Perceptual learning of liquids

Odette Scharenborg¹, Holger Mitterer¹, James M. McQueen¹,²

¹Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
²Donders 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.

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 [ʃ] will be learned as /s/ if heard in words such as platypus, but as /ʃ/ in words such as giraffe. This learning generalises [4], so that listeners hear the previously unheard word [nuʃ], for example, as nice or knife depending on the preceding exposure condition (platypus/ʃ/, vs. giraffe/ʃ/). 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. /z/ [5],[7]; /ʃ/ vs. /ʒ/ [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. [ʃ] 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 [ʃ/] 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 [Caʃ] and [Caʃ] 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]-[ʃ]-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 [ʃ] colour the pronunciation of the preceding vowel, the vowel preceding [l] and [ʃ] was kept constant, such that all words ended in /əl/ or /əʃ/. 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 /s/ can never be stressed in Dutch). For the 3-
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)
digitally recorded in a sound-attenuated booth at 44 kHz.
She also recorded the nonwords kwiplet 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/) 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 /l/ or /r/, but the
consonants before /l/ and /r/ varied. There were a total of 11
different final /C/l/ 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 /l/ and /r/ and for /s/ and /z/. 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 [z]-
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 [k3l] (top and third panel). This syllable was step 5
on the morphed continuum between the zero-padded natural
versions of the syllables [kl] (second panel) and [kzl] (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
[k3l] were concatenated with both /w/ (yielding winkel) and /we/
(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]-[z]-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 [z]. To aid the participants, the [l]-interpretation of
the stimulus was shown on the bottom left of the computer
screen, and the [z]-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 non-
word 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 [l]-[fa]-morphs had to be
done separately. Six subjects each heard 10 repetitions of each
[fa]-[fa] morph. The rest of the experimental set-up was
identical to the main pretest.

The total proportions of [l]-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 1 a natural [z]); for /m, d, t/ it
was step 3; for /l, n, p, m/ 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]
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, t/ = step 2; /l, n, p, m/ = 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 [z] or the selected ambiguous [l] or [z]). These stimuli could then
be used in the lexical decision task.

The test stimuli consisted of five versions of kwiplet//l.
These were created by concatenating five different versions of
the ambiguous [l] 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 [l]-[fa]-continuum created for the
pretest. Both the /final-/l- and final-/r- reading of the resulting
string is a non-word in Dutch.

<table>
<thead>
<tr>
<th>Word</th>
<th>English translation</th>
<th>Nonword</th>
<th>Word</th>
<th>English translation</th>
<th>Nonword</th>
</tr>
</thead>
<tbody>
<tr>
<td>stengel</td>
<td>stalk</td>
<td>stenger</td>
<td>hunger</td>
<td>hunga</td>
<td>hungel</td>
</tr>
<tr>
<td>amandel</td>
<td>almond</td>
<td>amander</td>
<td>zender</td>
<td>channel</td>
<td>zendel</td>
</tr>
<tr>
<td>tunnel</td>
<td>tunnel</td>
<td>doener</td>
<td>doer</td>
<td>doenel</td>
<td></td>
</tr>
<tr>
<td>meubel</td>
<td>furniture</td>
<td>membar</td>
<td>paber</td>
<td>pabel</td>
<td></td>
</tr>
<tr>
<td>tegel</td>
<td>tile</td>
<td>tiger</td>
<td>tiger</td>
<td>tiget</td>
<td></td>
</tr>
<tr>
<td>heuvel</td>
<td>hill</td>
<td>heaver</td>
<td>oever</td>
<td>oever</td>
<td></td>
</tr>
<tr>
<td>winkel</td>
<td>winker</td>
<td>wekker</td>
<td>alarm-clock</td>
<td>wekker</td>
<td></td>
</tr>
<tr>
<td>homel</td>
<td>heaven</td>
<td>hemaner</td>
<td>emmer</td>
<td>bucket</td>
<td>emnel</td>
</tr>
<tr>
<td>ezol</td>
<td>donkey</td>
<td>ezer</td>
<td>danser</td>
<td>dansel</td>
<td></td>
</tr>
<tr>
<td>schotel</td>
<td>dish</td>
<td>schoter</td>
<td>voter</td>
<td>shoelace</td>
<td>vetel</td>
</tr>
<tr>
<td>stempel</td>
<td>stamp/seal</td>
<td>stempor</td>
<td>kapper</td>
<td>hairdresser</td>
<td>kappel</td>
</tr>
</tbody>
</table>

Table 1. The word and nonword pairs used in the pretest.
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 \([l//g1867]\)) 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 \([//g1867]\), and the natural version of \([ka2]\), respectively.

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 task. 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 kwiptel/ 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).

### 2.3. Procedure

### 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/-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/-final items were accepted as words by the listeners exposed to /l/ in /r/-final words and 92.6% of the items by the listeners exposed to /l/ in /l/-final words. So most listeners accepted the stimuli ending in /l/ 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/- 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 /r/ 78 ms faster than to the /l/-items. This difference is significant (p < 0.001). Moreover, /l/-final words were responded to faster than /r/-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/ tended to be interpreted as /l/ by the group of listeners exposed to natural versions of the /l/ and /r/ words in the normally /l/-final words, whereas listeners who were exposed to /l/ in the context of normally /r/-final words and natural versions of the /l/-final words interpreted /l/ as /r/. Moreover, the results suggest that this tendency is somewhat stronger for the listeners who heard /l/ 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/ in the normally /r/-final words are indicated with the ‘r’s. The responses for the listeners who were exposed to /l/ 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/ in the normally /r/-final words and to natural versions of /l/ (\(w\)e\(k\)s\(l\)/ and \(w\)inkel) were strongly biased to label the sounds on the continuum as /r/, while those listeners who were
exposed to [l/]\(\text{\textit{w}}\)/ in the normally /l/-final words and to natural versions of /r/ (/\text{\textit{w}}/\text{\textit{k}}/\text{\textit{w}}/ and \text{\textit{w}}kek) 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 (\(p < 0.001\)).

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