Perceptual learning of liquids

Odette Scharenborg[1], Holger Mitterer[1], James M. McQueen[1,2]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, & Behavioural Science Institute, Radboud University Nijmegen, The Netherlands
{Odette.Scharenborg, Holger.Mitterer}@mpi.nl, J.McQueen@donders.ru.nl

Abstract

Previous research on lexically-guided perceptual learning has focussed on contrasts that differ primarily in local cues, such as plosive and fricative contrasts. The present research had two aims: to investigate whether perceptual learning occurs for a contrast with non-local cues, the /l/-/r/ contrast, and to establish whether STRAIGHT can be used to create ambiguous sounds on an /l/-/r/ continuum. Listening experiments showed lexically-guided learning about the /l/-/r/ contrast. Listeners can thus tune in to unusual speech sounds characterised by non-local cues. Moreover, STRAIGHT can be used to create stimuli for perceptual learning experiments, opening up new research possibilities.

Index Terms: perceptual learning, morphing, liquids, human word recognition, STRAIGHT.

10.21437/Interspeech.2011-69

1. Introduction

The flexibility to adjust to idiosyncratic speech, a form of lexically-guided perceptual learning [1], has been investigated widely (for a tutorial review, see [2]). Perceptual experiments showed that listeners use both lexical and phonotactic knowledge to retune their phonemic categories [1],[3]. For instance, an ambiguous sound between [s] and [f] ([s/f]) will be learned as /s/ if heard in words such as *platypus*, but as /f/ in words such as *giraffe*. This learning generalises [4], so that listeners hear the previously unheard word [naɪ[s/f]], for example, as *nice* or *knife* depending on the preceding exposure condition (*platypu*[s/f] vs. *gira*[s/f]). This adjustment is very fast: exposure to just 10 instances of an idiosyncratic sound results in adaptation to that sound [5]. The adjustment seems to be robust over time [6],[7] and thorough – these idiosyncratic sounds can be treated as if they are 'normal' instances of the phonemic category [8].

So far, this learning effect has been shown for contrasts that differ primarily in local acoustic cues, for instance, differences in voice onset times in stops (/t/ vs. /d/ [5]) and differences in fricative-noise spectra (/s/ vs. /ʃ/ [5],[7]; /s/ vs. /f/ [1],[4],[6],[8]). Due to coarticulation effects, however, many acoustic cues are non-local (e.g., those found in vowels preceding liquids [9],[10],[11]). For instance, [11] investigated coarticulatory "resonance effects" due to /r/ vs. /l/ in syllable onset position in Southern British English. Their analyses showed that anticipatory resonance effects can be seen up to 5 syllables (or 0.5-1s) before the /r/ or /l/.

The aim of the present research was to investigate whether lexically-guided perceptual learning also occurs for a contrast that is distributed in nature (i.e., with non-local cues). We chose the /l/ vs. /r/ contrast in Dutch (implemented as [l] vs. [ɹ] in the Western part of the Netherlands). A subordinate aim was to investigate whether the morphing software STRAIGHT [12] can be used to create ambiguous sounds on a continuum of this non-local contrast.

Our experiment consisted of two parts (following [1]). In the first part, a lexical decision task, listeners were exposed to an ambiguous [l/ɹ] in Dutch words ending in either /r/ or /l/ (the exposure phase). Listeners were divided into two groups with one group being exposed to the ambiguous sound only in /l/-final words and the other group being exposed to the ambiguous sound only in /r/-final words. The ambiguous sound was created by morphing [Cəɹ] and [Cəl] syllables, thus capturing the distributed nature of the contrast. In a subsequent phonetic categorisation task (the test phase), listeners were confronted with a range of ambiguous sounds from the [l]-[ɹ]-continuum and were asked to decide whether the ambiguous sound was /l/ or /r/. If a learning effect occurs, we expect more /r/-responses in the phonetic categorisation task for the group of listeners who were exposed to the ambiguous sound in /r/-final words compared to the group of listeners who were exposed to the ambiguous sound in /l/-final words. This research thus adds to the body of literature on the specificity of lexically-guided perceptual learning by testing for the first time a liquid contrast.

