CrossMark

Bilingualism: Language and Cognition 20 (1), 2017, 162–183 © Cambridge University Press 2015 doi:10.1017/S1366728915000620

# Phonological processing in late second language learners: The effects of proficiency and task*

ERIN JACQUELYN WHITE
*The Hospital for Sick Children; Toronto, Ontario*
*Centre for Research on Brain, Language and Music (CRBLM); Montreal, Quebec*
DEBRA TITONE
*Centre for Research on Brain, Language and Music (CRBLM); Montreal, Quebec*
*Department of Psychology; McGill University*
FRED GENESEE
*Centre for Research on Brain, Language and Music (CRBLM); Montreal, Quebec*
*Department of Psychology; McGill University*
KARSTEN STEINHAUER
*Centre for Research on Brain, Language and Music (CRBLM); Montreal, Quebec*
*School of Communication Sciences and Disorders; McGill University*

(Received: October 3, 2014; final revision received: August 1, 2015; accepted: August 10, 2015; first published online 7 September 2015)

*Using event related brain potentials (ERPs), we examined the neurocognitive basis of phonological discrimination of phoneme /h/ in native English speakers and Francophone late second language (L2) learners, as a function of L2 proficiency and stimulus/task demands. In Experiment 1, native and non-native (L2 only) phonological contrasts were presented as syllables during a task that directed attention to phonological form. Phonological categorization was assessed with MMN, N2b and P3b effects. In Experiment 2, the same contrasts were presented as words/ pseudowords during a task that directed attention to semantics. Phonological discrimination was assessed with N400 pseudoword effects. High proficiency L2 learners displayed similar ERPs as native speakers in both experiments; low proficiency L2 learners showed discrimination of non-native contrasts in Experiment 1 (directed attention task) only. Thus, L2 phonological discrimination by late learners may depend on stimulus/task factors and occurs in a wider range of contexts as L2 proficiency improves.*

Keywords: Second language acquisition, L2, bilingualism, phonological discrimination, attention, automatic processing, event-related brain potentials (ERPs), Mismatch Negativity (MMN), N2b, P300, P3b, N400, pseudoword

## Phonological processing in late second language learners: The effects of proficiency and task

The effect of age of acquisition on language learning is a controversial issue in second language (L2) acquisition research. According to the Critical Period Hypothesis, L2 attainment is restricted by sensitive or critical period(s) in neurocognitive development (Johnson & Newport, 1989; Lenneberg, 1967). While different sub-systems of language may display different sensitive periods (e.g., that for phonology may occur earlier in development than for morphosyntax; Long, 1990), in both cases L2 attainment is thought to be limited when learning begins after the period has closed. This is in contrast to meaning-based aspects of language, for which age of acquisition appears to play a less critical role (Newport, Bavelier & Neville, 2001). Consequently, L2 phonological and morphosyntactic processing, but not semantic processing, is thought to recruit different neural systems in late (i.e., post-sensitive period) L2 learners compared to native speakers and early L2 learners (for a discussion see White, Hakuta, Williams & Moreno, 2013).

However, recent neurocognitive studies indicate that, at least for some aspects of L2 morphosyntax, processing is affected primarily by L2 proficiency level rather than age of L2 acquisition (Morgan-Short, Steinhauer, Sanz & Ullman, 2012; Rossi, Gugler, Friederici & Hahne, 2006;

* We would like to thank Masha Westerlund, Brooke Levis, Julie Mercier, Heather Goad and Jennifer Goudey for their help with various aspects of this project. This work was supported by grants awarded from the Social Sciences and Humanities Research Council (SSHRC) to Fred Genesee, the Natural Sciences and Engineering Research Council of Canada (NSERC) to Karsten Steinhauer (# RGPGP 312835 and # RGPIN 402678-11), Debra Titone (CRC/CFI project # 205844; NSERC # 204609) and Erin White (Postgraduate Scholarship), by the Canada Foundation for Innovation to Karsten Steinhauer (CRC/CFI; project # 201876) and as well as the Center for Research on Brain, Language and Music (CRBLM). This study was conducted at McGill University.

Address for correspondence:
Erin Jacquelyn White, PhD, Postdoctoral Research Fellow, Neurosciences and Mental Health, The Hospital for Sick Children
*erin.white@sickkids.ca*

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

Steinhauer, White & Drury, 2009; White, Genesee & Steinhauer, 2012). These studies report neurocognitive processing profiles in highly proficient late L2 learners that are like those of native speakers, including automatic processes thought to be particularly difficult for late L2 learners (Ullman, 2001). Low proficiency L2 learners, in contrast, engage qualitatively different neurocognitive systems and recruit automatic processes to a lesser extent. Such findings challenge the notion that maturational changes in the brain place unsurpassable limits on L2 attainment for late learners (although see Pakulak and Neville, 2011 for evidence that age of acquisition, independent of proficiency, may impact auditory phrase structure processing in some cases). In the present study, we examined neurocognitive indices of phonological processing in late L2 learners with relatively high or low proficiency, during relatively easy and difficult listening conditions. Our goal was to elucidate the neurocognitive mechanisms implicated in late L2 phonological processing, and the circumstances under which they are engaged. Our study focuses on Francophone learners of English who often show persistent problems in the production of the non-native phoneme /h/ (e.g., *'oliday* for *holiday; heat* for *eat*) even after decades of exposure to English (Janda & Auger, 1992; John, 2006).

Age of acquisition may impact L2 phonological processing via adults' first language (L1) phonological system. Once the L1 phonological system is optimized for differentiating meaningful L1 sounds, discriminating L2 sounds that are not used contrastively in the L1 (non-native phonological contrasts)[1] may be difficult because they are assimilated into a single L1 category, or because features that are not phonologically relevant in the L1 are filtered out (Best & Tyler, 2007; Flege, 1995). The question is whether late L2 learners, with an already functional L1 phonological system, can establish new L2 phonological category representations and discriminate non-native L2 contrasts using similar processes as native speakers. Evidence for this is mixed. On the one hand, laboratory training studies show lasting improvements in adults' non-native speech perception (Logan & Pruitt, 1995), and highly proficient late L2 learners are better at discriminating L2 phonological contrasts than less proficient learners (Tees & Werker, 1984). However, other studies report that, in contrast to discrimination in native speakers, successful L2 discrimination depends on characteristics of the stimulus or task context, suggesting the use of different underlying processing mechanisms. For example, Yoshida (2004) found that while late L2 learners discriminated non-native phonological contrasts when listening to vowels during a discrimination task, the same learners failed to discriminate the same contrasts when listening to minimal pairs during a lexical decision task. This difference was not observed for contrasts phonemic in the L1 and L2, nor for native speakers. Thus, L2 discrimination success may depend on stimulus and task characteristics, in addition to L1-L2 similarities and L2 proficiency.

The Automatic Selective Perception (ASP) model (Strange, 2011) accounts for stimulus/task effects by discussing the role of attention in L2 speech perception. According to this model, L1 processing occurs automatically by recruiting highly over-learned 'Selective Perception Routines' that are attuned to acoustic-phonetic patterns that reliably differentiate L1 phonological categories. L2 phonological processing, in contrast, is less automatic and requires more attentional resources. Consequently, while L2 discrimination may be successful in situations that place phonological information within the focus of attention (e.g., simple discrimination tasks involving syllables), it can fail when processing more complex stimuli and when listeners are attending to other information, such as word meaning.

The ASP model predicts that late L2 learners may never develop fully automatic L2 Selective Processing Routines that can be employed in challenging listening conditions (Strange & Shafer, 2008). However, most research on this model has involved relatively low proficiency L2 learners, particularly with respect to pronunciation skills (Strange, 2011). This is surprising given that proficient L2 speakers (i.e., with little detectable foreign accent display more native-like patterns of speech perception compared to those with low proficiency (Flege & Schmidt, 1995). Moreover, neurocognitive studies of L2 morphosyntactic processing suggest that it is precisely automatic aspects of processing that become available with increasing proficiency (Steinhauer, 2014; Steinhauer et al., 2009). It is quite possible that as L2 learners become more proficient, L2 phonological processing becomes increasingly automatic, requiring fewer attentional resources, and can occur successfully in more contexts.

Here we used event-related brain potentials (ERPs) to compare neurocognitive indices of phonological processing by native English speakers and late L2 learners of English with relatively high or low English oral proficiency in two experimental conditions. In Experiment 1, we examined phonological processing in a relatively easy context: phonological contrasts are presented as individual syllables in a categorization task that directs attention to phonological properties (e.g., push a button when you hear a speech sound that is different from the rest). Phonological categorization was measured by MMN, N2b and P3b effects. In Experiment 2, the same contrasts were presented in a context that, according to

[1] We use the term 'phonological contrast' to indicate speech information that is 'relevant to minimal lexical differences in a given language' (Best & Tyler, 2007; p. 25). By using this term we are neutral as to whether a speech contrast is phonemic or phonetic for a particular listener.

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

Table 1. *Example Stimuli used in Experiment 1 and 2*

| | Non-native Phonetic Contrast | | | Native Phonetic Contrast | | |
|---|---|---|---|---|---|---|
| | /h/-Ø | | | /h/-Ø | | |
| | Standard | Deviant | Example | Standard | Deviant | Example |
| **Experiment 1** | Ha | A | $ha_1$ $ha_1$ $ha_1$ **$ha_2$ $a_1$** $ha_2$ | FA | A | $fa_1$ $fa_1$ $fa_1$ **$fa_2$ $a_1$** $fa_2$ |
| | A | Ha | $a_1$ $a_2$ $a_1$ $a_1$ **$a_2$ $ha_1$** $a_2$ | A | FA | $a_2$ $a_2$ $a_1$ $a_1$ **$a_2$ $fa_1$** $a_2$ |
| | **Word** | **Pseudoword** | **Example** | **Word** | **Pseudoword** | **Example** |
| Experiment 2 | H - initial | No H | Happy vs. 'Appy | F - initial | No F | Father vs. 'Ather |
| | No H | H - initial | English vs. H-english | No F | F - initial | Answer vs. F-answer |

In Experiment 1, non-native (/h/-Ø) and native (/f/-Ø) contrasts were presented as syllables in a categorical odd-ball paradigm (two tokens of each sound were presented as standards and deviants). Participants' task was to listen to the sounds and to identify a change in stimulus presentation. Experiment 1 represents a relatively simple stimulus context and a task that directs participants' attention to phonetic form.
In Experiment 2 the same contrasts were presented as words and derived pseudowords. The pseudowords were created by adding or removing a /h/ or /f/ from the start of a vowel initial words. Participants' task was a semantic categorization task (i.e., press a button when you hear a fruit or vegetable word) and were not told about the contrasts or pseudowords before testing began. Experiment 2 represents a relatively more difficult stimulus context and a task that directs participants' attention to semantic analysis, rather than phonetics.

the ASP model, should prevent successful discrimination: contrasts are presented as words (e.g., happy, English) and pseudowords (e.g., 'appy', 'h-english') during a semantic categorization task (e.g., push a button when you hear a fruit or vegetable word). Importantly, participants were not told about the contrasts or the pseudowords. Here, we used the N400 pseudoword effect to index phonological discrimination in a lexical context that oriented attention to word meaning rather than phonological form (see Table 1). By comparing results for the same participants and contrasts, we could assess the extent to which L2 phonological processing depended on L2 proficiency level and stimulus/task context.

In both experiments, we examined processing of an English contrast that is notoriously difficult for French speakers to both perceive (LaCharité & Prévost, 1999) and produce (Janda & Auger, 1992; John, 2006): /h/-Ø. Discriminating /h/ from silence may be difficult for French L2 speakers of English because /h/ does not exist in spoken French, either as a phoneme or allophonic variant of other speech sounds. This, together with its low perceptual salience, is thought to make it difficult for these L2 learners to form clear phonological representations for /h/ (Mah, 2011; Mah, Goad & Steinhauer, 2006; 2015). We explored how stimulus/task factors influence the processing of this difficult-to-perceive English-specific phoneme compared to one that exists in both languages (/f/), as an example of how context may influence L2 phonological processing more generally.

## Experiment 1

Most ERP research into the neurocognitive basis of L2 phonological processing examines the discrimination of phonological contrasts presented as syllables. This work has identified the Mismatch Negativity (MMN) as a sensitive measure of automatic and language-specific phonemic discrimination (Näätänen, Paavilainen, Rinne & Alho, 2007). The MMN is a fronto-central negativity, typically elicited by an infrequent 'deviant' sound within a sequence of repeating 'standard' sounds or sound patterns (the auditory odd-ball paradigm). Elicited 100–250 ms after the onset of the deviant, it reflects an early stage in detecting a change in auditory presentation. Two properties of the MMN are relevant here. First, the MMN elicited by phonological contrasts is enhanced when the contrasting sounds are highly familiar to the listener and correspond to members of distinct L1 phonemic categories (e.g., Kazanina, Phillips & Idsardi, 2006; Näätänen, Lehtokoski, Lennes, Cheour, Huotilainen, Iivonen, Vainio, Alku, Iimoniemi, Luuk, Allik, Sinkkonen & Alho, 1997; see however, Sharma, Kraus, Carrell & Nicol, 1993). Second, it can be elicited when participants are not attending to the stimuli and, therefore, is thought to reflect automatic ('pre-attentive') stages of sound discrimination (Näätänen, Paavilainen, Tiitinen, Jiang & Alho, 1993), although attention can influence its amplitude, particularly for difficult contrasts (Sussman, 2007).

