Adaptation to Accents by learning from Mistakes - Event Related Potentials during error-based Phonological Learning

Ronny Bujok
Supervisors: James McQueen [1], Mirjam Broersma[2], Sybrine Bultena[2]

[1]Donders Centre for Cognition, Nijmegen, The Netherlands
[2]Cenre for Language Studies, Nijmegen, The Netherlands

If people are exposed to an unfamiliar accent, they can usually learn to understand it and adapt to its specific patterns of pronunciation very quickly. However, the mechanisms involved in this learning process are still not fully understood. In this study we suggest that such phonological learning is based on external feedback processing and internal monitoring. When we make errors in our phonological perception and receive feedback, we process it and feed it into our internal monitoring, which should become better at detecting errors itself over time, and lead to better performance. We tested a group of Dutch native speakers in a novel accent learning task. Participants were presented with accented forms of Dutch words, pronounced according to an artificially created, novel accent. From two visually presented words, which differed from one another and the accented word only in their vowel, participants had to choose the correct word that corresponded to the accented word they had heard. They received corrective feedback on their choice to learn the mapping of the phonemes of the novel accent. We measured the electrophysiological activity (EEG) and analyzed response-locked (ERN & Pe) and feedback-locked (FRN & P300) data to investigate the mechanisms of this learning process. Participants adapted to the accent and they showed internal monitoring and feedback processing throughout the entire task. Internal monitoring was already present in the beginning of the experiment and no change over time could be found, indicating that learning had already taken place in the very beginning of the experiment and thus that phonological learning is very quick. The results suggest that phonological learning can indeed be described in terms of development of specific internal monitoring processes based on external feedback on erroneous responses.

Keywords: phonological learning, accent adaptation, error related negativity, feedback related negativity

Corresponding author: Ronny Bujok, E-mail: bujokronny@gmail.com

1. Introduction

Accents and dialects add a lot of diversity to the sound of a language. And while this diversity might be interesting or even beautiful to some, it can also lead to occasional misunderstandings. We need to adapt to accents first before we can understand them perfectly. But how exactly this adaption works is largely unknown. The purpose of this study is thus to shed light on the mechanisms that enable us to adapt to novel accents.

While it might not always be obvious to us at first speakers can vary in a multitude of ways, for example in their pitch, intonation or speaker rate (e.g. van Bezooijen, 1995; Gibbon, 1998; Jacewicz, O'Neill, & Salmons, 2009). But arguably one of the biggest parts of inter-speaker variability comes from our accents and dialects. The same word can sound very different depending on who is saying it. In the USA alone the differences in pronunciation of a single word can be very drastic depending on the regional dialect. In Midwestern American English the word "dead" is pronounced as "ded", while in Southern American English, and Northern American English it is pronounced as "dayed" and "dad" respectively (Labov, as cited by Maye, Aslin, & Tanenhaus, 2008). And if we look at non-native speech the variability becomes even larger, more exaggerated (Nygaard & Pisoni, 1998), and can, therefore, be very detrimental for speech comprehension.

Accented speech has been shown to lead to more word misidentifications (Lane, 1963), lower intelligibility (Wijngaarden, 2001) and slower processing speed (Munro & Derwig, 1995). However, it has also been shown that the processing costs are also modulated by the familiarity with an accent. Participants showed slower processing speed and made more errors in a sentence verification task only when the sentences were read in an accent, they were unfamiliar with (Adank, Evans, Stuart-Smith, & Scott, 2009). The evidence for increased processing costs for unfamiliar but not for familiar accents clearly demonstrates that experience with accented speech improves speech perception for that accent. In fact, research has shown that only limited experience is necessary and that adaptation to accented speech happens very rapidly. One study, which specifically aimed to quantify how much experience is necessary to adapt to a novel accent found very fast adaptation (Clarke & Garrett, 2004). Participants heard accented sentences and had to identify the final word in a word identification task. Their processing speed was measured as an indicator of their degree of adaptation. While they initially showed slower processing speed for accented speech, the difference disappeared after less than one minute of exposure. In some cases, it even happened after exposure to only two to four sentences.

Relatively fast adaptation also occurs when merely passively listening to accented speech. In one study participants listened to a 20-minute story read in a novel, artificially created accent, and then had to perform a lexical decision task (Maye et al., 2008). In the lexical decision task participants judged the phonetically altered words as acceptable representations of the correct words, indicating adaptation. Moreover, this adaptation generalized to previously unheard words. However, this study does not show what mechanisms led to the learning.

Some researchers have proposed the idea that top-down knowledge can help us normalize and understand the meaning of someone's speech (Davis et al., 2005; Maye et al., 2008; Norris, McQueen, & Cutler, 2003). More specifically, the perceptual learning of the phonemes could have been guided by the lexical knowledge. For example, in the study by Maye et al. (2008) participants listened to a story that was clearly about a witch. Although "witch" was pronounced as "wetch" the lexical and context information very clearly indicated that the intended word was “witch”. Based on this information participants could have linked the altered vowel to the intended pronunciation of the vowel or even retuned the perception of the accented form of the vowel.