2. Experiment

2.1. Participants

Fifty-two native Dutch speakers with no reported hearing problems, drawn from the MPI for Psycholinguistics subject pool, took part in the experiment. Sixteen participated in the pretest (see Section 2.2). The other 36 (8 male; mean age: 21.2) took part in the main experiment, 18 in the group who heard ambiguous /l/-final words during exposure and 18 in the group who heard ambiguous /r/-final words during exposure (see Section 2.3). All 52 were paid for their participation.

2.2. Materials

For the exposure phase, we selected 200 Dutch words from CELEX [13]. Forty words ended in /l/, and 40 ended in /r/; there were no /l/'s or /r/'s elsewhere in these 80 words. Since the sounds [l] and [ɹ] colour the pronunciation of the preceding vowel, the vowel preceding [l] and [ɹ] was kept constant, such that all words ended in /əl/ or /ər/. The number of syllables was matched between the two sets of critical items (i.e., /l/-final and /r/-final words): There were 25 words with two syllables (e.g., *ezel*, donkey), 10 with three syllables (e.g., *postzegel*, stamp), and five with four syllables (e.g., *sinaasappel*, orange). Word frequency was matched between the two sets (means: 2-syllable /l/-words: 23.9 per million, 3-syllable /l/-words: 11.8 per million, 4-syllable /l/-words: 10.6 per million, 2-syllable /r/-words: 23.0 per million, 3-syllable /r/-words: 12.2 per million, 4-syllable /r/-words: 11.2 per million).

Stress patterns were matched as far as possible. Stress was always on the first syllable for the bisyllabic words (a syllable containing /ə/ can never be stressed in Dutch). For the 3-

Copyright © 2011 ISCA

syllable words, 4/10 /l/-words had stress on the first syllable and 6/10 /r/-words had stress on the first syllable. For the 4-syllable words, 2/5 /l/-words and 3/5 /r/-words had stress on the first syllable, 2/5 /l/-words and 0/5 /r/-words had stress on the second syllable (no /r/-words with second syllable stress existed that fulfilled all criteria), and 1/5 /l/-words and 2/5 /r/-words had stress on the third syllable.

One hundred and twenty additional words were selected as filler words, and 200 filler nonwords were constructed. Both sets of fillers followed the same syllable-length distribution as the critical words (e.g. for the filler words, there were 75 with two syllables, 30 with three syllables and 15 with four syllables). The sounds /l/ and /r/ did not occur in any of these items. The nonwords followed Dutch phonotactic rules and tended to become nonwords (i.e., were no longer consistent with any real Dutch words) before their final phonemes.

All words were produced in isolation by a female native speaker of Dutch (from the Western part of the Netherlands) and digitally recorded in a sound-attenuated booth at 44 kHz. She also recorded the nonwords *kwiptel* and *kwipter* for use in the test phase (see below).

From the natural recordings we created versions of the 80 critical words ending in /l/ and /r/ with ambiguous final sounds. These ambiguous sounds [$^{l}/_{ɹ}$] and the test continuum for the phonetic categorisation task were selected using a phonetic categorisation pretest. The selection of the ambiguous sounds was done separately for each final syllable type present in the full set of 80 critical items for the lexical decision task. All critical items ended in /əl/ or /ər/, but the consonants before /əl/ and /ər/ varied. There were a total of 11 different final /Cə$^{l}/_{ɹ}$/ sequences in the set of 80 words. Note that due to devoicing of fricatives in Dutch, syllables beginning with /s/ and /z/ could be treated as the same sequence, likewise for /f/ and /v/ and for /x/ and /ɣ/. A subset of the bisyllabic words from the main experiment was then selected, one pair of words for each of the 11 sequences. Table 1 lists this subset: pairs of words ending in /l/ and /r/ for each sequence, with their English translations and nonword counterparts.

For each pair of words (e.g., *winkel* and *wekker*), the final syllable was excised using Praat [14]. All excised [l]- and [ɹ]-final syllables were zero-padded at onset and offset with 25 ms of silence to allow valid pitch estimation at the start and the end of the syllable. Subsequently, each syllable received the same stylised pitch contour (based on the naturally occurring pitch contour of the final syllables in the critical items) using Praat [14]. The resulting pairs of syllables were then each morphed to create equally-spaced 11-step continua using STRAIGHT [12] in Matlab. Figure 1 shows the ambiguous syllable [kə$^{l}/_{ɹ}$] (top and third panel). This syllable was step 5 on the morphed continuum between the zero-padded natural versions of the syllables [kəl] (second panel) and [kəɹ] (bottom panel). The ambiguous syllables were then concatenated, using Praat, as final syllables onto the first syllables of the matching /l/-final and /r/-final words. For example, the morphs for /kə$^{l}/_{ɹ}$/ were concatenated with both /wɪŋ/ (yielding *winkel*) and /wɛ/ (yielding *wekker*).