There is some evidence that L2 phonological discrimination elicits a MMN response similar to that of native speakers, but the evidence is mixed. This discrepancy might reflect differences in participants' L2 proficiency or the complexity of MMN paradigms used. For example, using a typical odd-ball paradigm (one vowel is the standard stimulus, another is the deviant) Winkler, Kujala, Tiitinen, Sivonen, Alku, Lehtokoski, Czigler, Csépe, Ilmonikemi & Näätänen (1999) found that fluent L2 speakers, but not naïve listeners, exhibited similar MMNs as native speakers when presented with a L2 vowel contrast. On the other hand, using a more

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

difficult categorical odd-ball paradigm[2] to investigate processing of the English /h/-Ø contrast (multiple tokens of the syllables 'hum' and 'um' serve as both standards and deviants), Mah et al., (2015; 2006) found that whereas native English speakers exhibited a clear MMN, French-L2 learners of English, albeit of variable L2 proficiency, did not.

In the present study, we examined the role of L2 proficiency using a similar paradigm as Mah, with the addition that participants performed a phonological categorization task. It may be that the L2 learners reported by Mah et al., (2015) did not successfully discriminate the English /h/-Ø contrast because it was presented in a passive/incidental listening context (participants' attention was drawn to a silent movie and no discrimination task was required). Conversely, in the present study, the phonological categorization task required participants to attend to the phonological features that differentiated the contrasting phonetic segments. Thus, we could examine the extent to which late L2 learners engage attention-based processes to categorize L2 phonological segments, as predicted by the ASP model. Attentional stages of phonological discrimination are associated with two ERP components, N2b and P3b, which are most prominent when participants attend to stimuli (Näätänen et al., 1993). The N2b is a central-posterior negativity and the P3b is a posterior positivity; they are typically found 200–300 ms and 300–500 ms after stimulus onset, respectively. The N2b is thought to reflect an attention orienting response (Näätänen & Gaillard, 1983), while the P3b is thought to reflect the outcome of decision making after the presentation of a task-relevant stimulus (Nieuwenhuis, Aston-Jones & Cohen, 2005).

Following the ASP model, we predicted that L2 speakers would engage attention-based processing of the English-specific /h/-Ø contrast, as evidenced by significant N2b/P3b responses. Of interest is whether they would also engage earlier stages of L2 phonological categorization (indexed by a MMN) and whether this would be more evident in highly proficient learners.

## *Method*

### *Participants*

Fifteen monolingual English speakers (19–35 years old, M = 22.6, 9 female) and 27 native French speakers who spoke English as a L2 (20–35 years old, M = 26.4, 15 female) participated. An additional 12 participants were tested, but were excluded because of technical problems during data acquisition (n=2), or excessive EEG artifacts in Experiment 1 or 2 that contaminated more than half of the trials in one or more conditions (e.g., eye movements; n=10). All were living in Montreal, a French–English bilingual city, at the time of testing. Participants gave written informed consent, were paid for their participation, and reported no history of hearing, language, speech or neurological disorders. All were right-handed (assessed via self-report and the Edinburgh handedness inventory; Oldfield, 1971) and in general reported comparable educational backgrounds (i.e., all but two had some post-secondary education).

### *L2 Proficiency*

L2 proficiency was based on English pronunciation as rated by two blind, independent native English speakers (see also Piske, MacKay & Flege, 2001). Following the EEG session, participants spoke freely on a chosen topic (e.g., "what makes a person healthy and happy?") while being recorded; an approximately 30-second speech sample was selected for assessment. Raters were asked to independently judge samples on 6 dimensions: overall pronunciation, pronunciation of /h/-initial words, vocabulary, grammar, fluency, and overall nativeness, using a 13-point scale (0= not at all like a native English speaker, 13= definitely a native English speaker). Bivariate correlations between the raters' six judgements ranged from 0.74 to 0.91 ($p < .001$), indicating similar proficiency ratings.

The L2 learners were divided into two approximately equal groups based on their 'overall pronunciation' ratings. The high proficiency group had ratings of 9/13 or higher by both raters (n=12; 20–35 years old, mean age= 27.3 years; 5 female; mean pronunciation rating: 10.6), and the low proficiency group had ratings of 8 or lower by at least one judge (n=15; 21–34 years old, mean age= 25.7 years; 10 female; mean pronunciation rating: 6.6). A MANOVA conducted on the groups' six ratings revealed a significant main effect of group [$F(6,20) = 9.88$, $p < .001$; Wilk's lambda = 0.25], indicating higher oral proficiency for the high proficiency group. Univariate $F$ tests showed significant group differences on all dimensions of proficiency (all $ps < .001$; see Table 2). The high proficiency group also outperformed the low proficiency group on a cloze test of English proficiency [$t(18.03) = 2.69$, $p < .05$; equal variances not assumed], and mean self-reported proficiency was also higher [$t(18.15) = 3.75$, $p < .005$; equal variances not assumed], together suggesting the L2 learners were grouped appropriately. See Table 2.

### *English exposure*

Age of L2 Acquisition (AoA) was defined as participants' first significant and intense exposure to English with

[2] The categorical oddball paradigm (see also Hisagi, Shafer, Strange & Sussman, 2010)) might be difficult for L2 learners because listeners must simultaneously ignore irrelevant acoustic differences between members of the same phonetic category, and respond only to phonologically-relevant differences between categories (Strange & Shafer, 2008).

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

Table 2. *Participant Information for the High and Low Proficiency L2 Speakers*

| | High Proficiency n=12 | | Low Proficiency n=15 | |
|---|---|---|---|---|
| | M | SD | M | SD |
| Age at Time of Testing[a] | 27.3 | 4.8 | 25.7 | 3.0 |
| Age of Acquisition in years (AoA) | 19.0 | 3.7 | 18.9 | 3.2 |
| Years of English use | 8.3 | 4.3 | 6.8 | 5.0 |
| Childhood English exposure (% of reported daily language use) | | | | |
| Age 0–4 home | 2.5 | 5.8 | 0.1 | 0.5 |
| Age 0–4 school | 0.8 | 2.9 | 0.0 | 0.0 |
| Age 5–11 home | 2.9 | 6.2 | 2.2 | 5.6 |
| Age 5–11 school | 5.4 | 5.0 | 4.2 | 4.6 |
| Age 12–14 home | 4.6 | 6.2 | 2.3 | 5.6 |
| Age 12–14 school | 10.4 | 5.4 | 12.0 | 5.9 |
| Age 15–18 home | 15.4 | 17.8 | 9.7 | 25.9 |
| Age 15–18 school | 29.2 | 22.7 | 15.8 | 9.3 |
| Current daily English use (%) with: | | | | |
| Friends | 34.6 | 10.8 | 15.1 | 21.7 |
| Family | 7.1 | 16.0 | 6.4 | 21.8 |
| School/Work | 51.7 | 30.0 | 28.1 | 23.6 |
| Spare Time | 44.6 | 21.1 | 33.3 | 30.3 |
| TV/Radio | 62.9 | 25.4 | 43.3 | 35.5 |
| Reading | 62.5 | 22.0 | 25.0 | 26.8 |
| Overall/Total** | **44.6** | 7.9 | **24.3** | 18.7 |
| Self Rate English abilities (7-point scale) | | | | |
| Pronunciation | 5.3 | 0.8 | 3.5 | 1.5 |
| Fluency | 5.5 | 0.9 | 3.9 | 1.8 |
| Listening | 6.2 | 0.6 | 4.7 | 1.4 |
| Reading | 6.5 | 0.5 | 5.1 | 1.6 |
| Vocabulary | 5.6 | 0.8 | 3.7 | 1.7 |
| Grammar | 5.4 | 0.7 | 3.9 | 1.7 |
| Overall/Total** | **5.8** | 0.5 | **4.1** | 1.5 |
| Cloze Test of English Proficiency (%)* | **89.1** | 8.3 | **71.4** | 23.8 |
| Average Speech Sample Ratings (13 point scale) | | | | |
| General Pronunciation | 10.6 | 0.7 | 6.6 | 1.7 |
| Pronunciation H words | 11.9 | 1.3 | 7.8 | 3.2 |
| Fluency | 11.7 | 1.3 | 8.2 | 2.5 |
| Vocabulary | 11.7 | 0.5 | 8.1 | 2.1 |
| Grammar | 11.8 | 0.7 | 7.6 | 2.5 |
| Overall Impression | 11.0 | 0.8 | 6.7 | 1.9 |
| Overall/Total*** | **11.5** | 0.6 | **7.5** | 1.8 |

[a]The Native speakers were younger than both groups of L2 learners, who did not differ from each other [$M = 22.6$ years, $SD = 4.3$; $F(1,2)$ 4.90, $p < .05$]. AoA is defined as age of first regular and intensive L2 exposure (White & Genesee, 1996). Years of English exposure is defined as current age minus AoA.
*$p < .05$; ** $p < .005$; *** $p < .001$

native-speakers (White & Genesee, 1996). Several inclusionary criteria ensured participants were late learners: (1) English acquired as a L2 after the age of 14 (i.e., well after the proposed critical period for phonology has closed, which has been placed during infancy by Kuhl, 2010; others place the close of a general critical period for L2 learning at age 12, see Birdsong, 2006 for a discussion); (2) very limited exposure to English or other languages reported at home or school before this age. Mean AoA for the high and low proficiency groups

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

were 19.0 and 18.9 years, respectively; a difference which was not significant [$t(25) = .05$, $p > .10$]. Life time English exposure was also assessed with a questionnaire in which participants reported English use at home and school (percentage of total language use) between the ages of 0–4, 5–11, 12–14, 15–16, 17–18, and 19+. To examine if groups differed in early English exposure, a repeated measures ANOVA was run using group as a between subjects variable; age range (0–4, 5–11, 12–14 and 15–16) and location (home, school) were within subjects variables. This revealed no significant main effect or interaction with the factor group ($p$s > .10), indicating a similar and limited amount of early English exposure for both groups as children and adolescents (see Table 2).

The questionnaire also asked about current English exposure as a percentage of daily language use in six contexts: with friends, family, at work/school, during their spare time, watching TV/listening to the radio, and reading for pleasure (see Table 2). A MANOVA revealed a significant effect of group [$F(6,19) = 9.88$, $p < .05$; Wilk's lambda = 3.39]: the high proficiency group used English significantly more than the low proficiency group in multiple aspects of their adult lives. Univariate $F$ tests showed significantly greater English use by the high proficiency group with friends ($p < .005$), at work/school ($p < .05$), and when reading ($p < .005$).

### *Stimuli*

The stimuli consisted of two tokens each of three different speech sounds, 'ha', 'a' and 'fa', recorded by a native English speaker (Table 1). One 'a' token was longer in duration (384 ms) and roughly matched the overall length of one of the 'ha' (402 ms) and 'fa' (395 ms) tokens (the second 'ha' and 'fa' tokens were 413 ms and 429 ms, respectively); the other 'a' token was shorter (332 ms) and comparable to the vowel length of the second 'ha' (331 ms) and 'fa' tokens (341 ms; the vowel length of the first 'ha' and 'fa' tokens were 349 ms and 354 ms, respectively). Matching in this way decreased the chance that the L2 speakers could use duration or vowel length as cues for detecting the /h/-Ø deviants, and that any observed ERP effects more likely reflected detection of phonologically-relevant acoustic differences (Kazanina et al., 2006).

The stimuli were presented in 4 five-minute blocks so that the 'ha', 'fa' and 'a' sounds served as both standards and deviants: 'ha' deviant with 'a', 'a' deviant with 'ha', 'fa' deviant with 'a', and 'a' deviant with 'fa'; block order was counter-balanced across participants. In each block, 50 'deviant' stimuli were presented with 235 standard stimuli (82.5 to 17.5 % standard-to-deviant ratio) in a pseudo-randomized order, such that each deviant was preceded by a minimum of 3 standards. Both tokens of each sound were used equally and randomly within a block.

### *Procedure*

Participants sat in a sound-attenuated room, approximately 70 cm in front of a computer screen. They were told they would hear a series of speech sounds, many of which were the same, but some of which were different. They were asked to listen carefully and to click a mouse button upon hearing a speech sound they judged to be different from the others. No explicit instructions were given about the /h/-Ø or /f/-Ø contrasts to avoid teaching L2 learners the contrasts and to examine spontaneous classification of tokens into phonological categories (Hisagi & Strange, 2011). To minimize eye movements, a fixation cross remained in the centre of the computer screen for the entire block. Stimuli were presented binaurally at 70 dB via inserted headphones (Etymotic Reseach) and set to a comfortable volume by each participant. They were presented with a variable stimulus onset asynchrony (SOA) of 1000, 1020, 1040, 1060, 1080, 1100, 1120, 1140, 1160, 1180, or 1200 ms that was randomly distributed across stimuli. A variable SOA ('jitter') prevented participants who were unable to perceive /h/ from using any perceived delay between the consonant and vowel stimuli as a reliable cue for detecting deviants (Phillips, Pellathy, Marantz, Yellin, Wexler, Poeppel, McGinnis & Roberts, 2000). A relatively long SOA was used to increase the likelihood that discrimination occurred by referencing phonological representations stored in long-term memory rather than the detection of acoustic differences using short-term sensory memory templates (Molnar, Polka, Baum & Steinhauer, 2009, 2014). Prior to the EEG session, participants completed the language-background questionnaires. Experiment 1 lasted for approximately 20–25 minutes, including short breaks and a practice block of 7 sounds.