However, lexically guided perceptual learning obviously relies on lexical content and it cannot always be resolved unambiguously. If words are presented in isolation or in a very non-restrictive context, pure lexical information would not be enough to make an unambiguous decision. For example, the word “wetch” could mean either “witch” or “watch” and without context it would be impossible to know which one was correct. If we do not know the correct answer but are forced to choose, we might need to guess. There is a chance that our guess is wrong but if we receive corrective feedback we might learn from this experience and avoid making the same error in the future. In fact, corrective feedback is often being used in second language acquisition and there is evidence that shows that it is effective in improving learners’ linguistic abilities (see Sheen & Ellis, 2011). Such error-based, feedback-guided learning could potentially be involved in phonological learning as well.

Error-based phonological learning lies at the heart of the newly proposed Learning by Error Detection model (LED) and it says that we learn phonological aspects of a language, including accents, by repeatedly making errors and learning from them (Broersma, 2015; Broersma, 2018). Often when we make an error, we receive some form of feedback. For example, when we hear someone with an unfamiliar accent, we might initially misunderstand what they are saying, for example when someone pronounces the word “witch” as /wɛtʃ/ we would understand “wetch” instead of “witch”. However, if we are visibly confused the person might explain to us which word they meant, giving us explicit, corrective feedback.

According to the LED model, we process this feedback and feed it into our internal monitoring mechanisms, which then become better at monitoring our errors over time. We basically learn to remap the phonetic input onto the correct phoneme representations. When the person says /wɛtʃ/ again and we initially, wrongfully understand “wetch”, our internal monitoring should correct our understanding of the word to “witch”, because it had learned that in that accent /ɪ/ is pronounced as /ɛ/. The LED model also predicts that external feedback processing and internal monitoring are reflected in specific neural signatures, which could thus be used to investigate the mechanisms underlying phonological learning (Broersma, 2018). The event related potential (ERP) components that are associated with internal monitoring and feedback processing are the error related negativity (ERN) and the feedback related negativity (FRN).

1.1 Internal Monitoring

When we make an error it generally indicates a failure in our performance. Therefore, an error triggers cognitive control functions that try to adapt our performance. These control functions are processes that detect the error, detect that control is needed, and implement the control. This could, for example, be done by readjusting attention or our strategies to prevent errors in the future (Gehring, Goss, Coles, Meyer, & Donchin, 1993; Gehring et al., 2011). These internal monitoring systems and cognitive control processes are generally believed to be reflected by the ERN (see Gehring et al., 2011).

The ERN is a negative ERP, peaking at around 100ms, which is time-locked to erroneous responses. It is strongest at the mid-frontocentral scalp, usually at the 10-20 electrode location FCz (Gehring, Liu, Orr, & Carp, 2011). A related but smaller response usually occurs on correct responses. That is the correct related negativity (CRN). The difference between the ERN and CRN makes the ERN effect which indicates the amount of internal error monitoring.

Initially, the ERN was found primarily in speeded decision tasks (e.g. Gehring et al, 1993; Herrmann, Römmler, Ehlis, Heidrich, & Fallgatter, 2004; Hohnsbein, Falkenstein, Hoormann, & Blanke, 1991). In these kinds of experiments errors are usually the result of time pressure and participants often become aware of their errors shortly after their response (Dambacher & Hübner, 2015). While some studies report that this recognition of errors is necessary for an ERN to appear (Hewig, Coles, Trippe, Hecht, & Miltner, 2011) others have clearly found ERNs even without conscious error awareness (O'Connell et al., 2007). The ERN is often also accompanied by an error positivity (Pe) which usually peaks between 200 and 300ms post response. Unlike the ERN, it is generally accepted that the Pe reflects internal error monitoring for errors the participant became aware of only (see Gehring et al. 2011).

Most research has focused on the ERN in relatively simple tasks, where errors were purely based on some perceptual information (see Gehring et al., 2011). But there have also been studies which demonstrated ERNs in more complex tasks where mere perception was not enough to make a correct response. In learning tasks, for example, memory retrieval is essential to respond correctly. While much fewer in number, there have been studies which demonstrated the ERN in such learning tasks, for example in face-name association learning (Hammer, Heldmann, & Münte, 2013) and, within the linguistic domain, in syntactic and lexical learning (Bultena, Danielmeier, Bekkering, & Lemhöfer, 2017; Davidson & Indefrey, 2011; Sebastian-Gallés, Rodriguez-Fornells, de Diego-Balaguer, & Díaz, 2006). This suggests that the ERN not only reflects internal monitoring processes for simple perceptual errors, but also for more complex erroneous processing and interpretation of a stimulus.

Sebastian-Gallés et al. (2006) presented the first evidence that shows that the same internal monitoring processes could also be found in tasks that involve phonological information. The study investigated the ability to perceive difficult vowel contrasts in Spanish – Catalan bilinguals. Participants were presented with words containing vowel contrasts that were either common in both languages or only in Catalan, and thus more difficult for Spanish dominant bilinguals. The results showed a lack of an ERN effect for the difficult contrast, meaning that the behavioral inability to discriminate difficult vowel contrasts resulted from the inability to perceive a difference. In contrast, the Catalan dominant bilinguals clearly showed internal monitoring as indicated by an ERN effect. The study showed that the ERN effect is in fact also present in a task involving phonological discrimination. However, it did not investigate phonological learning and if the internal monitoring develops over time as a result of corrective feedback.

