [5] or a lexical decision task [2,9], and then had to perform a test, usually a phonetic categorization task.

While it is now clear that native (L1) listeners employ lexically-guided perceptual learning as an adaptation mechanism to quickly tune into a speaker, research on this process in non-native listening is scarce. Non-native (L2) listeners typically have a smaller vocabulary and phoneme categories in L1 might differ from those in L2. Hence, three important pre-conditions for lexically-guided perceptual learning might be absent: the lexical information necessary for the retuning might not be exploited, an ambiguous sound might not be perceived as a phoneme, and/or the ambiguous sound might not be perceived as ambiguous enough to induce phoneme category retuning. It is thus not evident that non-native listeners are able to benefit from the same process of phoneme category retuning as native listeners during word recognition. To the authors' knowledge, only one study has looked at the process of lexically-guided perceptual learning by non-native listeners. This study demonstrated that lexical knowledge in a non-native language (Dutch) can induce category retuning in a native language (German) [10]. Reinisch et al. wanted to maximize the chance of finding the effect and therefore tested a sound-contrast (/f/ - /s/) that is similar in the native and non-native language of their participants. Since the used categories (/f/ - /s/) were similar across the two languages involved, the authors hypothesized that the listeners shifted their L1 phoneme categories using L2 lexical context. It remains to be seen, however, whether category retuning is possible for phoneme contrasts with larger differences in a native and non-native language.

The present study explores the possibility of lexically-guided perceptual learning by Dutch non-native listeners of English using the contrast: /r/ - /l/. The articulation of /l/ does not differ much between Dutch and English, but marked differences exist in the articulation of the /r/, despite the variability of pronunciation of /r/ in Dutch ([11], p. 179). The variant that approaches the English /r/ best is the bunched approximant, which, if it occurs in Dutch, only occurs in a post-vocalic position. The post-vocalic position was, however, not included in this study because of the non-rhotic property of many English accents. Dutch is a rhotic language, meaning that the /r/ is pronounced in all contexts.

Both PAM-L2 (the L2 version of the Perceptual Assimilation Model) [12] and SLM (the Speech Learning Model) [13,14] posit that L2 language phoneme categories that are similar to L1 phoneme categories are perceptually assimilated to the L1 phoneme category when there is a perfect match between the native and non-native phoneme or, when this match is less perfect, are established as allophones

14 – 18 September 2014, Singapore

of the native phonemes (PAM-L2). Since the pronunciation of /l/ in English and Dutch is similar, it is to be expected that the L2 English /l/ is assimilated into the L1 Dutch phoneme category. Due to the marked differences for /r/ between the two languages, however, Dutch listeners might have language-dependent phoneme categories for /r/. So, if lexically-guided perceptual learning is observed using the /l/-/r/ sound contrast, this may suggest that also L2 phoneme categories can be retuned on the basis of ambiguous L2 input.

2. Method

2.1. Participants

Fifty-two native Dutch speakers (9 males and 43 females, age 18-25; M = 21.2, SD = 2.0) from the Radboud University Nijmegen subject pool took part in the main experiment. All participants had learned English at school and university for at least five years. An additional 15 native Dutch participants, age 18-37 (M = 23.1, SD=4.7, 12 females and 3 males) took part in the pretest of the stimuli (see section 2.3). All participants were paid for their participation.

2.2. Materials

2.2.1. Exposure phase: Target words and short story

Nineteen English words containing one /l/ sound and 19 English words containing one /r/ sound with word frequencies of at least 100 per million were chosen from CELEX [15]. The words had no /r/ or /l/'s in other positions and contained no sibilants (see the first and third column in Table 1 for the list of the words). All /l/'s and /r/'s, apart from one (*Internet*, this was due to the small number of possible /r/-words that would fit into the story), appeared at the onset of the third or fourth syllable to ensure that the same allophonic /l/ and /r/ was used for all target words, as lexically-guided perceptual learning does not transfer to other allophonic variants of the same phoneme [6], at least for native listeners. These 38 target words were then used to compose a story, which contained no other words with an /l/ or /r/ apart from the target words. The final version of the story contained 333 words.