### *EEG recordings and ERP analyses*

Continuous EEG was recorded with a BioSemi ActiveTwo amplifier system (http://www.biosemi.com) from 64 cap-mounted Ag-AgCl electrodes according to the international 10–20 system, digitized online at 256 Hz and referenced to an additional active electrode (common mode sense; CMS). Eye movements were monitored by additional electrodes placed at the outer canthus of each eye (EOGH) and above and below the left eye (EOGV). Electrode offsets were kept below 25 μV.

Offline, the EEG was re-referenced to the averaged left/right mastoids and filtered with a 0.5–40 Hz band-pass filter using the EEProbe software package (Advanced Neuro-Technology, ANT; Enschede, the Netherlands). A trial was rejected whenever the standard deviation for a 200 ms moving window exceeded 25 mV in EOG or EEG channels included in ERP analysis (see below). EEG data for individual participants were also examined for eye movements and other artifacts; contaminated epochs were

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

excluded. Participants were included in further analyses if they contributed a minimum of 50 artifact-free trials in each of the Deviant and Standard conditions.

Four conditions were created for each participant: *Deviant H* (average response to 'ha' deviants among 'a' standards, and 'a' deviants among 'ha' standards), *Standard H* (average response to the corresponding standards), *Deviant F* (average response to the 'fa' deviants among 'a' and 'a' deviants among 'fa'), and *Standard F* (average response to the corresponding standards). Collapsing across sub-conditions in this way controlled for phonologically-irrelevant acoustic differences between stimuli (the same stimuli served as both standards and deviants), increased the signal-to-noise ratio, and decreased any potential block effects related to the participants' physical or mental state (Näätänen et al., 2007). Nevertheless, analyses were also conducted on the sub-conditions to test for differences between deviant types (e.g., deviant 'ha' with 'a' versus deviant 'a' with 'ha'). This revealed overall similar effects as the analyses of the collapsed conditions (differences between analyses, when present, are reported). ERPs were computed for standards immediately preceding a deviant (rather than all 235 Standards) so an approximately equal number of trials contributed to the deviant and standard conditions. ERPs were also computed for all standards (excluding those following a deviant). Analysis of these standards resulted in largely similar effects; differences, when present are reported.

For each participant, artifact-free ERP responses were averaged for each of the 4 conditions. This was done for a 650 ms interval, time-locked to the onset of the deviant/standard stimuli, including a 100 ms pre-stimulus baseline. Statistical analyses were conducted on the mean amplitude of the ERP waveform within time windows that were selected based on previous literature and visual inspection of the grand averages: 100–175 ms, 175–250 ms and 300–500 ms. For each time window, repeated-measures ANOVAs were performed on 12 representative lateral (F3, F4, F7, F8, C3, C4, T7, T8, P3, P4, P7, P8) and 3 midline (Fz, Cz, Pz) electrodes, as effects were most prominent at these electrodes. For the lateral electrodes, analyses involved Group as a between-subjects factor (native speakers, high and low proficiency L2 speakers); the within-subjects factors were: Condition (Deviant, Standard), Contrast (/h/-Ø, /f/-Ø), Anterior-Posterior ('AntPost'; Anterior, Central, Parietal), Hemisphere ('Hemi'; Left, Right) and Laterality (Medial, Lateral). For the midline sites, the factors were: Group, Condition, Contrast, and AntPost. Results are reported for main effects and interactions that involve the factor Condition. Interactions involving topographical variables (AntPost, Hemi, Laterality) were followed by simple effects analyses for each level of the variables involved in the interaction. Midline analyses are reported for results not already revealed by analyses of lateral electrodes. The Greenhouse-Geisser correction was applied to all analyses involving the AntPost factor (as it involves more than one degree of freedom), and corrected *p* values are reported.

## *Results*

### *Accuracy results*

Accuracy on the discrimination task (i.e., correct identification of a deviant) was analysed with a repeated measures ANOVA using Group as a between subjects variable; Contrast (/h/-Ø vs. /f/-Ø), and Deviant Type ('a' deviant among 'ha'/ 'fa' vs. 'ha'/'fa' deviants among 'a') were within-subjects variables. This revealed a main effect of Contrast [$F(1, 39) = 38.14$, $p < .001$] and Deviant Type [$F$ (1, 39) = 10.47, $p < .005$], indicating higher performance for the /f/-Ø compared to the /h/-Ø contrast and when 'a' was the deviant, compared to 'ha' or 'fa' deviants (Table 3). These were qualified by a Deviant Type x Contrast interaction [$F$ (1, 39) = 14.17, $p <$ .005]. Follow-up analyses showed a significant difference in Deviant type for the /h/-Ø [$F$ (1, 39) = 12.53, $p < .005$] but not the /f/-Ø contrast [$F$ (1, 39) = 3.14, $p < .09$]: detecting the deviant 'ha' with 'a' was more difficult than detecting the deviant 'a' with 'ha' [M = 65.7 % vs 87.5 %], whereas for the /f/- Ø contrast, performance was similar regardless of whether 'fa' or 'a' was the deviant [M = 96.9 % vs 95.3 %]. No group differences were observed (see Table 3 for group averages).

### *ERP results*

The ERP grand mean waveforms for the native speakers and the high and low proficiency L2 speakers for the /h/-Ø and /f/-Ø contrasts are presented in Figures 1–3 **.** These figures show a central-posterior negativity followed by a posterior medial positivity in response to the deviant stimuli for all groups and both contrasts. Statistical results are presented in Table 4 and described below.

Between 100–175 ms, a main effect of Condition was observed, indicating a significantly larger negativity for the deviant, compared to the standard, as seen in Figures 1–3. A Condition x AntPost interaction was also found. Follow up analyses showed that this negativity was present at central [$F(1, 39) = 26.32$, $p < .001$] and posterior [$F(1, 39) = 50.62$, $p < .001$], but not anterior ($p > .10$) electrodes. Finally, a Condition x Contrast interaction was also observed, which was followed by analyses for each contrast separately. For the control /f/-Ø contrast, a significant main effect of Condition [$F(1, 39) = 25.43$, $p < .001$] and a Condition x AntPost interaction [$F(2, 38) = 25.16$, $p < .001$] pointed to a negativity that was present at central [$F(1, 39) = 34.47$, $p < .001$] and posterior [$F(1, 39) = 33.27$, $p < .001$] electrodes. For the /h/-Ø contrast, a main effect of Condition that approached significance [$F(1, 39) = 3.47$, $p < .08$] and a significant Condition x

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

Table 3. *Accuracy in detecting deviant stimuli in Experiment 1*

| | | | Native Speakers | | High Proficiency | | Low Proficiency | |
|---|---|---|---|---|---|---|---|---|
| | Standard | Deviant | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Non-native Contrast | ha | a | 89.5 | 11.5 | 92.5 | 11.1 | 80.7 | 26.6 |
| /h- Ø/ | a | ha | 65.1 | 36.3 | 70.5 | 35.6 | 61.5 | 37.1 |
| Native Contrast | fa | a | 94.0 | 4.7 | 96.2 | 3.1 | 95.6 | 7.1 |
| /f- Ø/ | a | fa | 96.1 | 3.0 | 9.8 | 3.1 | 97.7 | 2.3 |

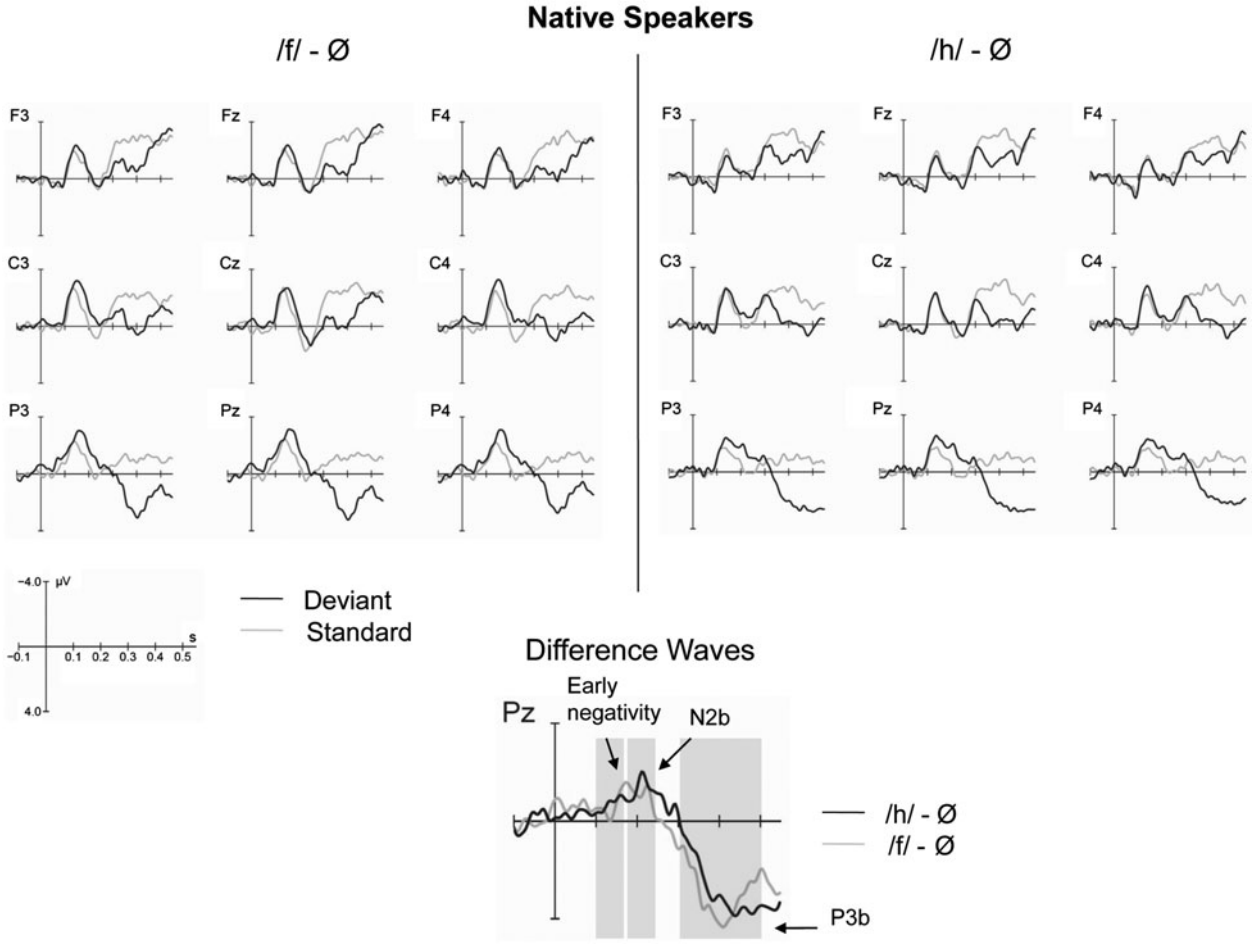



Figure 1. Grand average ERPs in response to the standard and deviant stimuli for the Native Speakers for the /f/-Ø and /h/-Ø contrasts. In the bottom panel, difference waves (deviant-standard) show the responses in the three time intervals.

AntPost interaction [$F(2, 38) = 5.69$, $p < .001$], pointed to a significant negativity at posterior electrodes [$F(1, 39) = 12.49$, $p < .001$]. No interactions with group were observed[3] .

Between 175–250 ms, very similar results were observed. A main effect of Condition was found, indicating a larger negativity for the deviant, compared to the standard stimuli. A Condition x AntPost interaction was also observed. Follow up analyses again revealed this

[3] Comparison of the deviant and standard stimuli (defined as all standards that did not follow a deviant, rather than only those standards that directly proceed a deviant, as reported previously), revealed largely similar effects. One exception is that follow-up analysis of the significant Condition x AntPost interaction, revealed a broadly distributed negativity that was significant at anterior [$F(1, 39) = 11.28$, $p < .005$], central [$F(1, 39) = 44.31$, $p < .001$] and posterior [$F(1, 39) = 34.15$, $p < .001$] electrodes, albeit still largest at central-posterior sites. No group differences were observed.