1.2 Feedback Processing

If participants receive feedback informing them about their error, they can use this feedback to improve their performance in the future. The FRN is one of the most studied feedback related ERP components that is believed to reflect feedback processing. It is time-locked to the feedback presentation and occurs after about 200 – 300ms after the feedback. Its time-course and mid-frontocentral scalp distribution are very similar to the ERN and thus it has also been called the feedback ERN (Gehring et al., 2011; Luft, 2014; Nieuwenhuis, Holroyd, Mol, & Coles, 2004).

The FRN is most often associated with evaluation of expected and actual outcome (Bellebaum & Daum, 2008; Hajcak, Moser, Holroyd & Simmons, 2007; Holroyd, & Coles, 2002; San Martin, 2012). Therefore, it is often modulated by the expectancy of the feedback; more unexpected feedback leading to greater FRNs (Bellebaum & Daum, 2008; San Martin, 2012). In the context of learning this makes sense, as more unexpected feedback is more informative about what is and is not a correct response and can thus help adjust behavior.

Another feedback related ERP, which usually accompanies the FRN is the so called P300. It follows the FRN and is biggest in centro-parietal regions around the 10-20 position CPz. It usually occurs 300 – 600ms post feedback (Bultena et al., 2017; San Martin, 2012). It is associated with memory processing and attention to facilitate learning (Ernst & Steinhauser, 2012; Polich, 2012).

Together the FRN and P300 could be described as indicators of an external monitoring process that guides our learning process by re-evaluating our knowledge of the rules of a learning task in the light of new, unexpected feedback.

1.3 Neural signatures during learning

The neural signatures associated with internal monitoring (ERN & Pe) and external feedback processing (FRN & P300) change throughout learning. Before any learning takes place, only a small or no ERN and Pe are present. After learning, they increase in size. This change is believed to reflect increased internal monitoring of a person’s behavior (Bellebaum & Daum, 2008; Bultena et al., 2017). In contrast, the FRN and the P300 are bigger before learning, because, at that time external feedback is needed to make correct decisions, and gets smaller after learning (see Gehring et al., 2011). Learning can thus be characterized by shifting reliance on external feedback to reliance on internal monitoring (Bultena et al. 2017). Once the participant fully learns and can monitor their behavior, feedback becomes redundant and thus less surprising, usually leading to a decrease in the FRN (Heldmann, Rüsseler, & Münte, 2008). This typical decrease of the FRN, however, was not found in some other studies (e.g. Bultena et al., 2017; Heldmann et al., 2008; Holroyd & Coles, 2002). Some believe that this absence of a decrease in the FRN is the result of uncertainty during the task which led to the feedback remaining important even after learning had taken place (Bultena et al., 2017; Heldmann et al., 2008).

In a study on morphosyntactic learning, for example, (Bultena et al. 2017) participants had to assign correct determiners, based on gender, in their L2. No decrease of the FRN was found even after successful learning of the correct determiners, evidenced by improved behavioral accuracy and an emergent ERN effect. Because participants had to learn the determiners on a case by case basis and the correct syntactic structures in their L1 interfered with the correct structures in their L2, they might have perceived more uncertainty about the correct decisions and thus have still attributed importance to the feedback. In an absence of uncertainty, a decrease in the FRN should be observed.

In summary, we know that learning can be expressed in terms of feedback processing and internal monitoring mechanisms. Initially participants rely on external feedback to tune their internal monitoring to the rules of a task and, once successful, they finally rely on internal monitoring to guide their behavior. ERPs associated with these components can be used to demonstrate this learning mechanism. If this learning mechanism also applies to phonological learning is still unclear.

First evidence of internal monitoring, as evidenced by an ERN effect, in a lexical decision task involving manipulation of phonemes suggests that internal monitoring can indeed help in making decisions about phonology (Sebastian-Gallés et al., 2006). The presence of internal monitoring implies that it must have developed at some point before. No study to date has investigated this development in more detail.

1.3 Current study

This study aims to test the LED model and shed some light on the mechanisms of phonological learning. It aims to answer if phonological learning and accent adaptation can be explained by means of external feedback processing and internal monitoring mechanisms. In line with previous research the internal monitoring (ERN and Pe) should increase throughout the task and the external feedback processing (FRN & P300) should decrease. It is the first to investigate the error-related and feedback-related ERPs during a rule-based phonological learning task and as such it will contribute to a broader understanding of these ERP components.

In this study we tested native monolingual Dutch participants by presenting them with a novel accent. Because uncertainty about the correct response has been found to modulate the FRN, the novel accent contained vowel shifts consistent with a clear rule. Therefore, after successful learning no uncertainty should remain. The novel accent contained six artificially created downward vowel shifts with respect to the F1-F2 vowel space. These vowel shifts were applied to a list of monosyllabic Dutch words to create accented versions of these words. Recordings of the accented words were presented individually to participants. In a two-alternative forced choice task participant had to decide which word they had heard. They received corrective feedback on every trial.