Table 1. *The /l/ and /r/ words used in the short story and their most ambiguous step (see section 2.3).*

/l/ target words	Step	/r/ target words	Step
ability	4	*accurate*	4
accumulated	6	*acquired*	8
adequately	9	*admiring*	6
capabilities	5	*apparent*	4
catalogue	5	*camera*	4
dialogue	7	*category*	5
equality	6	*coherent*	4
happily	4	*cooperate*	4
humiliating	7	*entering*	5
immediately (2 times)	5	*ignorant*	7
inability	5	*ignoring*	4
independently	1	*inherent*	5
inequality	6	*Internet*	5
mobility	5	*memory*	5
openly	3	*moderate*	5
quietly	8	*neighboring*	6
undoubtedly	5	*operated*	7
utility	6	*wandering*	4
		wondering	5

2.2.2. Creating the ambiguous stimuli

Due to coarticulation effects, many acoustic cues due to /r/ and /l/ are non-local [16]. Since in our stimuli the target sounds occurred in the middle of the words – and in different phonetic environments – ambiguous versions of not just the target sounds but of the entire words were created in order to keep this coarticulatory information available to the listener. To that end, two versions of each target word were morphed: one natural version of the target word (e.g., *memory* and *equality)* and another one where the target sound was substituted with its counterpart (e.g., *memoly* and *equarity*).

To ensure that the words were pronounced at a comparable speech rate, intonation, and style as the rest of the story, the target words were recorded as a part of the short story. Three different versions of the story were recorded by a native male speaker of British English in a sound-damped booth using a Sennheiser ME 64 microphone. In one version of the text the speaker substituted all /r/ sounds by an /l/; in the second version all the /l/ sounds were substituted by /r/ sounds; in the third version, the speaker produced all target words in their natural way. All versions of the text were recorded twice.

Following [7,9], to create the ambiguous items, the target words were excised at the positive-going zero crossings from the audio files of the short stories, using Praat [17]. Subsequently, the excised words were zero-padded, so that there was 25 ms of silence at the beginning and at the end of the words. The pitch contours of the two items from each pair (e.g., *memory-memoly*) were equalized and the resulting words were morphed with the STRAIGHT algorithm [18] in MATLAB [19] to create an 11-step continuum from a version of the word where the interpretation of the ambiguous [l/r] sound was /l/ like (step 0) or /r/ like (step 10).

2.2.3. Test phase: Minimal pairs

The stimuli that were used in the test phase of the experiment consisted of two minimal pairs. To avoid bias towards either the /l/ or the /r/, two minimal pairs were chosen with opposite patterns of word frequency (obtained from CELEX [15]): *collect – correct* and *alive – arrive*. For *alive-arrive* the /l/-word is more frequent (*arrive* – 157 per million, *alive* – 1135 per million), whereas for *collect-correct*, the /r/-word is more frequent (*collect* – 117 per million, *correct* – 804 per million). The words were recorded by the same speaker who also produced the short story, and morphed according to the procedure described in the previous sub-section.

2.3. Pretest of stimuli

A pretest was carried out to determine the most ambiguous version of both the target words and the minimal pairs. During the pretest, five steps (steps 1, 3, 5, 7, 9) of each target (38 words) or test word (4 words) were presented to 15 participants in a sound-proof booth using headphones (Sennheiser HD 215 MKII DJ). Participants had to indicate whether they thought they had heard the /l/- or /r/-interpretation of the item by pressing one of two buttons on a button box. The left button corresponded to the item with an /l/ sound, whereas the right button corresponded to the item with an /r/ sound. To help the participant, the /l/-reading of the item was presented in the left-bottom corner of a computer screen while the /r/-reading of the item was shown in the right-bottom corner during auditory presentation of the stimulus. The experiment consisted of two blocks, so that each step of the

continuum was presented twice. In total, participants had to categorize 400 items. There was a self-paced pause after each 50 items. The experiment was implemented using Presentation software [20] and took about 15 minutes to complete.

After the pretest, total proportions of /l/ and /r/ responses to the five steps of each of the continua were calculated. The most ambiguous step of each word individually was that step on the continuum that received approximately 50% of /r/ and 50% of /l/ responses. If the 50% point fell in between two of the tested steps, the step in-between the two tested steps was chosen as the most ambiguous step.