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

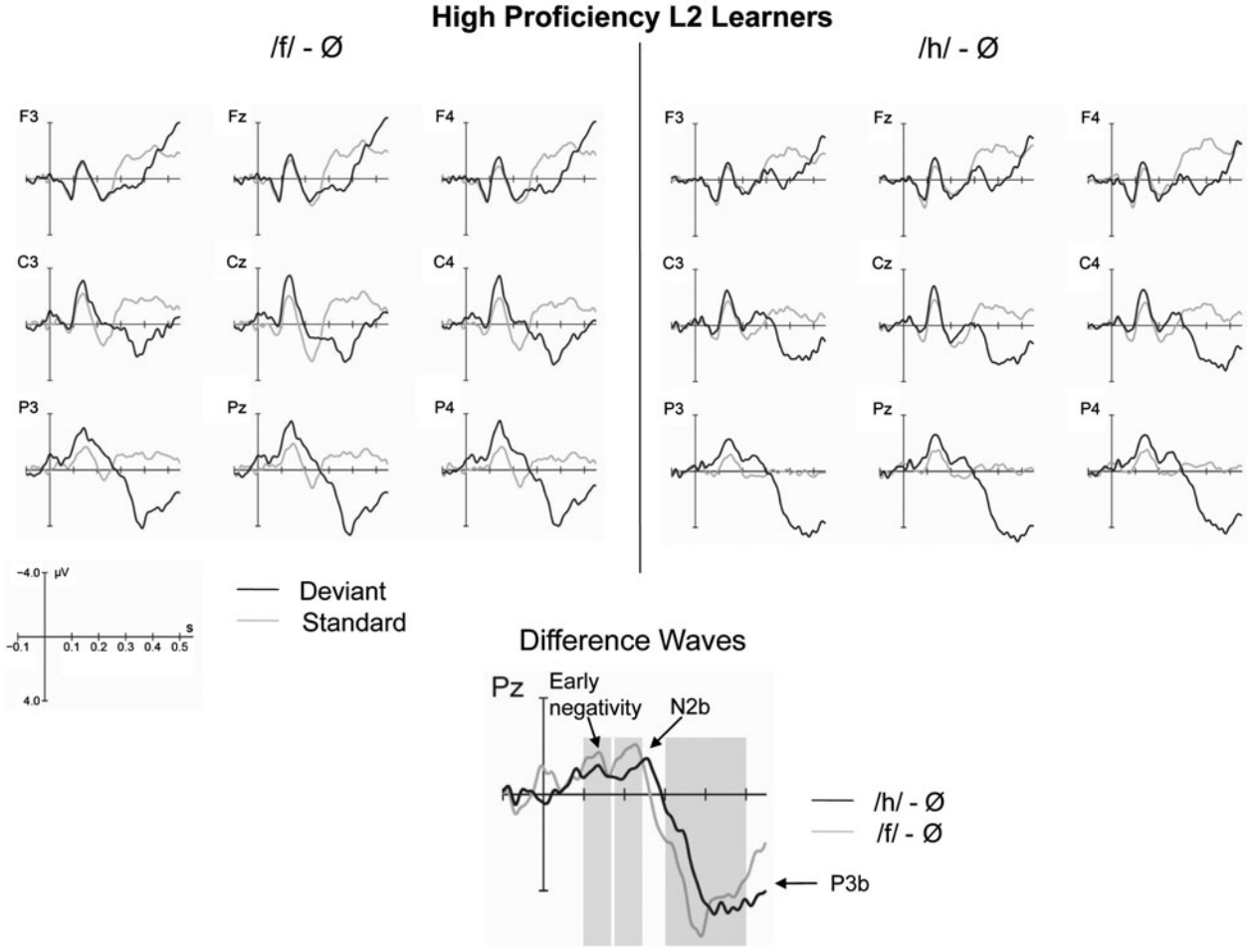



Figure 2. Grand average ERPs in response to the standard and deviant stimuli for the High Proficiency L2 learners for the /f/-Ø and /h/-Ø contrasts. In the bottom panel, difference waves (deviant-standard) show the responses in the three time intervals.

negativity was present at central [$F(1, 39) = 24.45$, $p < .001$] and posterior [$F(1, 39) = 78.03$, $p < .001$] electrodes. Two significant interactions with the factor Condition and Contrast were observed (Condition x Contrast x Hemi; Condition x Contrast x Hemi x Lat), suggesting the effect might have different distributions for the two contrasts (Table 4). Follow up analyses conducted for each hemisphere revealed a significant Condition x Contrast interaction for right hemisphere [$F(1, 39) = 14.75$, $p < .001$], but not left hemisphere electrodes ($p > .10$). This pointed to a strongly right-lateralized negativity for the /f/-Ø contrast and a more bilaterally distributed negativity for the /h/-Ø contrast. Again no interactions with group were observed.

Between 300–500 ms, a main effect of Condition was found, indicating a larger postivity for the deviant compared to the standard stimuli, as well as multiple interactions with electrode location factors (Table 4 ). Follow up analyses revealed highly significant effects of Condition at each level analyzed, although the largest effects were found at posterior medial sites [$F(1,39) = 84.99$, $p < .001$], consistent with P3b effects. Two significant interactions with the factor Condition and Contrast were observed (Condition x Contrast x AntPost x Hemi; Condition x Contrast x AntPost x Lat). These revealed a positivity for both contrasts that was highly significant at each level analysed, with one exception, at medial central sites it was significant for /h/-Ø ($F(1, 39) = 5.63$, $p<.05$) but not /f/-Ø ($p > .10$). One marginal interaction with the factor Group and Contrast was observed in this time window [Condition x Contrast x Hemi x Group: $F(2,39) = 2.55$, $p = .091$]; however, no further significant differences between groups or contrasts was found. As can be seen in Figures 1–3 , all groups exhibited a large posterior medial positivity for both contrasts ($ps < .005$).

### *Analysis of mastoid electrodes*

To better differentiate MMN and N2b effects, additional analyses were conducted on ERP data recorded from

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

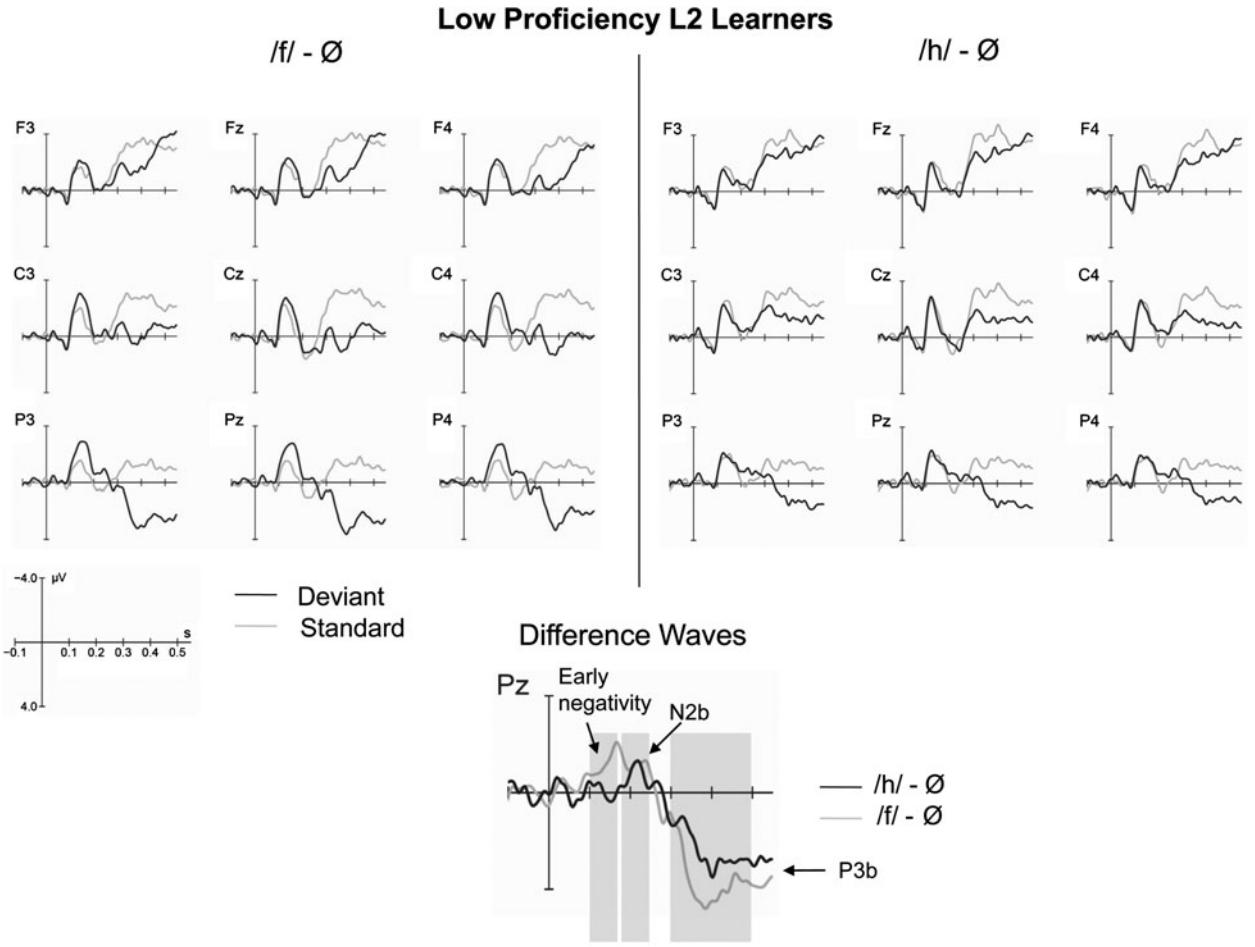



Figure 3. Grand average ERPs in response to the standard and deviant stimuli for the Low Proficiency L2 learners for the /f/-Ø and /h/-Ø contrasts. In the bottom panel, difference waves (deviant-standard) show the responses in the three time intervals.

mastoid electrodes between 100–175 ms and 175–250 ms (nose referenced). The MMN, but not the N2b, tends to reverse in polarity along the sylvian fissure (Novak, Ritter, Vaughan & Wiznitzer, 1990); a positivity at mastoid sites would indicate effects due to the MMN alone. Repeated measures ANOVAs were conducted to compare effects at mastoid electrodes with effects at electrode Pz (which showed the largest negativity in the analyses reported earlier). The factors were Condition (Deviant, Standard), Contrast (/h/-Ø, /f/-Ø), Electrode (Pz, left mastoid, right mastoid), and Group (native speakers, high and low proficiency L2 speakers). Greenhouse-Geisser corrected $p$ values are reported for interactions involving the factor Electrode.

Between 100–175 ms, a Condition x Electrode [$F(2,38) = 16.73$, $p < .001$] interaction was observed. Analysis at electrode Pz revealed a significant main effect of Condition [$F(1,38) = 10.29$, $p < .005$], indicating a negativity, as reported earlier. At left and right mastoids, response to the deviant stimuli was more positive than to the standard, although this was not significant ($F$s < 1). Between 175–250 ms, a main effect of Condition [$F(1,39) = 17.04$, $p < .005$] and a Condition x Electrode interaction [$F(2,38) = 32.47$, $p < .001$] revealed significant negativities at electrode Pz [$F(1,38) = 43.72$, $p < .001$] as well as the left [$F(1,38) = 4.22$, $p < .05$] and right [$F(1,38) = 4.80$, $p < .05$] mastoids. Thus, a polarity reversal at mastoid electrodes was observed between 100–175 ms but not between 175–250 ms, as seen in Figure 4. No group differences were observed.

### *Correlation Analyses*

To further examine whether there may be a relationship between L2 oral proficiency and these ERP components, we correlated the amplitude of the difference wave (i.e., ERP response for deviant – standard stimuli) within the three time windows at electrode Pz (where effects were largest) for all L2 learners for the /h/-Ø contrast with the 6 proficiency ratings (i.e., general pronunciation, pronunciation of H words, vocabulary,

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

Table 4. *Global ANOVA for ERP effects in Experiment 1*

| | *df* | 100–175 ms *F value* | 175-250 ms *F value* | 300-500 ms *F value* |
|---|---|---|---|---|
| **Lateral Electrodes:** | | | | |
| Condition | 1, 39 | 19.81**** | 18.35**** | 47.91**** |
| Condition x Hemi | 1, 39 | 1.54 | >1 | 17.30**** |
| Condition x AntPost | 2, 38 | 18.82**** | 40.14**** | 47.11**** |
| Condition x AntPost x Hemi | 2, 38 | 2.60 | 1.60 | 32.70**** |
| Condition x Hemi x Lat | 1, 39 | >1 | >1 | 32.65**** |
| Condition x AntPost x Lat | 2, 38 | 2.01 | >1 | 36.44**** |
| Condition x AntPost x Hemi x Lat | 2, 38 | >1 | >1 | 30.81**** |
| Condition x Cont | 1, 39 | 8.16** | 3.72+ | >1 |
| Condition x Cont x Lat | 1, 39 | 2.93+ | >1 | >1 |
| Condition x Cont x Hemi | 1, 39 | >1 | 4.51* | >1 |
| Condition x Cont x AntPost x Hemi | 2, 38 | 1.22 | 1.03 | 11.26**** |
| Condition x Cont x AntPost x Lat | 2, 38 | >1 | 2.11+ | 4.07* |
| Condition x Cont x Hemi x Lat | 1, 39 | >1 | 4.70* | 2.50 |
| Condition x Cont x AntPost x Hemi x Lat | 2, 38 | >1 | >1 | 3.19+ |
| Condition x Cont x Hemi x Group | 2, 39 | >1 | >1 | 2.55+ |
| **Midline Electrodes:** | | | | |
| Condition | 1, 39 | 12.67**** | 8.22** | 56.46**** |
| Condition x AntPost | 2, 38 | 12.66**** | 23.52**** | 27.88**** |
| Condition x Cont | 1, 39 | 6.74* | 2.16 | >1 |
| Condition x Cont x AntPost | 1, 39 | 1.58 | 1.05 | 2.17+ |

Condition (Deviant vs. Standard), Cont = Contrast (/h-Ø vs. /f/-Ø), Lat = laterality, Hemi = hemisphere, AntPost= Anterior-Posterior, Group (Native Speakers, High Proficiency, Low Proficiency)
* p < .05, ** p < .01, *** p <.005, **** p < .001

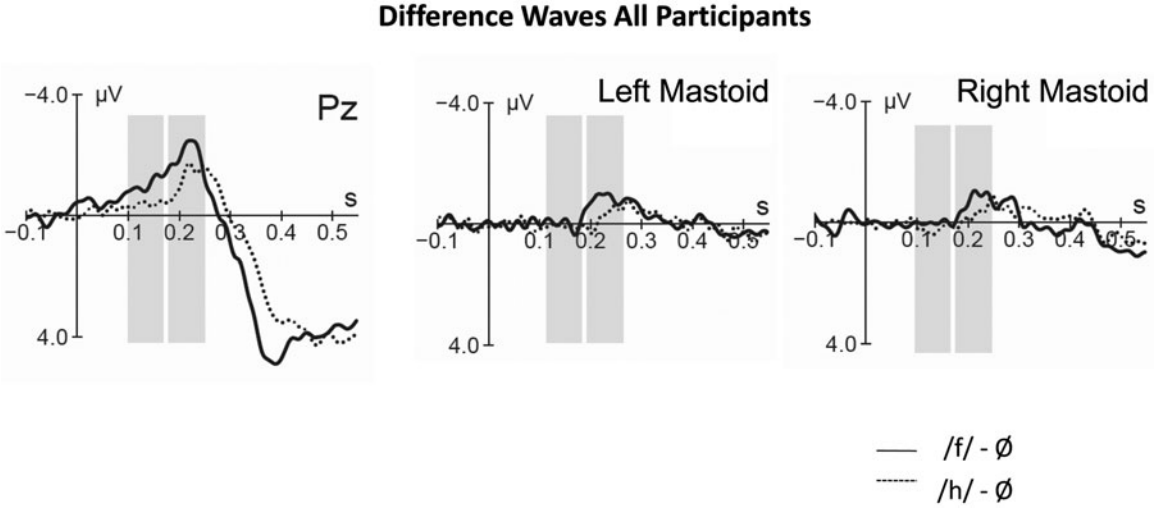



Figure 4. Difference waves for all participants combined, comparing the response at electrode Pz and the left and right mastoids. For both contrasts, a negativity can be seen at mastoid electrodes in the second, but not the first, time interval.