During the task we measured EEG to look at the ERN and Pe, reflecting internal error monitoring, and FRN and P300, reflecting feedback processing. Because other studies in the linguistic domain have shown that these neural signatures can be used to investigate learning (e.g. Bultena et al., 2017, Hammer, Heldmann, & Münte, 2013, Heldmann et al., 2008; Holroyd & Coles, 2002), we wanted to see if it also applies to phonological learning. As such we had the following questions:

1. If there is no initial internal monitoring, can sufficient accent adaptation be developed throughout the duration of the task, leading to internal error monitoring at the end, indicated by the error related ERPs (ERN and Pe)?
2. How is corrective feedback processed and how, if at all, does feedback processing (FRN & P300) change throughout the task?

Regarding the internal monitoring mechanisms and error detection we expect no ERN and Pe effects (difference between the error and correct conditions) to be present in the beginning of the experiment. Since these error-related ERPs reflect internal error monitoring and the accent participants are presented with is novel, no internal error detection for this particular accent should take place. The initial absence of these effects has been documented in other linguistic tasks (Bultena, Danielmeier, Bekkering, & Lemhöfer, 2017; Davidson & Indefrey, 2011; Sebastian-Gallés, Rodriguez-Fornells, de Diego-Balaguer, & Díaz, 2006) and it is most likely that it is similar for phonological learning. In line with previous studies on accent learning (Adank et al., 2009; Clarke & Garrett, 2004) participants should adapt to the novel accent. As a result, the internal monitoring (ERN & Pe) should increase throughout the experiment as participants learn.

Feedback, as a relevant source of information, should be processed in the beginning of the experiment before any learning has taken place. Because the FRN is associated with the evaluation of expected vs. actual outcome (Bellebaum & Daum, 2008; Hajcak, Moser, Holroyd, & Simons, 2007; Luft, 2014; San Martin, 2012) and the P300 with memory processing and attention to facilitate learning (Ernst & Steinhauser, 2012; Polich, 2012) it should be biggest in the beginning of the experiment when the feedback is still unexpected and attention to learn is still relevant. Throughout the experiment, as the participants learn, feedback should become more expected and less relevant. Therefore, feedback processing (FRN & P300) should decrease over time. While the FRN has been found to remain the same size throughout learning tasks with uncertainty (e.g. Bultena et al., 2017; Heldmann et al., 2008), the present study uses a novel accent with rule-based vowel shifts, which should eliminate uncertainty and hence the decrease in the FRN is still expected.

2. Method

Prior to the main EEG study two behavioral pilot experiments were run. The purpose of the pilot experiments was to test if participants could learn to understand the novel accent. Moreover, since we were interested in error related ERPs, we had to ensure that the task was not too easy and hence that participants made enough errors throughout the experiment to be able to analyze and compare the ERPs at different time points in the experiment. Based on previous literature it was decided that five errors would be a required minimum, but seven or more errors would be preferred (Olvet & Hajcak, 2009). The pilot experiments led to an optimized item list and methodology. All experiments were approved by the Ethics Assessment Committee of the Radboud University Faculty of Arts.

2.1 Participants

Participants were recruited at the Radboud University via the Radboud Research Participation System. All participants were Dutch native speakers from the Netherlands under 35, who did not grow up in any of the following regions: Noord-Holland, Zuid-Holland, Flevoland, Utrecht and/or Limburg. This restriction was set because some of the regional dialects may contain vowel shifts like the ones used for this experiment and thus would not reflect novel accent learning. Furthermore, no participant reported having dyslexia, hearing problems and/or history of mental or neurological issues, and all participants had normal or corrected-to-normal vision. All participants gave their written consent to participate in this study and were rewarded with €20 or 2 Course Credit.

31 participants were tested. Seven participants had to be excluded for various reasons. Two participants did not meet the pre-screening criteria (they grew up in Limburg), three participants were excluded because of technical difficulties, one participant did not make enough errors, and one participant did not following task instructions. The remaining sample thus consisted of 24 participants (13 female) with an average age of 22 years (SD: 3.96, range: 18 – 33).

2.2 Design

The task consisted of a total of six blocks divided into three rounds. Each round consisted of two types of blocks (training block and mixed block) that were presented one after another without any explicit indication of change. After each round there was a pause. A training block contained only training items and its purpose was to familiarize the participants with the novel accent. A mixed block contained all items, training and test items, and was used to test participants. The different item types are explained in the following section.

2.3 Materials

2.3.1 Stimuli

For this experiment we used three monophthong (/i/ → /ɛ/; /ɪ/ → /ʏ/; /u/ → /ɔ/) and three diphthong vowel shifts (/eː/ → /ɛi/; /øː/ → /œy/; /oː/ → /ɑu/) to create a novel accent (see Figure 1). All 6 vowel shifts were downward shifts with respect to the F1-F2 vowel space, and all selected vowels were part of the participants’ L1 phoneme inventory. We applied the vowel shifts to a list of monosyllabic and highly frequent Dutch words. The resulting accented words were presented as auditory stimuli and, therefore, are going to be referred to as such. The original words were considered targets. They were the unaccented form of the auditory stimuli and hence the intended meaning of them. Another Dutch word, which was a minimal pair of the target and the auditory stimulus, acted as a distractor. It did not conform with any of the novel accent’s vowel shift rules, that is there was no systematic relationship between the distractor’s vowel and the pronounced vowel in the auditory stimulus. The target and distractor were always presented visually on screen. There were two different item types: training items and test items.