After the ambiguous step of each target word was determined, the 38 most ambiguous versions of the target words were spliced back into the short story, such that one version of the story only contained the target words where all /l/ sounds were ambiguous, and the other version only contained the target words where the /r/ sounds were ambiguous. For the test phase, the most ambiguous step and the two steps directly preceding and following the most ambiguous step were chosen for each of the two minimal pairs. The most ambiguous step for the *alive-arrive* continuum was step 5; hence, steps 3-7 were included in the task. The most ambiguous step for the *collect-correct* continuum was step 4; therefore, steps 2-6 were used.

2.4. Procedure

Participants were tested individually in a sound-proof booth. They were randomly assigned to one of two 'exposure groups', which differed in the version of the text they received. The intensity level of all stimuli (short story and words in the phoneme categorization task) was set at 60 dB SPL and was the same for all participants. The experiment was administered with Presentation software [20], and audio stimuli were presented binaurally through the same headphones as used for the pretest. Participants were comfortably seated in front of a computer screen in a sound-proof booth. They saw an instruction on the computer screen informing them that they would be listening to a short story in English. To start the story, participants had to press a button. Once the story was finished, a message came up on the screen to indicate that the story had finished and that they had to press another button if they were ready for the next task. Subsequently, the test phase started.

In the test phase, participants were presented with 120 items divided over four blocks. Each block consisted of the five steps of each minimal pair each presented three times. The task of the participants was to decide as accurately and fast as possible which word they thought they had heard and press the corresponding button. The left button of the button box corresponded to the /l/ reading of the word (*collect* or *alive)*, while the right button corresponded to the /r/ reading of the word (*correct* or *arrive*). To help the participants, the words were visually presented on a computer screen in the same way as during the pretest. The whole procedure lasted approximately 15 minutes.

3. Results

On the basis of the number of /r/ and /l/ responses in the phonetic categorization task, four participants from the /l/- and one participant from the /r/- group were excluded from further analysis since out of 120 presented items, they categorized 8 or fewer as containing an /l/ sound (mean number of /l/

responses: 44.0 (SD=19.4); the lowest number of /l/ -responses in the remainder of the data: 16).

The data of the remaining 47 participants (24 in the /r/ group and 23 in the /l/ group) were analyzed using generalized linear mixed effect models (e.g., [18]). We started the analysis with the most complex model, i.e., the model containing all predictors and all possible interactions. Subsequently, interactions and predictors that were not significant were one-by-one removed from the model. The results presented here were obtained with the best-fitting model with the exposure condition (exposure to the version of the story where all words with an /l/ or an /r/ sound were manipulated), the step of the /r/ -/l/ continuum (the ambiguous steps for both continua were relabeled so that step 1 is the most /l/ like step, step 3 is the most ambiguous step, and step 5 is the most /r/ like step), and the minimal pair as fixed predictors. Table 2 displays the estimates of the fixed effects and their interactions in the best-fitting model. Step was included as a nominal variable.

Table 2. *Fixed-effect estimates of performance in the phonetic categorization task.*

| Fixed effect	β	SE	$p <$
Intercept	-2.055	0.247	.001
Exposure condition	1.072	0.325	.001
Step 2	1.352	0.199	.001
Step 3	3.351	0.209	.001
Step 4	4.783	0.246	.001
Step 5	5.102	0.281	.001
Minimal pair	-1.024	0.201	.001
Exposure condition × Step 2	-0.700	0.229	.01
Exposure condition × Step 3	-0.876	0.241	.001
Exposure condition × Step 4	-1.632	0.265	.001
Exposure condition × Step 5	-1.541	0.331	.001
Exposure condition × Minimal pair	0.666	0.153	.001
Step 2 × Minimal pair	0.116	0.220	*ns*
Step 3 × Minimal pair	0.070	0.226	*ns*
Step 4 × Minimal pair	-0.343	0.249	*ns*
Step 5 × Minimal pair	1.338	0.323	.001
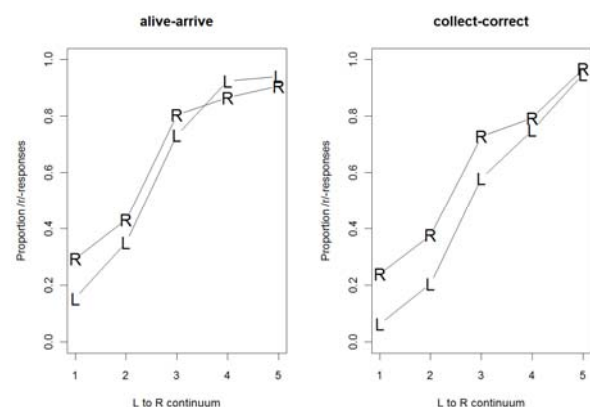
Figure 2: *Proportion of the /r/ responses of the two exposure conditions for the minimal pairs "alive-arrive" (left panel) and "collect-correct" (right panel).*