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

grammar, fluency and overall impression; averaged across raters). This revealed no significant relationship between L2 proficiency and ERP amplitudes within any time window (all *ps* >.10).

### *Discussion*

Experiment 1 found no significant differences in the neurocognitive processes recruited by native speakers and late L2 learners when categorizing syllables involving a non-native phonological contrast. This is noteworthy because the /h/-Ø contrast is notoriously difficult for native French speakers to perceive and produce in spoken word contexts (LaCharité & Prévost, 1999; Janda & Auger, 1992; John, 2006). Our results indicate that, when tested with simple syllable stimuli and during a task that focuses attention to the sound of the speech signal, even relatively low proficiency L2 learners can accurately discriminate this non-native contrast.

No significant group differences were observed in accuracy performance: all groups showed high performance for the /f/-Ø contrast and for the /h/-Ø contrast when 'ha' was the standard. Unexpectedly, even the native speakers' performance was rather variable when 'ha' was the deviant; perhaps reflecting the fact that we did not tell participants about the contrasts or which sounds to respond to. Given the low acoustic saliency of /h/, this may have made it more difficult for all participants to differentiate 'ha' from the standard 'a' than to differentiate the more salient 'fa' from 'a'. This reinforces why /h/-Ø is a difficult contrast for L2 learners to acquire.

Regarding the ERP data, the ASP model predicts that L2 speakers can recruit attention-driven mechanisms to aid in the processing of non-native phonological contrasts. Consistent with this, we found no significant group differences in the ERP effects exhibited during the later time windows (which we interpret to reflect N2b and P3b effects), nor did these effects differ substantially for the native /f/-Ø and non-native /h/-Ø contrasts. Thus, as predicted, later stages of L2 phonological processing that involve attention (i.e., attention orientation, response preparation) were similar in L2 learners and native speakers. More surprising was the finding that the negativity in the early time window was also similar between groups and contrasts (albeit more broadly distributed for /f/-Ø). This, along with its central-posterior maximum, allows us to speculate as to neurological processes this negativity may reflect.

One possibility is that the early negativity may reflect N100 recovery from refractoriness effects. The N100 amplitude tends to decrease when the eliciting stimulus is repeated and there is a short interval between presentations, because the underlying neural generators have not yet fully recovered from their response to a previous stimulus presentation (Budd, Barry, Gordon, Rennie & Michie, 1998). According to this account, neurons that respond to the standard stimuli may have experienced a larger decrease in excitability upon repeated stimulus presentations, because the inter-standard interval was larger than the inter-deviant interval, thus resulting in a larger negativity for the deviant stimuli. The timing and central-posterior distribution of our early negativity support this interpretation. Applied to the present study, this implies that the population of neurons that responded to the two tokens of the 'a' stimulus were different than those that responded to the 'ha' stimulus. That no significant group differences were observed, suggest that the L2 learners differentiated the phonological categories at an early/low-level of perceptual processing in a similar way as native speakers.

Another possibility is that the early negativity reflects later, attention-based stages of processing and includes early contributions from the following N2b component. Indeed, the negativity in both time windows was largest at central-posterior electrodes (more typical for N2b effects; Novak et al., 1990), and the polarity reversal at mastoid electrodes, although significant, was small. At first glance, both negativities appear to resemble N2b effects. However, this distribution is also consistent with MMN effects reported in other studies. For example, although a polarity reversal at mastoid sites is typically used to differentiate MMN and N2b effects, it is not found in all studies. It was not found for the MMN reported by Molnar, Baum, Polka & Steinhauer, (2009, 2014) who also used a categorical odd-ball paradigm and controlled for phonologically-irrelevant acoustic differences between standards and deviants by comparing the response to a vowel to itself when it served as both a standard and a deviant, as in the present study. It may be that while mastoid positivities reflect activity in the auditory cortex elicited when processing acoustic differences between stimuli, activity at scalp electrodes reflects higher order cognitive processes, such as phonological categorization (Sussman, Kujala, Halmetoja, Lyytinen, Alku & Näätänen, 2004). It also seems unlikely that large and very early N2b effects would be found in listeners who did not first engage in earlier processing stages (e.g., as reflected by N100 or MMNs).

Thus, a third possibility is that the early negativity is a central-posterior MMN. Although MMNs are typically largest over frontal sites (Näätänen et al., 2007), similar posterior MMNs have also been reported in studies that used syllables or word stimuli, possibly reflecting both lexical and phonological activation (Kazanina et al., 2006; Shytrov, Hauk & Pulvermüller, 2004). A broadly distributed MMN was also observed by Mah et al. (2015) for native speakers in response to the /h/- Ø contrast, and appeared more posteriorly for 'linguistic' stimuli (syllables 'hum' vs. 'um') than for non-linguistic stimuli (sounds 'hf' vs. 'f'). The fact that Mah et al. report

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

MMN effects for the 'hum'/'um' contrast in English native speakers, but not in French L2 learners, whereas no group differences were found in the present study is note worthy, given both studies examined the /h/- Ø contrast in similar categorical odd-ball paradigms. The major difference between studies is that whereas here participants attended to the stimuli in order to perform the phonological categorization task, in Mah's study participants' attention was drawn to a silent movie and no discrimination task was required. Similar findings were also reported by Hisagi et al. (2010), who report that the MMN exhibited by naïve listeners in response to a non-native phonological contrast was larger when the contrast was presented during a task that required participants to attend directly to the auditory features of the stimuli, as opposed to during a task that required attention to a visual display. Together, these findings support the notion that L2 phonological processing, as indexed by early negativities, is enhanced in contexts that direct attention to the contrasting sounds.

A number of studies suggest that attention can sharpen the representation of auditory information, and thus allow for MMNs to be observed in response to difficult-to-discriminate contrasts (see Sussman, 2007). How can this be, given the MMN is thought to reflect automatic stages of processing? Sussman (2007) suggests that attention can modulate the detection of a deviant by enhancing the encoding of the standard stimulus. Forming a sensory memory trace of the 'standard' stimulus involves organizing and grouping sounds in order to establish a sound context. When multiple groupings are possible, such as in a categorical oddball paradigm in which physically different tokens can be grouped into one or more phonological categories (as was the case in the present study), then directing attention to these groupings can influence the sound context against which the deviant is compared, thereby modulating the MMN response. This can increase the saliency of a phonological contrast, particularly when the contrast is difficult to discriminate (Gomes, Molholm, Ritter, Kurtzberg, Cowen & Vaughan,2000). As in the current study, attention to the contrasting phonological features could also account for Hisagi, Garrido-Nag, Datta and Shafer's (2015) finding that even L2 speakers with relatively low L2 oral proficiency can nevertheless perform well on difficult L2 phonological contrasts in some conditions. In Experiment 2, we investigate the role of attention further by examining phonological processing during a task that directs attention to the meaning of the speech signal, rather than to its sound.

## Experiment 2

During spoken word comprehension, phonetic segments must be automatically differentiated so listeners can focus on word meaning (Strange, 2011; Strange & Shafer, 2008). The ASP model predicts that, in these situations, without fully automatic L2 phonological processing routines, late L2 learners will fail to discriminate non-native phonological contrasts. Consequently, they may perceive minimal pairs and subtle mispronunciations (pseudowords) that involve non-native phonological contrasts (e.g., hair-air; happy-'appy) as identical, even though they can discriminate the same contrasts in contexts that focus attention to its phonological constituents.

In Experiment 2, we tested this by means of the N400 pseudoword-effect. The N400 is a negative-going ERP wave that peaks approximately 400 ms after the presentation of a word or word-like string and is a reliable index of lexical representation, activation, and access (Lau, Phillips & Poeppel, 2008). The N400 pseudoword-effect is the overall larger amplitude and longer duration N400 elicited by pseudowords (orthographically legal and pronounceable word-like strings) compared to words. It is thought to reflect a prolonged memory search for potentially low-frequency, low-familiarity words that may match the pseudoword (Holcomb & Neville, 1990; Holcomb, Grainger & O'Rouke, 2002). The N400 pseudoword-effect provides a unique index of phonological discrimination because it requires listeners to perceive a difference between words and pseudowords. If, however, pseudowords differ only with respect to a difficult to discriminate non-native contrast, then both might activate the same lexical representation and elicit a similar N400 response (Friedrich, Lahiri & Eulitz, 2008). Thus, the N400 pseudoword effect characteristic of native speakers – a larger N400 to pseudowords compared to words – can index whether L2 learners discriminate non-native phonological contrasts in a lexical context (c.f., Sebastian-Gallés, Rodríguez-Fornells, Diego-Balaguer & Díaz, 2006).

The same participants from Experiment 1 heard English words and pseudowords that were created by manipulating the /h/-Ø and /f/-Ø contrasts (by removing the /h/ or /f/ from real h- or f-initial words and by adding an /h/ or /f/ to vowel-initial words; Table 1). To ensure attention was directed away from the phonological form of the stimuli, participants performed a semantic categorization task ("press a button when you hear a fruit or vegetable word"), and were not told about the pseudowords or phonological contrasts. We expected native English speakers would differentiate both contrasts and display N400 pseudoword effects. The ASP model predicts that in this lexical context without attention to phonological form, L2 speakers should have difficulty differentiating words and pseudowords involving the /h/-Ø contrast, although they may differentiate those involving the /f/-Ø contrast, as it exists in French. We test whether L2 proficiency may influence processing outcomes.

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

### *Method*

#### *Participants*

The same participants took part in Experiment 1 and 2.

#### *Stimuli*

Two lists of stimuli were created, each containing 90 words (30 /h/-initial, 30 /f/-initial, and 30 vowel-initial) and 120 pseudowords (30 missing /h/, 30 missing /f/, 30 with an additional /h/, and 30 with an additional /f/). Pseudowords were derived from the words in two ways: by removing the initial consonant from the /h/- and /f/-initial words and by adding an /h/ or /f/ to vowel initial words (see Table 1). This is akin to mispronunciations reported by French L2 speakers of English, in which both deletion and hypercorrection (e.g., 'elp for 'help' and 'h-as' for 'as') have been reported (Janda & Auger, 1992). The words were one to four syllables in length; average duration 688 ms (SD = 159 ms). Only high frequency content words familiar to beginning L2 learners of English were included; French–English cognates and inter-lingual homophones were excluded.

Lists were counterbalanced so that a given item was presented as a word in one list and a pseudoword in the other. Lists were equivalent in terms of word frequency (Kučera & Francis, 1967; Hyperspace Analogue to Language, Lund & Burgess, 1996), phonological neighbourhood density (English Lexicon Project Database; Balota,Yap, Cortese, Hutchison, Kessler, Loftis, Neely, Nelson, Simpson & Treimna, 2007), word and pseudoword duration, and the first vowel of the vowel-initial words and /h/-less and /f/-less pseudowords. Because the pseudowords for one list were derived from the words of the other list, this limited any phonological differences to the contrasts of interest (/h/-Ø/ and /f/-Ø/). It also ensured that any N400 differences between words and pseudowords observed in the L2 learners were truly due to the word/pseudoword status and not differences in word frequency and phonological neighbourhood density between words and any pseudowords misidentified as words.

Two measures were taken to direct participants' attention to word meaning. First, stimuli were presented in a semantic priming paradigm: words that were semantically related to the test words (or words from which the pseudoword was derived) were presented following each word and pseudoword, as well as following semantically unrelated words (210 in total, presented twice). Second, participants' task was to respond to word meaning (i.e., with a mouse click when they heard a fruit or vegetable word). To this end, 43 fruit and vegetable words were presented as part of a pair with a semantically unrelated word and were randomly intermixed with the critical words and pseudowords. Including these 'filler' items ensured that participants attended to word meaning. These items did not begin with /h/ or /f/ and were not included in ERP analyses.

Stimuli were recorded by a female native Canadian English speaker in a sound attenuating chamber using a digital recorder and edited into individual tokens using Cool Edit 1996 software package (Syntrillium Software; Phoenix, AZ). A 15 ms silence was left before and after each word. All stimuli were then normalized to 70 dB using Praat software package version 5.2.35 (Boersma & Weenink, 2009). Stimulus order within each list was randomized and equally distributed into 6 presentation blocks of approximately 9 minutes each. Block and item order was reversed for half of the participants.