2.3.1.1 Training items

In training items, the vowel shift created a nonword auditory stimulus. It was a minimal pair of the target word (e.g. “zin” /zɪn/ → /zʏn/). The target and auditory stimulus were then paired with a distractor that was a minimal pair of the two (e.g. “zien” /zin/). The distractor’s vowel was always one of the following: /i/, /ɪ/, /u/, / eː/, /øː/, /oː/, and within each vowel shift set several different vowels were used to avoid systematic occurrences of any distractor vowels. Moreover, since none of the vowel shifts in the novel accent shifted to these vowels, participants could assume that they were unshifted. This made them relatively unambiguous regarding their pronunciation.

2.3.1.2 Test items

In test items, the vowel shift created another Dutch word. In half of the test items the correct Dutch representation of the auditory stimulus also acted as the distractor (e.g. “list” /lɪst/ → /lʏst/, Distractor: “lust” /lʏst/). We assumed that these items would be more difficult for participants than the training items because of the interference with their lexical knowledge. According to

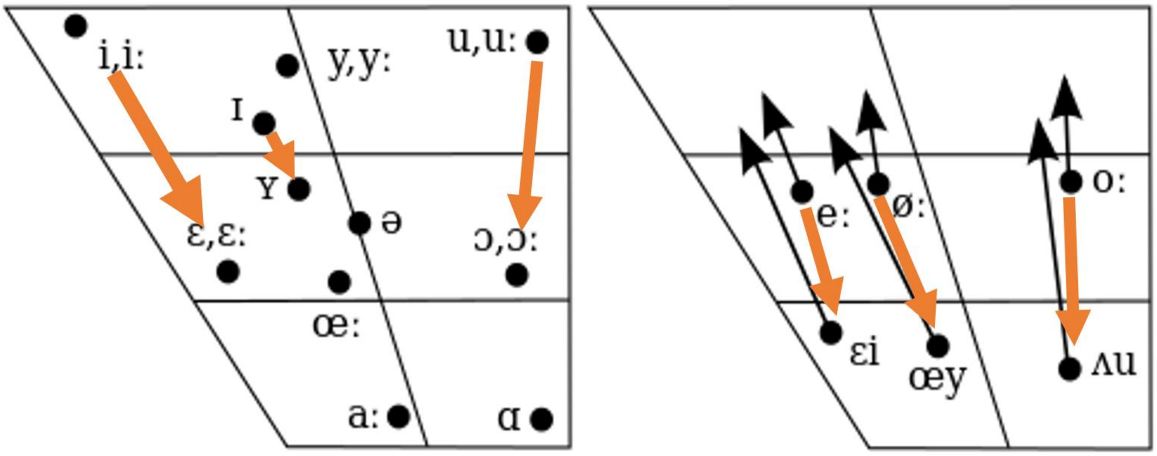

Fig 1. Overview over the vowel shifts in the novel accent. All vowels shown in this chart are part of the Dutch phoneme inventory. Left chart shows monophthongs, right chart shows diphthongs. Orange arrows indicate the direction of the shift

standard Dutch pronunciation the distractor should be the correct choice, which should make the target relatively counterintuitive. employment of the strategy “it is never the word you hear”. To mask this relationship the second half of the test items (in every vowel shift category) had a distractor with a non-identical vowel (/i/, /ɪ/, /u/, / eː/, /øː/, /oː/), which was also a word (e.g. “kist” /kɪst/ → /kʏst/, Distractor: “kiest” /kist/). The decision to change only half of the test items’ distractors was in part due to the limitation of finding enough unique minimal pairs. Moreover, we still assumed test items with distractors identical to the auditory stimulus should provoke errors in speeded decisions if they were not too frequent. It was ensured that in the non-identical distractors all distractor vowels would appear equally often.

Eight items per vowel shift per item type were chosen (see Appendix A), totaling 96 unique sets of targets and distractors. An additional six unique sets, one per vowel shift, were chosen as practice items. The words and nonwords that resulted from the vowel shifts (e.g. /lʏst/ and /zʏn/) were recorded and presented as auditory stimuli while the targets and distractors were presented visually on screen. Words occurred only once because multiple occurrences of a word in different positions and with different pairings could have interfered with the learning process.

2.3.2 Recordings

Auditory stimuli were recorded in a soundproof booth at 44,100 Hz in stereo with a Sennheiser e914 microphone. The list was read by a female Dutch native speaker. The average length of the words was 632ms (SD= 87ms, Range= 406 – 825ms).