The pretest stimuli were presented in three blocks, each consisting of 132 items, in a newly randomised order in each block. Participants heard in each block six [l]-[ɹ]-continuum steps (steps 1, 3, 4, 6, 7, 9), for each of the 11 syllables. These steps were chosen to sample perception of the entire continuum (excluding the endpoints). Since each morph was concatenated with both an /l/- and an /r/-final word, each morph was heard twice per block, and thus six times in total.

The task for the participants was to indicate by button press as quickly and as accurately as possible whether they heard [l] or [ɹ]. To aid the participants, the [l]-interpretation of the stimulus was shown on the bottom left of the computer screen, and the [ɹ]-interpretation of the stimulus on the bottom right. If the [l]-interpretation was a word, the right option was a nonword, and vice versa. Table 1 shows the word and nonword pairs that were used. Each stimulus was presented over headphones 500 ms after trial onset. Due to an error in the testing software, the pretest for the [fəl]-[fəɹ] morphs had to be done separately. Six subjects each heard 10 repetitions of each [fəl]-[fəɹ] morph. The rest of the experimental set-up was identical to the main pretest.

The total proportions of [ɹ]-responses to each of the tested morphs were calculated, and the most ambiguous morph was determined for each of the 11 syllables. The most ambiguous morph for syllables starting with /k, x, b/ was step 5 (where step 0 is a natural [l] and step 10 a natural [ɹ]); for /m, d, f/ it was step 3; for /t, ŋ, n/ it was step 6; for /p/ it was step 2; and for /z/ it was step 7. However, after six participants had been tested on the lexical decision task (see Section 2.3), the results showed that most of the /l/-words ending in ambiguous [$^{l}/_{ɹ}$] were not recognised as words. We therefore decided to change the ambiguous morphs by selecting the next more [l]-like step: /k, x, b/ = step 4; /m, d, f/ = step 2; /t, ŋ, n/ = step 5; /p/ = step 1; and /z/ = step 6.

The selected morphed syllables were then concatenated as final syllables onto the non-final syllables of the matching /l/-final and /r/-final words, in the same manner as was done to create the stimuli for the pretest. This resulted in 80 stimulus pairs consisting of the same word ending in either a natural [l] or [ɹ] or the selected ambiguous [$^{l}/_{ɹ}$]. These stimuli could then be used in the lexical decision task.

The test stimuli consisted of five versions of *kwipte*$^{l}/_{ɹ}$. These were created by concatenating five different versions of the ambiguous [$^{l}/_{ɹ}$] sound as final syllables onto the first syllable *kwip* (excised from a recording of the nonword *kwipter*). The steps (i.e., steps 2, 4, 5, 6, and 8; where step 5 was judged to be the most ambiguous sound in the pretest) were taken from the [təl]-[təɹ]-continuum created for the pretest. Both the /final-/l/ and final-/r/ reading of the resulting string is a non-word in Dutch.

Table 1. The word and nonword pairs used in the pretest.

/l/-final			/r/-final		
word	English translation	nonword	word	English translation	nonword
stengel	stalk	*stenger*	*honger*	hunger	*hongel*
amandel	almond	*amander*	*zender*	channel	*zendel*
tunnel	tunnel	*tunner*	*doener*	doer	*doenel*
meubel	furniture	*meuber*	*puber*	teenager	*pubel*
tegel	tile	*teger*	*tijger*	tiger	*tijgel*
heuvel	hill	*heuver*	*oever*	shore	*oevel*
winkel	shop	*winker*	*wekker*	alarm-clock	*wekkel*
hemel	heaven	*hemer*	*emmer*	bucket	*emmel*
ezel	donkey	*ezer*	*danser*	dancer	*dansel*
schotel	dish	*schoter*	*veter*	shoelace	*vetel*
stempel	stamp/seal	*stemper*	*kapper*	hairdresser	*kappel*