#### *Procedure*

Experiment 2 was conducted before Experiment 1, so participants would not be aware of the contrasts. Participants were told they would hear a series of English words presented in pairs. They were asked to respond with a mouse click when they heard a fruit or vegetable word. No explicit mention was made of the pseudowords before testing.

Each trial began with the presentation of a fixation cross that remained in the centre of the computer screen throughout stimuli presentation. One second later, a word or pseudoword was presented orally, followed by an inter-stimulus interval of 600 ms and the second word in the pair. The symbol '???' appeared on the screen one second after the offset of the second word, prompting participants to respond. After 2 seconds, this was replaced by the symbol '–', prompting for eye blinks, which disappeared after 2 seconds and was replaced by the fixation cross, signalling the start of the next trial. The experiment lasted approximately 1 – 1 1/2 hours, including short breaks and a practice block of 9 word/pseudoword pairs. See Figure 5 for the trial structure of experiment 2. ERP analyses were conducted on the first word of the pair only; analyses of the second word in the pair (i.e., priming effects) will be reported in a separate paper.

#### *EEG Recordings and ERP Analysis*

EEG recording and ERP pre-processing (mastoid re-referencing, filtrating, artefact rejection) procedures were the same as in Experiment 1. ERPs were computed for each participant for 4 conditions: H WORD (average of /h/-initial and vowel-initial words), H PSEUDOWORD (average of /h/-less and /h/-extra pseudowords), F WORD (average of /f/-initial and vowel-initial words), and F PSEUDOWORD (average of /f/-less and /f/-extra pseudowords; see Table 1). By collapsing across sub-conditions, we controlled for: (1) length differences between words and pseudowords and between vowel- and consonant-initial items; and (2) for items' initial letter. These factors could influence the latency and amplitude of the N400 and earlier ERP components that could

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

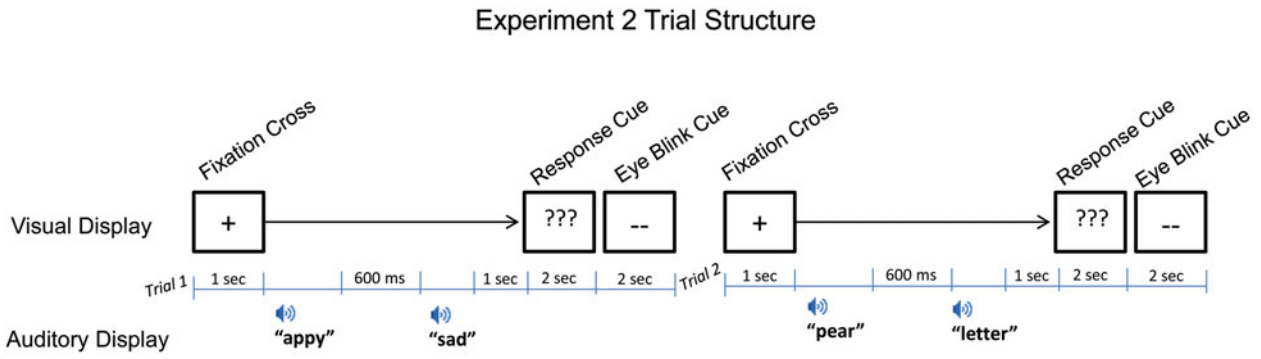



Figure 5. (Colour online) The trial structure of Experiment 2. Each trial consisted of an orally presented stimulus pair. The first stimulus was either a word or a pseudoword; the second was a semantically related or unrelated word. Participants' semantic monitoring task was to respond to fruit or vegetable words that were randomly intermixed with the test stimuli. The ERP responses to the first stimulus in a pair (word/pseudowords) are reported.

carry into the N400 time window. Thus, we can be more confident that any ERP effects observed are due to processing of lexical status.

Participants were included in ERP analyses if they contributed a minimum of 30 artifact-free trials in each of these 4 conditions. For each participant and condition, artifact-free ERP responses were time-locked to the onset of the first word of each pair. This was done for a 1400 ms interval, including a 100 ms pre-stimulus baseline. Statistical analyses were conducted on the mean amplitude of ERPs within two time windows – 500–700 ms and 700–900 ms – which were selected based on previous literature of N400 pseudoword effects (Friedrich, Lahiri & Eulitz, 2008; Holcomb & Neville, 1990) and visual inspection of the grand averages. As in Experiment 1, analyses were conducted on twelve lateral (F3, F4, F7, F8, C3, C4, T7, T8, P3, P4, P7, P8) and 3 midline (Fz, Cz, Pz) electrodes, where effects were most prominent. For each time window, repeated-measures ANOVAs were performed on lateral and midline sites separately. The factors were Condition (word, pseudoword), Contrast (/h/-Ø or /f-Ø/), Group (Native Speakers, High and Low Proficiency), Anterior-Posterior ('AntPost'; anterior, central, parietal; Greenhouse-Geisser correction applied and corrected *p* values reported), Hemisphere (left, right) and Laterality (medial, lateral). Results are reported for effects involving the factor Condition; midline analyses are discussed for effects not already revealed in analysis of lateral electrodes.

## *Results*

### *Accuracy results*

Accuracy on the semantic categorization task was defined as correct identification of fruit and vegetable words (percent hits). Group means were compared using a univariate ANOVA; the Welch F-ratio is reported in light of unequal variance between groups. This revealed no significant difference between groups [$F(224.19) = 1.76$, $p > .10$], (native speakers: $M = 80.31$ % $SD = 10.73$, range: 58–95%; high proficiency: $M = 84.29$ % $SD = 12.02$, 53–98%; low proficiency: $M = 71.01$ % $SD = 23.63$; 35–98%)[4].

### *ERP results*

ERPs for the native speakers, high proficiency and low proficiency groups for the /h/-Ø and /f/-Ø contrasts are presented in Figures 6 to 8. These figures show a larger and longer lasting N400 for pseudowords compared to words (i.e., the N400 pseudoword effect) in all cases except the low proficiency group in response to the /h/-Ø contrast. Statistical analyses are described below.

Between 500–700 ms, a main effect of Condition was observed [$F(1,39) = 7.10$, $p < .05$]: a larger N400 was found for pseudowords compared to words. This was qualified by a highly significant Condition x Group interaction [$F(2,39) = 9.05$, $p \leq .001$]. Analyses for each group revealed a significant effect of Condition for the native speakers [$F(1,14) = 12.99$, $p < .005$] and high proficiency [$F(1,11) = 5.68$, $p < .05$] groups only. For the low proficiency group, the effect of Condition did not reach significance [$F(1,14) = 3.77$, $p < .08$].

Between 700–900 ms, a main effect of Condition was again observed [$F(1,39) = 27.42$, $p < .001$]: a larger N400 was found for pseudowords compared to words. Importantly, a Condition x Contrast x AntPost x Group interaction was also observed (see Table 5 ), which was followed by analyses for each group and contrast separately. For the native speakers, a main effect of Condition was observed for both the /f/-Ø [$F(1,14) = 4.88$, $p < .05$] and /h/-Ø [$F(1,14) = 7.43$, $p < .05$] contrasts. Similarly, for the high proficiency group, a significant

[4] The low proficiency group's variable performance may reflect the significantly lower word frequency of the fruit and vegetable items compared to that of the test stimuli (i.e., the /h/- /f/- and vowel-initial words; $F(3, 210) = 2.93$, $p < .05$).

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

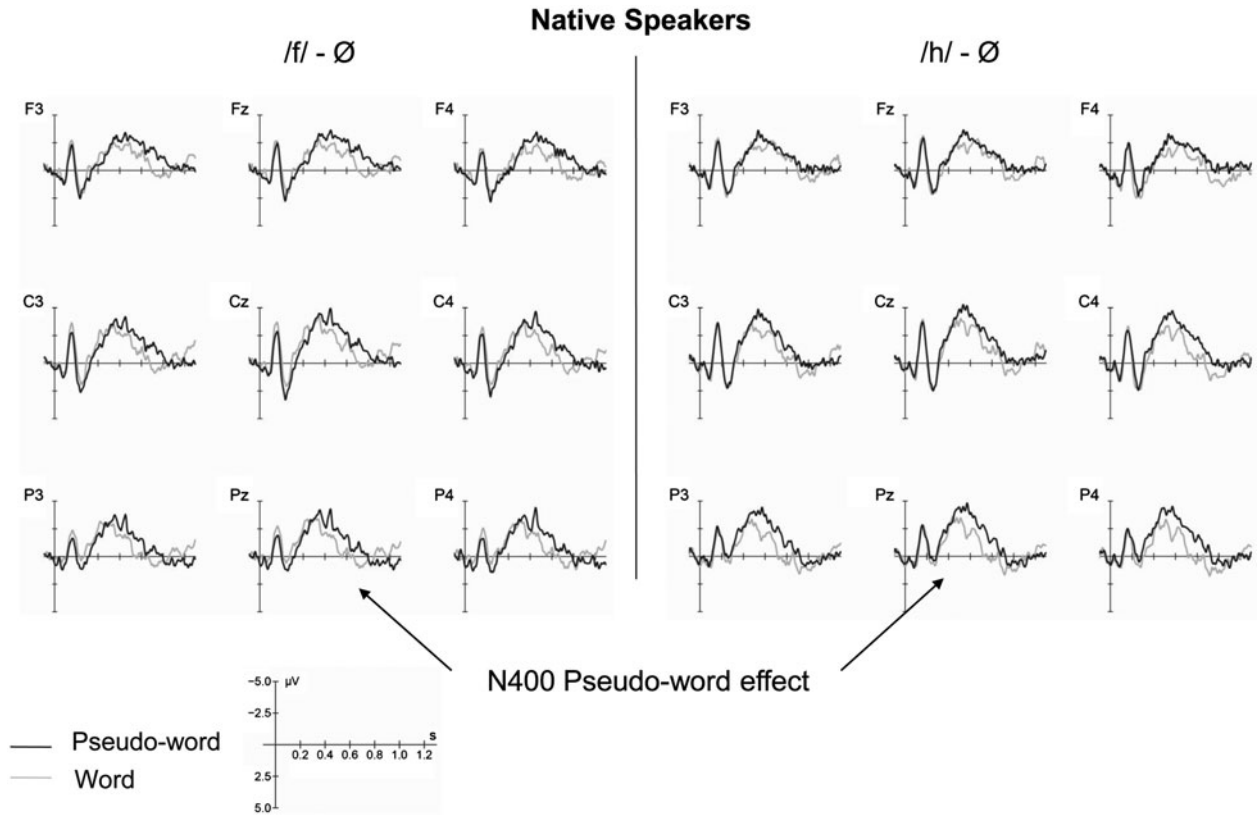



Figure 6. Grand average ERPs for the Native Speakers in response to words and pseudowords for the /f/-Ø and /h/-Ø contrasts in Experiment 2. An N400 pseudoword effect can be seen for both contrasts.

Table 5. *Global ANOVA for N400 pseudoword effects within the 700–900 ms time window*

| | df | F value |
|---|---|---|
| **Lateral Electrodes:** | | |
| Condition | 1, 39 | 24.42**** |
| Condition x Lat | 1, 39 | 11.46*** |
| Condition x AntPost | 2, 38 | 4.81$^{+}$ |
| Condition x Cont x Lat | 1, 39 | 5.73* |
| **Condition x Cont x AntPost x Group** | 4, 78 | 1.92* |
| Condition x Cont x Hemi x Group | 2, 39 | 2.65$^{+}$ |
| Condition x Cont x Lat x Hemi x Group | 2, 39 | 2.58$^{+}$ |
| **\Midline Electrodes:** | | |
| Condition | 1, 39 | 25.45**** |
| Condition x AntPost | 2, 38 | 1.91$^{+}$ |
| **Condition x Cont x AntPost x Group** | 4, 78 | 3.47$^{+}$ |

Lex = Lexical Status, Lat = Laterality, AP = Anterior-Posterior, Hemi = Hemisphere, Cont = Contrast (/h-Ø vs. /f/-Ø), Group (Native Speakers, High Proficiency, Low Proficiency)
* p < .05, ** p < .01, *** p < .005, **** p < .001

main effect of Condition was found for both the /f/-Ø [$F(1,11) = 42.74$, $p < .001$] and /h/-Ø [$F(1,11) = 15.61$, $p < .005$] contrasts. This is shown in Figure 7 as a large and broadly distributed negativity for pseudowords (and a smaller negativity and/or positivity for the words). For this group, a significant Condition x Laterality interaction was also observed for the /f/-Ø contrast [$F(1,11) = 7.16$, $p < .05$], pointing to a somewhat larger effect over medial [$F(1,11) = 40.69$, $p < .001$] than lateral [$F(1,11) = 32,62$, $p < .001$] electrodes, as is often found for N400 effects. In contrast, for the low proficiency group no significant effects or interactions with the factor Condition were observed for the /h/-Ø contrast ($p > .10$). For /f/-Ø, two interactions were observed: Condition x AntPost [$F(2,13) = 8.50$, $p < .005$] and Condition x Laterality [$F(1,14) = 7.59$, $p < .05$]. Follow up analyses revealed a significant effect of Condition at posterior medial electrodes [$F(1,14) = 5.55$, $p < .05$]. Similarly, at midline electrodes, a significant negativity was observed for this group at electrode Pz for /f/-Ø/ [$F(1,14) = 5.61$, $p < .05$] but not for /h/-Ø [$ps > .10$]. As seen in Figure 8, for the low proficiency group, words and pseudowords involving the /h/-Ø contrast resulted in similar N400 responses (i.e., no N400 pseudoword effect) and a small effect at posterior medial electrodes for the /f /-Ø contrast.