2.4 Procedure

Participants were welcomed by the experimenter at the Centre for Language Studies Labs at Radboud University and However, as our first pilot showed, the relationship between target and distractor was relatively clear and led to the
given an information sheet for the experiment and a consent form. After they had given their written consent, they sat down in a soundproof booth and the experiment was started. The experiment was run in Presentation® software (Version 19.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) and displayed on a 24” Full HD screen (Ben Q XL 2420T) at 60Hz approximately 80cm away from the participants. Auditory stimuli were presented in stereo on high quality headphones (Sennheiser HD 215 MKII DJ). Responses were given with an in-house designed button box with three active buttons (left, right and START). They received instructions on screen to listen to the auditory stimulus and then, of two words presented visually, decide which one they had heard and respond with the corresponding button on the button box. They were explicitly asked to respond as fast as possible. They were informed that the speaker had an unusual pronunciation. First, participants were presented with the six practice items in randomized order. They were given the possibility to ask any questions regarding the task. If all questions had been answered, participants began with the main task.

The participants were presented with the stimuli in three rounds, each round consisting of a training block where all 48 training trials were shown, followed by, without explicit mentioning of a change, a mixed block consisting of the same 48 training plus 48 test trials. This resulted in a total of 432 trials across the three rounds. There were different time limits for each round: Round 1 = 1300ms, Round 2 = 1200ms, Round 3 = 1100ms. On the one hand, these time limits appeared realistic and doable based on the pilot experiments, allowing for learning to take place. On the other hand, they would force participants to respond fast and potentially lead to more errors. Corrective feedback was not

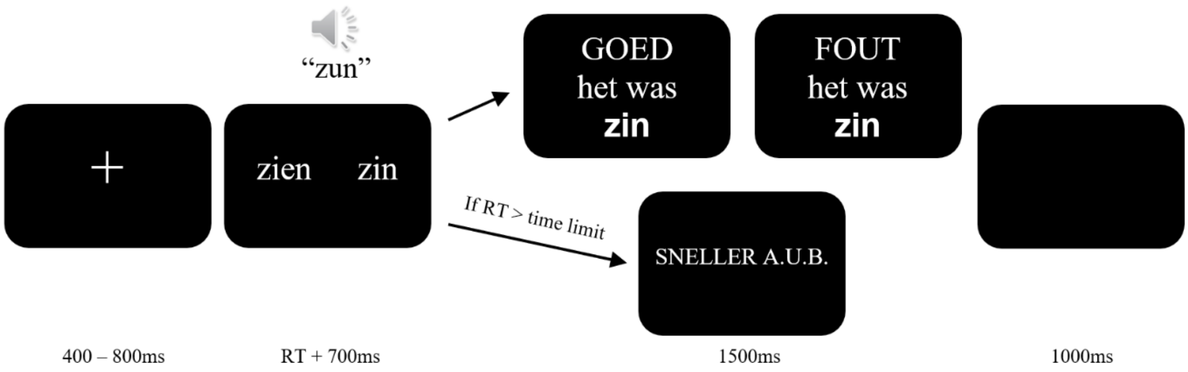

Fig 2. Example of the Trial Sequence. Positive ("GOED") or negative ("FOUT") feedback was shown depending on the participant's response. If the RT exceeded the time limit no corrective feedback but instead a reminder to respond faster ("SNELLER A.U.B.") was shown. Time limits were different for each round (Round 1: 1300ms, Round 2: 1200ms, Round 3: 1100ms)

presented on trials with reaction times greater than the time limit. Because feedback was crucial to improve performance this change encouraged faster responses.

Trials were randomized within each block, with no vowel repeating twice in a row, and an equal number of test and training items in both halves of the mixed block. Each participant received a unique randomization. After every round, participants were shown a Pause screen and could resume when they felt ready to do so. This design, of presenting a training block first in each round, was chosen to ensure that participants could learn and internalize the vowel shift more easily before being exposed to the supposedly more difficult test trials.

Each trial was constructed in the same way (see Figure 2): First a fixation cross was presented for a jittered duration between 400 and 800ms. Then the auditory stimulus and the visual stimuli (target and distractor) were presented simultaneously. Visual stimuli remained on screen until a response was given. The response was given with either the left or right button on the button box, corresponding to the left or right word on screen. After the response there was a 700ms delay before corrective feedback was presented in the middle of the screen. Feedback was presented for 1500ms. If the participants' response time was longer than the time limit, they were not given corrective feedback and instead received a visual reminder to respond faster for 1500ms. After feedback there a black screen was presented for 1000ms before the next trial. All visual stimuli were presented in white on a black background. Targets and distractors were presented in lower case Times New Roman, font size 40, at -100 and 100 pixels off center on the x-axis. Feedback ("GOED" (Engl.: correct) or "FOUT" (Engl.: incorrect)) was presented in the center of the screen in upper case Times New Roman, font size 30, with the additional information indicating which word was the target, in lower case ("het was [target]" (Engl.: it was [target])). The target was lower case Arial Bold, font size 40 to highlight its relevancy. The visual reminder to respond faster was presented in the center of the screen in upper case Times New Roman, font size 30 ("SNELLER A.U.B." (Engl.: faster please)).

After the experiment, participants filled out a language background questionnaire including open questions about the strategies they had used during the experiment, and if they had had any difficulties reading the visual stimuli and/or hearing the auditory stimuli. Finally, the participants were briefed about the purpose of the study, questions were answered by the experimenter, and the participants received their reward. The average duration of the study was ca. two hours.