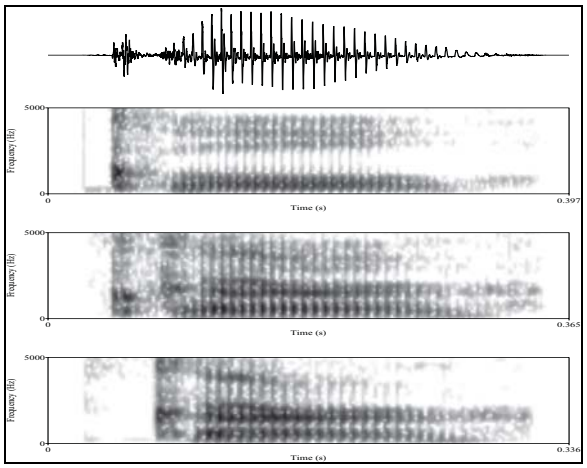

Figure 1. The top panel shows the acoustic signal for the zero-padded ambiguous syllable [kə$^{l}/_{ɹ}$]*; panels 2-4 show the spectrograms of the natural versions of* [kəl]*, the ambiguous* [kə$^{l}/_{ɹ}$]*, and the natural version of* [kəɹ]*, respectively.*

Table 2. Performance on the lexical decision task, mean percentage of 'no' responses and mean RTs for 'yes' responses in the two exposure conditions, for the natural and the ambiguous versions of the /l/- and /r/-final words.

	Natural liquids		Ambiguous liquids	
	/l/-final	/r/-final	/l/-final	/r/-final
Mean % 'no'	1.8	1.0	7.4	14.1
Mean RT 'yes'	205	190	280	248

2.3. Procedure

Two experimental-word lists were created in which the test items appeared in a pseudo-randomised running order, one for each of the two test conditions. The restrictions were that no critical item (i.e., no word ending in [$^{l}/_{ɹ}$]) was allowed to appear in the first six words, and no two critical items could appear within a range of four words. Each word list consisted of 400 words, i.e., the 200 nonwords, 120 filler words, 40 words ending in a clear [l] or [ɹ], and the 40 critical items, i.e., the /r/-final or /l/-final words ending in [$^{l}/_{ɹ}$]. The difference between the two word lists was that one list contained only natural /r/-final words and /l/-final words ending in [$^{l}/_{ɹ}$], the other list contained the natural /l/-final words and the /r/-words ending in [$^{l}/_{ɹ}$].

Half of the participants were presented with the list with ambiguous /l/-words, the other half with the list with ambiguous /r/-words. They were asked to press a button as fast and accurately as possible when they heard a word (left button) or a non-word (right button). They were not informed about the presence of ambiguous sounds. Reaction times (RTs) were measured from item onset and adjusted by subtraction of item durations prior to analysis so as to measure from item offset.

Subsequently, participants were tested using a phonetic categorisation test. They were asked to decide, by button press, as fast and accurately as possible whether the stimulus ended in /l/ or in /r/. The five ambiguous versions of *kwipte*$^{l}/_{ɹ}$ were each presented six times per block, and were newly randomised for each of a total of three blocks (90 items in total). To aid the participants, the /l/-interpretation of the stimulus (*kwiptel*) was shown on the bottom left of the computer screen, and the /r/-interpretation of the stimulus on the bottom right (*kwipter*).

3. Results

3.1. Lexical decision

We first examined performance, in terms of overall acceptance rate or percentage 'yes' responses, during the training phase. We excluded from further analyses subjects who judged only 20 or less of the 40 ambiguous items as words. This resulted in the exclusion of one subject who heard ambiguous /r/-final words and natural /l/-final words (only 5/40 of the [$^{l}/_{ɹ}$]-items were accepted as 'word'). Overall, 96.9% and 97.2% of the filler words were accepted as words by the listeners who were exposed to the ambiguous sounds in /l/-final words and in /r/-final words, respectively.

Table 2 shows the mean percentage of 'no' responses for the natural and the ambiguous versions of the /l/- and /r/-final words for the listeners who were exposed to the ambiguous sound in /r/-final or /l/-final words. 85.9% of the [$^{l}/_{ɹ}$]-final items were accepted as words by the listeners exposed to [$^{l}/_{ɹ}$] in /r/-final words and 92.6% of the items by the listeners exposed to [$^{l}/_{ɹ}$] in /l/-final words. So most listeners accepted the stimuli ending in [$^{l}/_{ɹ}$] as words. Analyses were carried out using generalised linear mixed-effects models, with the word-final liquid and stimulus condition (natural or ambiguous) as fixed predictors and the target word and subject as random predictors. The difference in percentage 'yes' answers between the natural and the ambiguous stimuli was significant ($p < 0.001$). There were more 'no' responses for /r/-final stimuli than for /l/-final stimuli, but not significantly so ($p > 0.5$).