The Condition x Contrast x AntPost x Group interaction described above was also explored with a set of analyses comparing the native speakers to each

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

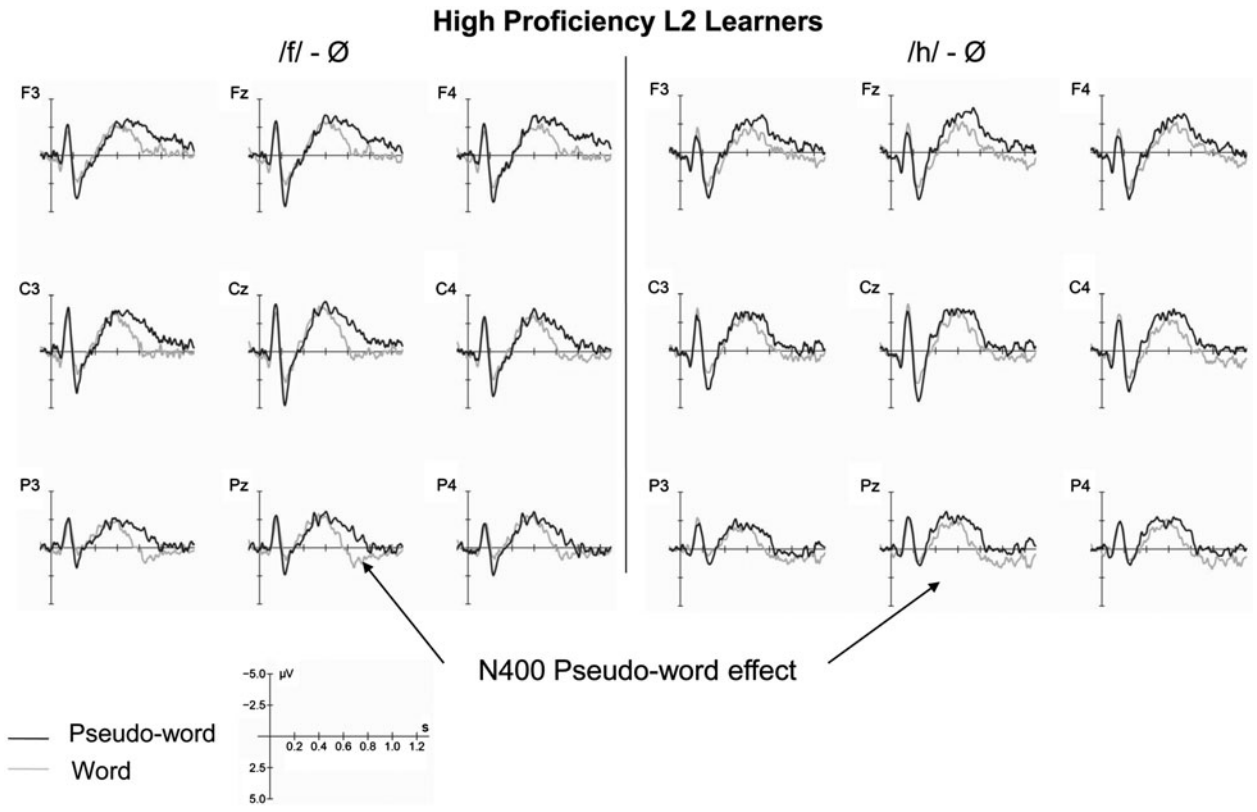



Figure 7. Grand average ERPs for the High Proficiency L2 learners in response to words and pseudowords for the /f/-Ø and /h/-Ø contrasts in Experiment 2. An N400 pseudoword effect can be seen for both contrasts.

of the high and low proficiency groups. Comparing the native speakers with the high proficiency group revealed no significant effect of Group ($p < .10$). In contrast, comparing the native speakers with the low proficiency group revealed a significant Condition x Contrast x AntPost x Group interaction [$F(2,27) = 3.75$, $p < .05$], and significant group differences at posterior electrodes for the /h/-Ø contrast [Condition x Group: $F(1,28) = 5.21$, $p < .05$] but not for the /f/-Ø contrast [Condition x Group: $F < 1$]. Thus, the low proficiency group differed from the native speakers in their processing of the English-specific /h/-Ø contrast, and no differences were observed between native speakers and high proficiency L2 learners for either contrast.

### *Correlation analyses*

To further investigate the relationship between L2 oral proficiency and the N400 pseudoword effect, we correlated the amplitude of the N400 difference wave for all L2 learners for the /h/-Ø contrast (at electrode Pz within the 700–900 ms time window) with the 6 proficiency ratings (i.e., general pronunciation, pronunciation of H words, vocabulary, grammar, fluency and overall impression; averaged across raters). N400 amplitude was correlated with most aspects of oral L2 proficiency ($p$ values corrected for multiple comparisons): participants with a large N400 pseudoword effect were those who received high ratings on overall pronunciation [$r = -.680$, $p < .001$], nativelikeness [$r = -.605$, $p \leq .001$], vocabulary [$r = -.551$, $p = .003$], pronunciation of 'h' words [$r = -.509$, $p = .007$] and fluency [$r = -.507$, $p = .007$]. Individuals with higher accuracy on the semantic monitoring task also showed larger N400 effects, although this relationship was not significant using corrected $p$ values [$r = -.376$, $p = .027$]. Figure 9 displays the relationship between pronunciation ratings and the N400 pseudoword effect for the /h/-Ø contrast.

### *Discussion*

In Experiment 2 we found that high proficiency late L2 learners, like native speakers, exhibited significant N400 pseudoword effects for both contrasts, while low proficiency L2 speakers displayed an effect for the native /f/-Ø contrast, but not the non-native /h/-Ø contrast. These results suggest that L2 proficiency is an important factor in discriminating non-native phonological contrasts in more difficult stimulus/task contexts.

The results for the high proficiency group are noteworthy. It is unlikely that attention to the phonological difference between /h/-Ø or /f/-Ø could have facilitated discrimination – participants were not told about the

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

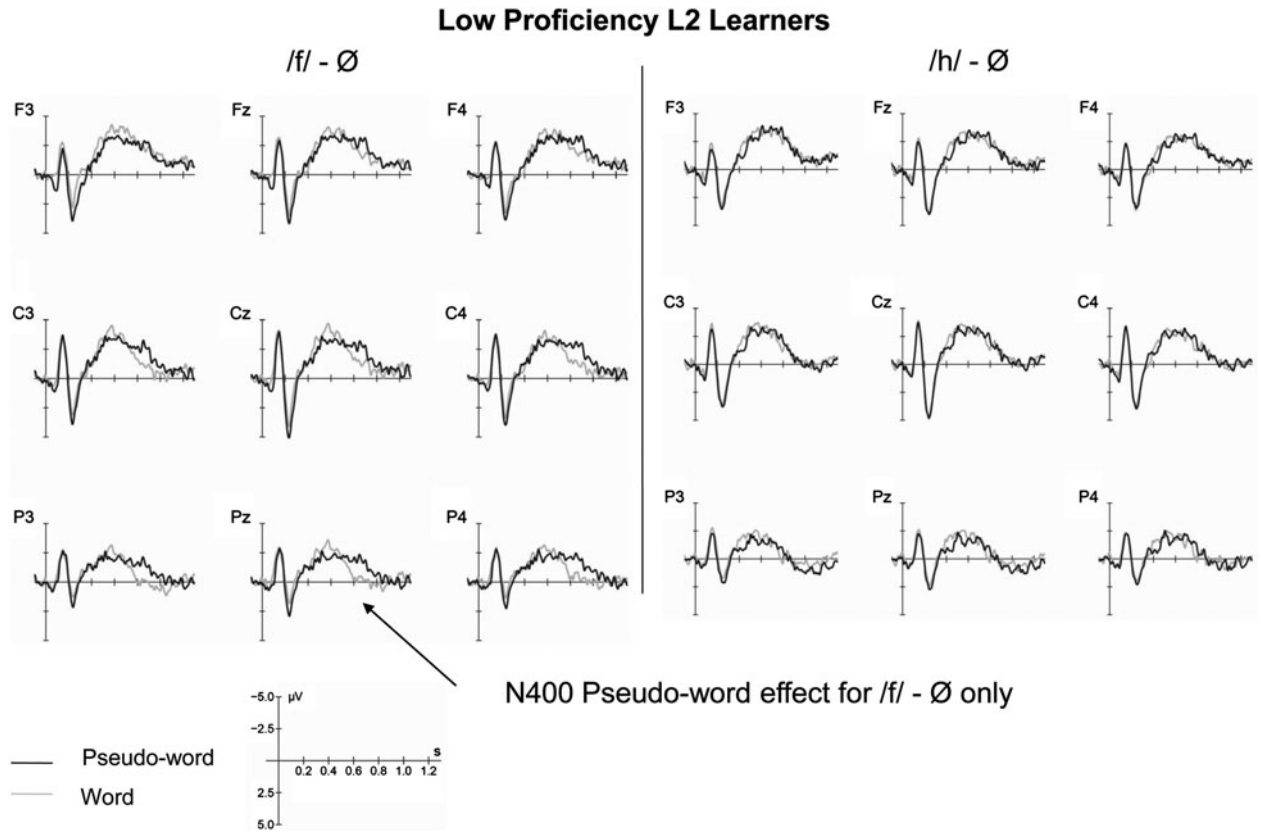



Figure 8. Grand average ERPs for the Low Proficiency L2 learners in response to words and pseudowords for the /f-Ø/ and /h-Ø/ contrasts in Experiment 2. An N400 pseudoword effect can be seen for the /f/-Ø contrast only.

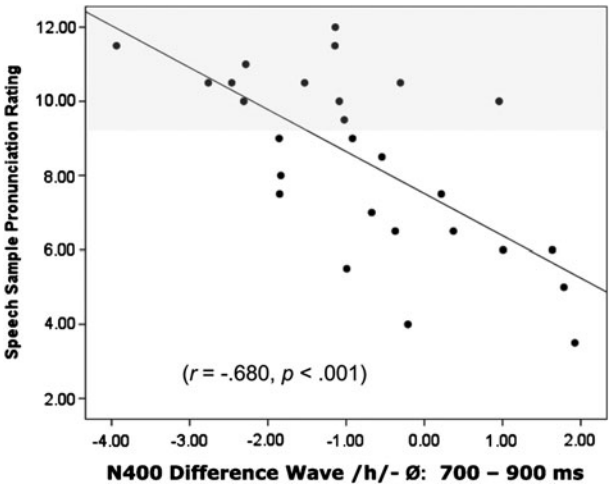



Figure 9. Correlation between L2 learners' pronunciation and the amplitude of their N400 pseudoword effect for the /h/-Ø contrast (at electrode Pz within the 700–900 ms time window).

pseudowords or contrasts and were required to listen for word meaning. Despite this, there was no difference in their response for the two contrasts, nor compared to native speakers; in all cases, a larger, longer lasting N400 was observed for pseudowords compared to words. That these effects may be related to L2 proficiency is underscored by the significant correlation between pronunciation ratings and the amplitude of the N400 pseudoword effect for the /h/-Ø contrast, suggesting that the mechanisms underlying this response become increasingly available as L2 oral proficiency improves.

The low proficiency group, in contrast, displayed a small N400 pseudoword effect for the /f/-Ø but not the /h/-Ø contrast. Could this difference be due to words in the /h/ condition being less frequent or familiar than words in the /f/ condition? Unlikely, as they were matched in word frequency. Moreover, if words in the /h/ condition were less familiar, one would expect overall increased N400s for both words and pseudowords, because unfamiliar words would have been equally difficult to retrieve and, thus, processed in a similar way as pseudowords. Instead, the ERP plots show similar N400 amplitudes for words for both contrasts and a larger effect for pseudowords for the /f/- Ø contrast only. Indeed, paired t-tests (conducted on ERPs recorded at electrode Pz within the 700–900 ms time window) revealed a significant difference between the /f/- Ø and /h/- Ø contrast pseudowords [$t(14) = 3.11$, $p < .01$], and no significant difference between /f/- Ø and /h/- Ø contrast words ($p > .10$). Thus, it appears that the

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

lack of N400 pseudoword effects for the /h/-Ø contrast reflects a true difficulty differentiating pseudowords from words. It may also suggest that for this group pseudowords such as 'appy and h-ugly are processed as words, and activate the corresponding lexical entries ('happy' and 'ugly', respectively) as successfully as the actual words. This is in contrast to the high proficiency group, for whom a larger N400 for pseudowords suggests that their lexical representations specified enough relevant phonological detail about this non-native contrast for them to successfully differentiate pseudowords from words.

Discriminating the /h/-Ø contrast during spoken word comprehension may have been difficult for the low proficiency learners for two reasons. First, as predicted by the ASP model, processing non-native phonological contrasts may be less automatic than processing native contrasts, and so attention-guided mechanisms may have been necessary for accurate integration and categorization of L2-specific phonological information. Consequently, the ability to discriminate L2-specific phonological categories suffered in Experiment 2 when attention was not focused on the phonological difference between /h/ and Ø. Second, this group displayed a rather small and delayed N400 pseudoword effect even for the native /f/-Ø contrast, suggesting that L2 words in general (regardless of whether they contain phonetic features used in the L1) may be less clearly represented and require greater and/or prolonged lexical activation compared to the high proficiency group. Arguably, the low proficiency group may have needed to focus more attention on lexical-semantic processing in order to discern word meaning, leaving even fewer resources available for phonological-level processing.