2. 5 Analysis

Data were analyzed with Microsoft Excel (Microsoft Office 365 ProPlus, Version 1903) and IBM SPSS Statistics (IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.). Data were analyzed with a three-factor repeated measures ANOVA with correctness (two levels: error vs. correct response), block (six levels) and item type (two levels: training item vs. test item) as factors. Moreover, another three-factor repeated measures ANOVA with correctness (two levels: error vs. correct response), mixed block (three levels) and test item type (two levels: identical test item vs. different test item) as factors was performed. If a sphericity assumption was violated the corrected results were reported according to the Greenhouse-Geisser correction.

Two participants were excluded from the analysis because they ignored the task instructions (i.e., they ignored the feedback). Analysis was thus performed on the remaining 18 participants. For inferential statistics error rates were transformed using an empirical logit transformation (Warton & Hui, 2011) to fulfil the assumptions for ANOVAs.

2.6 EEG Data Collection and Processing

We measured the participants' brain activity with EEG while they were performing the task. The EEG signal was recorded from 26 active electrodes attached to an elastic cap according to the extended international 10-20 system (Fp1, Fp2, F3, Fz, F4, FC5, FC1, FCz, FC2, FC6, C3, Cz, C4, CP5, CP1, CPz, CP2, CP6, P7, P3, Pz, P4, P8, O1, Oz, O2). The vertical electro-oculogram (EOG) was recorded below and above the right eye, the horizontal EOG was recorded from electrodes placed on the temples close to the eyes. Two more electrodes were positioned on the participants' mastoids, with the electrode on the left mastoid functioning as the reference electrode. The ground electrode was placed at position AF3. The impedances were kept below 5 kOhm.

EEG and EOG activity were measured continuously. They were converted with a BrainAmp DC amplifier and BrainVision Recorder (Brain Products GmbH, Gilching, Germany) with a 500Hz sampling rate and 16-bit resolution.

The EEG data were preprocessed and analyzed in MATLAB (Mathworks, Natick, MA) using EEGLAB (Delorme and Makeig, 2004). The data was re-referenced to a common average of all electrodes and then two filters, a high pass filter at 0.1 Hz and a low pass filter at 30 Hz, were applied to filter out any slow drifts and high frequency artifacts. Long epochs time-locked to stimulus onset were created for each trial, containing the response and the feedback presentation. The epochs were baseline-corrected on the 200ms prior to stimulus onset. The response-locked and the feedback-locked data thus had the same baseline. Epochs with large artifacts were removed. Large artifacts were identified with the EEGLAB joint probability tool. This tool identified epochs with 5 standard deviations from the channel means as improbable data. An additional visual inspection confirmed these artifacts. On average this deleted 15.51 epochs (SD: 6.91) per participant.

Then an independent component analysis was performed on the epoched data for each participant. 26 components, based on the 26 scalp electrodes, were computed. These components were screened for any artifacts, such as eye-movements, muscle activity and heartbeats by visual inspection of the power spectrum, the activity across trials and the topography. Components that were identified as artifacts were removed from the data. On average 6.48 components were removed per participant (SD= 1.91).

The data was re-epoched into separate response-locked and feedback-locked epochs and another baseline correction was performed on the 200ms before the response or feedback onset to be closer to the effect. These epochs were averaged for each participant, per round, per incorrect vs. correct response. At least five trials per combination were required for the participant to be included in the analysis. Trough-to-peak amplitudes were computed to quantify the ERPs for the ERN and FRN (Bultena et al., 2017). The Pe and P300 were quantified by using mean amplitudes in specific time windows (San Martin, 2012; Wessel, Danielmeier, & Ullsperger, 2011). These time windows were chosen based on visual inspection of the grand average.

2.7 Data Analysis

To investigate the learning progress and the electrophysiological responses with regards to the correctness of the responses, we analyzed the error rates and RTs over rounds. The behavioral data were analyzed with a three-factor repeated measures ANOVA with correctness (two levels: error vs. correct response), round (three levels: block 1 & 2 vs. block 3 & 4 vs. block 5 & 6) and block type (two levels: training block vs. mixed block) as factors.

The response-locked and the feedback-locked ERPs were examined to investigate our main assumptions about internal error monitoring and external feedback processing. They were also analyzed with a three-factor repeated measures ANOVA with correctness (two levels: error vs. correct response for response-locked data, or negative vs. positive feedback for feedback-locked data), round (two levels: block 1 & 2 vs. block 3 & 4) and block type (two levels: training block vs. mixed block) as factors.

First a series of analyses was performed for rounds one and two only. The third round (block 5 & 6) was initially excluded due to relatively low error rates in these blocks and thus inability to compute proper average ERPs for many participants. Additionally, another series of analyses on a smaller subset of participants, who did make enough errors in the fifth and sixth block (n= 17), was performed with all three rounds included.

Any other additional analyses that deviated from this pattern are explicitly reported. In case of sphericity violation corrected values according to the Greenhouse-Geisser correction are reported. The results report only relevant significant findings.