Table 2 also shows the performance in terms of mean RTs for 'yes' responses, measured from stimulus offset, for the natural and the ambiguous versions of the /l/-final and /r/-final words for the listeners who were exposed to the ambiguous sound in /l/-final or /r/-final words. Analyses with linear mixed-effects models, with the word-final liquid and stimulus condition (natural or ambiguous) as fixed predictors and the target word and subject as random predictors, showed that, on average, participants responded to natural versions of [l] and [ɹ] 78 ms faster than to the [$^{l}/_{ɹ}$]-final stimuli. This difference is significant ($p < 0.001$). Moreover, /r/-final words were responded to faster than /l/-final words, but not significantly so ($p > 0.5$; p-values based on Monte Carlo Markov Chain sampling).

The results of the lexical decision task showed that the ambiguous liquid [$^{l}/_{ɹ}$] tended to be interpreted as /l/ by the group of listeners exposed to natural versions of the /r/-final words and [$^{l}/_{ɹ}$] in the normally /l/-final words, whereas listeners who were exposed to [$^{l}/_{ɹ}$] in the context of normally /r/-final words and natural versions of the /l/-final words interpreted [$^{l}/_{ɹ}$] as /r/. Moreover, the results suggest that this tendency is somewhat stronger for the listeners who heard [$^{l}/_{ɹ}$] in the normally /l/-final words.

3.2. Phonetic categorisation

Figure 2 shows the proportion of /l/ and /r/ responses for the five ambiguous stimuli in the phonetic categorisation task. The responses for the listeners who were exposed to [$^{l}/_{ɹ}$] in the normally /r/-final words are indicated with the 'r's. The responses for the listeners who were exposed to [$^{l}/_{ɹ}$] only in the normally /l/-final words are indicated with 'l's. As Figure 2 shows, there is a clear effect of exposure condition on phonemic categorisation. Listeners who were exposed to [$^{l}/_{ɹ}$] in the normally /r/-final words and to natural versions of /l/ (/wɛkə$^{l}/_{ɹ}$/ and *winkel*) were strongly biased to label the sounds on the continuum as /r/, while those listeners who were

exposed to [$^{l}/_{ɹ}$] in the normally /l/-final words and to natural versions of /r/ (/wɪŋkə$^{l}/_{ɹ}$ and *wekker*) were far less likely to do so. This difference was shown to be significant ($p < 0.001$) using a generalised linear mixed-effects model, with the exposure condition (i.e., exposed to the ambiguous /l/-words or exposed to the ambiguous /r/-words), the stimulus step on the continuum (i.e., steps 2, 4, 5, 6, and 8) and test block (i.e., test block 1, 2 or 3) as fixed and subject as random predictor. Additionally, there was an interaction of exposure condition and block ($p < 0.001$).

To understand the nature of this interaction, the difference in /r/-responses for the two exposure conditions was investigated for each of the three test blocks independently. The analysis using generalised linear mixed-effect models showed that although the effect reduces somewhat from block 1 ($b = 4.94$), to block 2 ($b = 4.16$), to block 3 ($b = 3.21$), in the third block, the group with exposure to ambiguous /r/-final words still gave about 25.8% more /r/-responses than the group with exposure to ambiguous /l/-final words ($p < 0.001$).

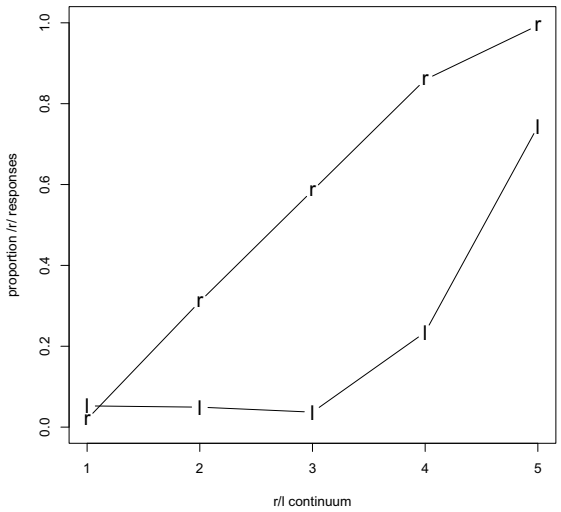

Figure 2. Total proportion of /r/ responses in the two exposure conditions for the five ambiguous test stimuli: r *indicates the listeners who learned to map [$^{l}/_{ɹ}$] onto [ɹ];* l *indicates the listeners who learned to map [$^{l}/_{ɹ}$] onto [l].*