## General discussion

We investigated the role of L2 proficiency, task and stimulus effects on L2 phonological processing in two experiments. In Experiment 1, a native and non-native phonological contrast was presented as syllables during a task that required participants to attend directly to the phonological features of the stimuli. Here, similar ERP responses were observed for both contrasts in native speakers and in high and low proficiency late L2 learners. In Experiment 2, the contrasts were presented as words and pseudowords in a paradigm that directed attention away from a phonological level of analysis, in favor of semantic processing. Here, high proficiency L2 speakers, like native speakers, showed evidence of successful discrimination of both contrasts, while low proficiency L2 speakers discriminated the native, but not the non-native contrast. These results both support predictions made by the ASP model and suggest a revision to underscore the important role played by L2 proficiency and the listening context in which L2 phonological processing occurs.

Results from the low proficiency group support a central prediction of the ASP model – discriminating non-native phonological contrasts depends on stimulus/task context. While the low proficiency group successfully discriminated the non-native /h/-Ø contrast in Experiment 1 when the contrast was presented as syllables in an attended paradigm, they failed to discriminate this contrast in Experiment 2 when it was presented as words and pseudowords and when attention was directed away from phonological analysis. Consistent with the ASP model, L2 phonological discrimination by low proficiency L2 learners may require a listening situation that allows learners to focus their attention on relevant phonological information.

The results of the high proficiency group, however, were not predicted by the ASP model. Like the native speakers, these learners showed evidence of successful discrimination even when attentional resources were directed to other levels of processing and phonological analysis was left to automatic processing routines. This suggests that a qualification should be made to the ASP model's claim that "phonetic perception of non-native contrasts may never become as automatic and robust as perception of native contrasts" (Strange & Shafer, 2008, pp. 185) – rather, automatic perception appears possible for high proficiency L2 learners, even for some difficult L2 contrasts. Our findings are consistent with previous ERP studies of L2 grammar processing that report native-like automatic processing in high, but not low, proficiency late L2 learners (e.g., Steinhauer et al., 2009). Our results are also consistent with speech motor learning studies, which report a link between speech perception and production following auditory feedback training (Shiller, Sata, Broacco & Baum, 2009). Together this suggests that the ASP model should be extended to account for the possibility of native-like automatic L2 phonological processing by high proficiency L2 learners, even in cognitively-demanding situations.

More generally, our findings do not support the notion that the neurocognitive mechanisms supporting L2 processing are qualitatively different in native speakers and late L2 learners. Despite learning English as a L2 after the hypothesized critical period, ERP responses of the high proficiency late L2 learners did not differ significantly from those of the native speakers. Noteworthy differences were instead observed between the high and low proficiency groups, as seen in Figure 9. These findings support the notion that the neurocognitive correlates of L2 phonological discrimination change as L2 proficiency improves, even in older learners, and add to the growing body of literature demonstrating changes in neurocognitive processing associated with late L2 learning (McLaughlin, Tanner, Pitkänen, Frenck-Mestre, Inoue, Valentine & Osterhout, 2010; Tanner, McLaughlin & Osterhout, 2013; White et al., 2012). Future research could examine whether there might be an upper-limit

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

to late L2 learners' ability to engage in the same neurocognitive processes used by native speakers and the listening conditions that may pose the greatest challenge due to their attentional demands. For example, would L2 learners have difficulty with dichotic listening paradigms in which they must simultaneously discriminate contrasts and ignore irrelevant background noise?

Another research direction may be to investigate the factors that allow some learners to become highly proficient and, consequently, develop more efficient processing routines. One important factor may be current L2 use. Here the high proficiency group reported nearly double the amount of daily L2 use as the low proficiency group, although groups did not differ in their English use during childhood or adolescence. This is consistent with Piske, MacKay and Flege (2001) who report that current L1-L2 use exerts a significant effect on L2 pronunciation skills, independent of the age at which the L2 was acquired. Future studies should explore whether high attainment is possible for all individuals given enough L2 use, or if it reflects underlying cognitive differences, such as the ability to maintain new phonological sequences in short-term memory (O'Brien, Segalowitz, Collentine & Freed, 2006).

A limitation of our study is that we examined only one non-native contrast (/h/-Ø), which involved a phone not used in the L1. Future studies should test whether our results generalize to other non-native contrasts, particularly those that are allophonic in the L1 and thought to be difficult for L2 learners to discriminate because both contrasting sounds are assimilated into one L1 phonological category (Best & Tyler, 2007), such as the Catalan /e/-/E/ by Spanish speakers (Sebastian-Gallés et al., 2006). Such work could reveal whether some L2 listeners can achieve automaticity with regards to the full L2 phonological inventory.

In conclusion, we found that highly proficient late L2 learners processed the non-native phonological contrast /h/-Ø using similar neurocognitive mechanisms as native speakers, even in more difficult listening contexts. In contrast, for low proficiency L2 learners, discrimination was successful only when attention was directed to the phonological characteristics of the contrasting sounds. We argue that L2 phonological processing depends on participants' level of L2 proficiency and the listening context in which processing occurs. An implication of our findings is that automatic phonological processing is possible for late L2 learners and appears to become increasingly available as L2 proficiency improves.

## References


Balota, D. A, Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B. Neely, J. H., Nelson, D. L., Simpson, G. B., & Treimna, R. (2007). English Lexicon Project. *Behavioral Research Methods, 39*(*3*), 445–459.

Best, C. T., & Tyler, M. D. (2007). Non-native 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 & Philadelphia: John Benjamins.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*(*s1*), 9–49.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.2. 35)[Computer program]. Retrieved January 13, 2009.

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. *refractoriness. International Journal of Psychophysiology*, *31*(*1*), 51–68.

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.

Flege, J. E., & Schmidt, A. M. (1995). Native speakers of Spanish show rate-dependent processing of English stop consonants. *Phonetica, 52*, 90–111.

Friedrich, C. K., Aditi, L., & Eulitz, C. (2008). Neurophysiological Evidence for Underspecified Lexical Representations: Asymmetries with Word Initial Variations. *Journal of Experimental Psychology: Human Perception and Performance, 34*(*6*), 1545–1559.

Gomes, H., Molholm, Ritter, W., Kurtzberg, D., Cowan, N., & Vaughan, H. G. Jr. (2000). Mismatch negativity in children and adults, and effects of an attended task. *Psychophysiology, 37* (*6*), 807–816.

Hisagi, M., Garrido-Nag, K., Datta, H., & Shafer, V. L. (2015) ERP indices of vowel processing in Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *18*(*2*), 271–289.

Hisagi, M., Shafer, V. L., Strange, W., & Sussman, E. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: Behavioral and electrophysiological measures. *Brian Research, 1360*, 89–105.

Hisagi, M., & Strange, W. (2011). Perception of Japanese temporally-cued contrasts by American English listeners. ille, H. J. 1990. Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Speech, 54*(*2*), 241–264.

Holcomb, P. J. Grainger, J., & O'Rouke, T. (2002). An Electrophysiological Study of the Effects of Orthographic Neighborhood Size on Printed Word Perception. *Journal of Cognitive Neuroscience, 14* (*6*), 938–950.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*(*4*), 281–312.


https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

Janda, R. D., & Auger, J. (1992). Quantitative evidence, qualitative hypercorrection, sociolinguistic variables–And French speakers' 'eadhaches with English h/Ø. *Language & Communication, 12*(*3-4*), 195–236.

John, P. (2006). Variable h-epenthesis in the interlanguage of francophone ESL learners. Unpublished master's thesis. Concordia University, Montreal, Quebec.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60–99.

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *PNAS, 103*(*30*), 11381–11386.

Kučera, H., & Francis, W. N. (1967). Computational analysis of present-day English. Providence, RI: Brown University Press.

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron*,*67*(*5*), 713–727.

LaCharité, D., & Prévost, P. (1999). The role of L1 and teaching in the acquisition of English sounds by francophones. In A. Greenhill, H. Littlefield & C. Tano (Eds.), *Proceedings of BUCLD 23* (pp. 373–385). Somerville: Cascadilla.

Lenneberg, E. H. (1967). *The biological foundations of language*. New York: Wiley.

Lau, E. F. Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience, 9*, 920–933.

Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech Perception and linguistic experience: Issues in cross-language research* (pp. 351–378). Timonium, MD: York Press.

Long, M. H. (1990). Maturational constraints on language development. *Studies in second language acquisition*, *12*(*3*), 251–285.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic space from lexical co-occurence. *Behavioral Research Methods, Instruments & Computers, 28*(*2*), 203–208.

Mah, J. (2011). Segmental representations in interlanguage grammars: the case of francophones and English /h/. Unpublished doctoral dissertation. Department of Linguisitics, McGill University.

Mah, J., Goad, H., & Steinhauer, K. (2006). The trouble with [h]: Evidence from ERPs. In: O'Brien, M.G. (ed.), *GASLA-2006 Proceedings*, 80–87. Somerville, MA, USA: Cascadilla Proceedings Project.

Mah, J., Goad, H., & Steinhauer, K. (2015). The perception of English /h/ by French speakers: evidence from ERPs. Manuscript submitted for publication.

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, *60*(*s2*), 123–150.

Molnar, M., Baum, S. Polka, L., & Steinhauer, K. (2009). Automatic auditory discrimination of vowels in simultaneous bilingual and monolingual speakers as measured by the mismatch negativity (MMN). *Journal of the Acoustical Society of America, 125*(*4*), 2754-2754.

Molnar, M., Polka, L., Baum, S., & Steinhauer, K. (2014). Learning two languages from birth shapes pre-attentive processing of vowel categories: Electrophysiological correlates of vowel discrimination in monolinguals and simultaneous bilinguals. *Bilingualism: Language and Cognition*, *17*(*3*), 526–541.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24* (*4*), 933–947.

Näätänen, R., & Gaillard, A. W. K. (1983). The orienting reflex and the N2 deflection of the event-related potential (ERP). In A. W. K. Gaillard and W. Ritter (Eds.), *Tutorials in event related potential research: Endogenous components* (pp. 119–141). Amsterdam: North Holland.

Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K., (1993). Attention and mismatch negativity. *Psychophysiology, 30*, 436–450.

Näätänen, R. Paavilainen, P. Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*, (*12*), 2544–2590.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Iimoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432–434.

Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In E. Dupoux (Ed.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler* (481-502). Cambridge, MA: The MIT Press.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus–norepinephrine system. *Psychological bulletin*, *131*(*4*), 510.

Novak, G., Ritter, W., Vaughan, H. G. Jr., & Wiznitzer, M. L. (1990). Differentiation of negative event-related potentials in an auditory discrimination task. *Electroencephalography Clinical Neurophysiology, 7*, 255–75.

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical narrative and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics, 27*(*3*), 377–402.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Endingburgh Inventory. *Neurophysiologica, 9*, 97–113.

Pakulak, E., & Neville, H. J. (2011). Maturational constraints on the recruitment of early processes for syntactic processing. *Journal of Cognitive Neuroscience*, *23*(*10*), 2752–2765.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., & Roberts, T. (2000). Audiroty cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience, 12*(*6*), 1038–1055.

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics, 29* (*2*), 191–215.

Rossi, S., Gugler, M.F., Friederici, A.D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience 18*, 2030–48.

Sebastian-Gallés, N. Rodríguez-Fornells, A. Diego-Balaguer, R., & Díaz, B. (2006). First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neurosceince, 18*, (*8*), 1277–1291.

Sharma, A., Kraus, N., McGee, T., Carrell, T., & Nicol, T. (1993). Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroencephalograph and clinical neurophysiology, 88* (*1*), 64–71.

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, *125*(*2*), 1103–1113.

Shytrov, Y., Hauk, O., & Pulvermüller, F. (2004). Distributed neuronal networks for encoding category-specific semantic information: the mismatch negativity to action words. *European Journal of Neuroscience, 19*, 1083–1092.

Steinhauer, K. (2014). Event-related potentials (ERPs) in second language research: a brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2. *Applied Linguistics*, amu028.

Steinhauer, K., White, E. J., & Drury, J.E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research, 25*(*1*), 13–41.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, doi:10.1016/j.wocn.2010.09.001

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards, & M.L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153–191). Philadelphia: John Benjamins.

Sussman, E. S. (2007). A new view on the MMN and attention debate: the role of context in processing auditory events. *Journal of Psychophysiology, 21* (*3-4*), 164–175.

Sussman, E. Kujala, T. Halmetoja, J. Lyytinen, H. Alku, P., & Naatanen, R. (2004). Automatic and controlled processing of acoustic and phonetic contrasts. *Hearing Research, 190*, 128–140.

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(*2*), 367–382.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38*(*4*), 579–590.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition, 4*, 105–122.

White, E. J., Hutka, S. A., Williams, L. J., & Moreno, S. (2013). Learning, neural plasticity and sensitive periods: implications for language acquisition, music training and transfer across the lifespan. Frontiers in systems neuroscience, 7.

White, E. J., Genesee, F., & Steinhauer, K. (2012). Brain Responses before and after Intensive Second Language Learning: Proficiency Based Changes and First Language Background Effects in Adult Learners. *PloS one*, *7*(*12*), e52318.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research, 12*, 233–265.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmonikemi, R. J., & Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*, 638–642.

Yoshida, K. A. (2004). *Shallow vs. Deep: Bilingual contrast processing*. Unpublished master's thesis. University of British Columbia, Vancouver, BC.

https://doi.org/10.1017/S1366728915000620 Published online by Cambridge University Press