Figure 2 shows the total proportion of /r/ responses for the five steps of the continua used in the phonetic categorization task, for the two exposure groups and the two minimal pairs separately. The responses of the participants exposed to the version of the story where all /r/ sounds were ambiguous are labeled with "R", the responses of the other group are labeled

"L". The left panel shows the data for the minimal pair *alive-arrive* and the right panel shows the data for the minimal pair *collect-correct*. The difference in proportion of /r/ responses as indicated by the two lines is referred to as the lexically-guided perceptual learning effect.

As Figure 2 shows, participants exposed to the story where all /r/ words were replaced by the ambiguous versions chose the word containing an /r/ sound in the phonetic categorization task more often than the participants exposed to the story where all words with an /l/ sound were replaced by the ambiguous version. This difference in exposure condition was significant (see Table 2): our participants thus showed lexically-guided perceptual learning. The degree of lexically-guided perceptual learning, however, varied depending on the minimal pair (Exposure condition × Minimal pair), as can be seen in Figure 2 when comparing the left and the right panel.

Further, as expected, the factor Step was significant: the number of /r/ responses significantly increased for more /r/-like stimuli (i.e., towards the /r/ end of the /l/-/r/ continuum). This effect was less strong for the exposure group where all /r/ words were ambiguous (Exposure condition × Step interactions). Significantly fewer /r/ responses were given to the items from the *correct-collect* minimal pair than to the items from the *alive-arrive* minimal pair (Minimal pair), although this was not the case for the final step of the continuum (Step 5 × Minimal Pair).

The effect of minimal pair is surprising considering that normally it is assumed that the learning effect is found irrespective of the test stimuli used, and thus asks for further analysis. We conducted a separate analysis for each minimal pair. Although the previous analysis suggested that the lexically-guided perceptual learning effect differed between the two minimal pairs, the lexically-guided perceptual learning effect was observed for both minimal pairs (*correct-collect*: β = 2.247, *SE* = 0.631, *p* < .001; *alive – arrive*: β = 1.124, *SE* = 0.401, *p* < .01). This analysis confirmed that the lexically-guided perceptual learning effect was indeed smaller for the *alive-arrive* minimal pair.

4. General Discussion and Conclusions

In the present study, we investigated whether lexically-guided perceptual learning occurs during non-native listening with a sound contrast that is different in the native and non-native language of the participants. In a two-phase experiment, Dutch native participants were first exposed to one of two versions of a short story in English, where either the words with an /l/ or those with an /r/ sound were manipulated in such a way that the /l/'s and /r/'s were replaced by an ambiguous sound halfway between /l/ and /r/. In a subsequent phonetic categorization task, the participants exposed to the short story where words with an /r/ sound were ambiguous gave significantly more /r/ responses than participants exposed to the short story where the words with an /l/ sound were ambiguous, thus demonstrating a lexically-guided perceptual learning effect. This result adds to a large literature of how listeners flexibly adapt their phoneme categories depending on their listening situation, but extends it in an important way.

Lexically-guided perceptual learning in non-native listening was previously demonstrated by [10] using a sound-contrast similar in the native (L1) and non-native (L2) languages of the participants (/s/-/f/). The authors suggested that since the phoneme representations of /f/ and /s/ are similar

across their L1 and L2 languages, L2 lexical context retuned L1 phoneme categories rather than L2 phoneme categories. The sound contrast /l/-/r/ used in the present study, however, is less similar in our L1 (Dutch) and our L2 (English) languages. Although Dutch /l/ is fairly similar to the English /l/, this is not the case for /r/, for which, due to the differences between English and Dutch, language-specific phoneme categories may have been established by the listeners. Our lexically-guided perceptual learning results thus seem to imply that L2 lexical context retunes L2 phoneme categories of /r/.