4. Discussion

The results of our experiment show that lexically-guided perceptual learning also occurs for contrasts that differ more than locally, such as the [l]-[ɹ] contrast in Dutch. In the experiment, listeners were exposed to an ambiguous [$^{l}/_{ɹ}$] in Dutch words ending in either /r/ or /l/. The ambiguous sound was created by morphing [Cəɹ] and [Cəl] syllables to capture the distributed nature of the contrast. A subsequent phonetic categorisation test revealed a significant difference in proportion of /r/-responses to an [l]-[ɹ] continuum between the two listener groups that learned to interpret the ambiguous sound as either /r/ or /l/. There were more /r/-responses by the listeners who had been exposed to ambiguous [$^{l}/_{ɹ}$] in the /r/-final words. Listeners thus adapt to contrasts with non-local acoustic cues, that is, those that are more distributed in nature.

Moreover, this adaptation is preserved over time. The perceptual learning effect was still present in the third test block (items 61-90), albeit to a somewhat lesser extent than in the first block. This suggests that exposure to less ambiguous instantiations of [$^{l}/_{ɹ}$] does not immediately undo the adaptation.

These results allow us to investigate whether learning generalises over allophonic differences and over position. If so, exposure should also influence the perception of another implementation of the contrast using a trill for /r/ ([l]-[r]) in both post- and pre-vocalic position (where the approximant is not attested in Dutch).

Finally, the stimuli used in the experiment were created using STRAIGHT. The results presented here show that STRAIGHT can be used to create continua for non-local contrasts that can then be used to investigate lexically-guided perceptual learning. This opens up new research possibilities, by making possible investigations of perceptual learning beyond those on contrasts that differ only in local cues.

5. Acknowledgements

The research by Odette Scharenborg was partly sponsored by the Max Planck International Research Network on Aging. We thank Denise Moerel, Laurence Bruggeman, Lies Cuijpers, Michael Wiechers, Willemijn van den Berg, and Zhou Fang for assistance in preparing and running these experiments and Marijt Witteman for recording the stimuli.

6. References

[1] Norris, D., McQueen, J.M., Cutler, A., "Perceptual learning in speech", Cognitive Psychology, 47(2):204-238, 2003.

[2] Samuel, A.G., Kraljic, T., "Perceptual learning for speech", Attention, Perception, & Psychophysics, 71(6):1207-1218, 2009.

[3] Cutler, A., McQueen, J.M., Butterfield, S., Norris, D., "Prelexically-driven perceptual retuning of phoneme boundaries", Proceedings of Interspeech, 2056-2056, 2008.

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

[5] Kraljic T., Samuel, A.G., "Perceptual adjustments to multiple speakers", Journal of Memory and Language, 56:1-15, 2007.

[6] Eisner F., McQueen, J.M., "Perceptual learning in speech: Stability over time", Journal of the Acoustical Society of America, 119:1950-1953, 2006.

[7] Kraljic T., Samuel, A.G., "Perceptual learning for speech: Is there a return to normal?" Cognitive Psychology, 51:141-178, 2005.

[8] Sjerps, M.J., McQueen, J.M., "The bounds on flexibility in speech perception", Journal of Experimental Psychology: Human Perception and Performance, 36:195-211, 2010.

[9] West, P., "The extent of coarticulation: an acoustic and articulatory study", Proceedings of ICPhS, 1901-1904, 1999.

[10] West, P., "Long-distance coarticulatory effects of British English /l/ and /r/: An EMA, EPG and acoustic study", Proceedings of the 5th Seminar on Speech Production: Models and Data, Kloster Seeon, Bavaria, Germany, 105-108, 2000.

[11] Heid, S., Hawkins, S., "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, 2000.

[12] Kawahara, H., Masuda-Katsuse, I., Cheveigne, A., "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, 1999.

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

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