The approach we employed in the present study differs from previous work. Firstly, a short story was used as a method of exposure for non-native listening, while previously it was employed only for natives [5]. Secondly, since we wanted to make coarticulatory information of /l/ and /r/ available to the listener, the stimuli were generated by morphing whole words, rather than single phonemes or syllables. This method ensured that our participants were able to use the non-local, coarticulatory information of /l/ and /r/ present in the speech signal during speech perception. The results presented here show that whole word morphs can be used to investigate lexically-guided perceptual learning, making it possible to investigate non-local phoneme contrasts in both native and non-native languages.

Interestingly, the found effect depended on the word-pair used during testing. There was a significant interaction between the exposure condition and the word pair, meaning that the lexically-guided perceptual learning effect was more pronounced for *collect-correct* than for *alive-arrive*. This difference cannot be explained by the frequency of occurrence of the words in the minimal pairs, since for the pair *alive-arrive*, "alive" is more frequent, while for the pair *collect-correct*, "correct" is more frequent, and the word frequencies in both minimal pairs were relatively similar. Acoustic analysis of the test stimuli suggests that this difference between the minimal pairs might be due to the steps of the continuum that were chosen for the phonetic categorization task. Despite the fact that the steps of the continuum used in the experiment were chosen on the basis of a pretest, the first step of the *collect-correct* continuum was closer to the clear version of "collect" than the first step of the continuum *alive-arrive* was to the clear version of "alive". In other words, the first step of the *alive-arrive* continuum was more /r/-like than the first step of the *collect-correct* continuum. This difference could cause the greater number of /r/ responses for the *alive-arrive* continuum for both groups and diminish the effect of the exposure condition for this minimal pair. More research is needed to investigate this result.

To conclude, our results show a lexically-guided perceptual learning effect by Dutch listeners using the English /l/-/r/ contrast. These results, for the first time, seem to suggest that listeners not only can use L2 lexical knowledge to retune their L1 phoneme categories but also their L2 phoneme categories.

5. Acknowledgements

6. References

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

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

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

[5] Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *Journal of the Acoustical Society of America*, 119, 1950–1953.

[6] Mitterer, H., Scharenborg, O., & McQueen, J.M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition.*

[7] Scharenborg, O., Mitterer, H., McQueen, J.M. (2011) Perceptual learning of liquids. *Proceedings of Interspeech,* Florence, Italy.

[8] McQueen, J., & Mitterer, H. (2005). Lexically-driven perceptual adjustments of vowel categories. In V. Hazan & P. Iverson (Eds.), *Proceedings of ISCA Workshop on Plasticity in Speech Perception* (pp. 233–236). Grenoble, France: ISCA.

[9] Scharenborg, O., & Janse, E. (2013). Comparing lexically-guided perceptual learning in younger and older listeners. *Attention, Perception, and Psychophysics.* doi: 10.3758/s13414-013-0422-4.

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

[11] Collins, B., & Mees, I. M. (1996). *The phonetics of English and Dutch*. Leiden, The Netherlands: Brill.

[12] Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam, The Netherlands: John Benjamins.

[13] Flege, J. E. (1992). Speech learning in a second language. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development. Models, research, implications* (pp. 565–604). Timonium, MD: York Press.

[14] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

[15] Baayen, R., Piepenbrock, R., Gulikers, L. (1995) The CELEX lexical database (release 2), PA: Linguistic Data Consortium, University of Pennsylvania.

[16] Heid, S., Hawkins, S. (2000) An acoustical study of long-domain /r/ and /l/ coarticulation, *Proceedings of the 5th Seminar on Speech Production: Models and Data*, Kloster Seeon, Bavaria, Germany, 77-80.

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

[18] 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.

[19] Matlab "R 2013a" (Software). *The MathWorks Inc.*

[20] Presentation "17.0" (Software). *Neurobehavioural Systems Inc.*

[21] 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.