3. Results

3.1 Behavioral Results

The results of the language background questionnaire confirmed that all participants fulfilled the requirements for participation in the experiment. Participants were also asked to describe the strategies they used during the task. 24 out of 31 participants explicitly reported that they noticed a pattern in how certain sounds were produced and tried to choose the correct word based on this pattern. Despite our efforts to counteract the strategy six participants still noticed and reported using the strategy “it is never the word you hear” among others.

Participants’ RTs decreased over rounds [$F(2,56)= 92.95$, $p< .001$, $\eta_p^2 = .769$] and a significant difference between block types was found [$F(1,28)= 9.62$, $p= .004$, $\eta_p^2 = .256$] (see Table 1). Generally, RTs in training blocks were faster than in mixed blocks. Within mixed blocks there was no item type main effect. The round effect remained significant however [$F(1.56,43.79)= 64.85$, $p< .001$, $\eta_p^2 = .689$].

Table 1. Mean Reaction Times over Blocks.

Block	Mean Reaction Times	Std. Deviation
1	928	88
2	927	88
3	807	99
4	843	76
5	761	49
6	787	61

An additional analysis of the mixed blocks with regards to the different test item types revealed a main effect [F(1,28)= 50.44, p< .001, η_p^2 = .643]. Identical test items (i.e. where the pronounced form and the distractor were identical, e.g. “list” /lɪst/ → /lʏst/, Distractor: “lust” /lʏst/) were generally reacted to faster (see Figure 3). The error rates of the participants decreased over rounds [F(2,56)= 36.3, p< .001, η_p^2 = .565] and participants generally made more errors in the mixed blocks when compared to the training blocks [F(1,28)= 5.05, p< .033, η_p^2 = .153] (see Figure 4). An analysis on mixed blocks only revealed an item type main effect whereas the error rates on test items were higher than for training items [F(1,28)= 4.8, p= .037, η_p^2 = .146]. The error rates within mixed blocks also dropped over time [F(2,56)= 39.84, p< .001, η_p^2 = .587]. No significant difference between test item types could be found with regard to errors.

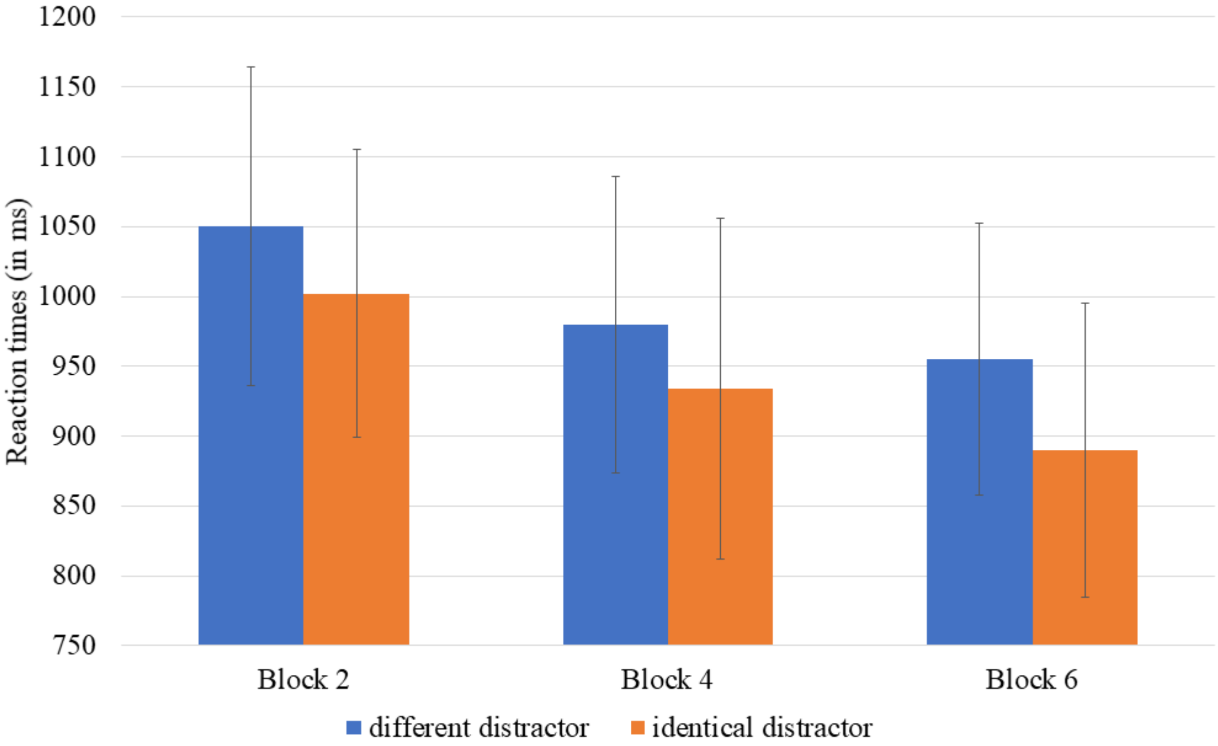

Fig 3. Comparison of RTs between the two different test item types across all mixed blocks.

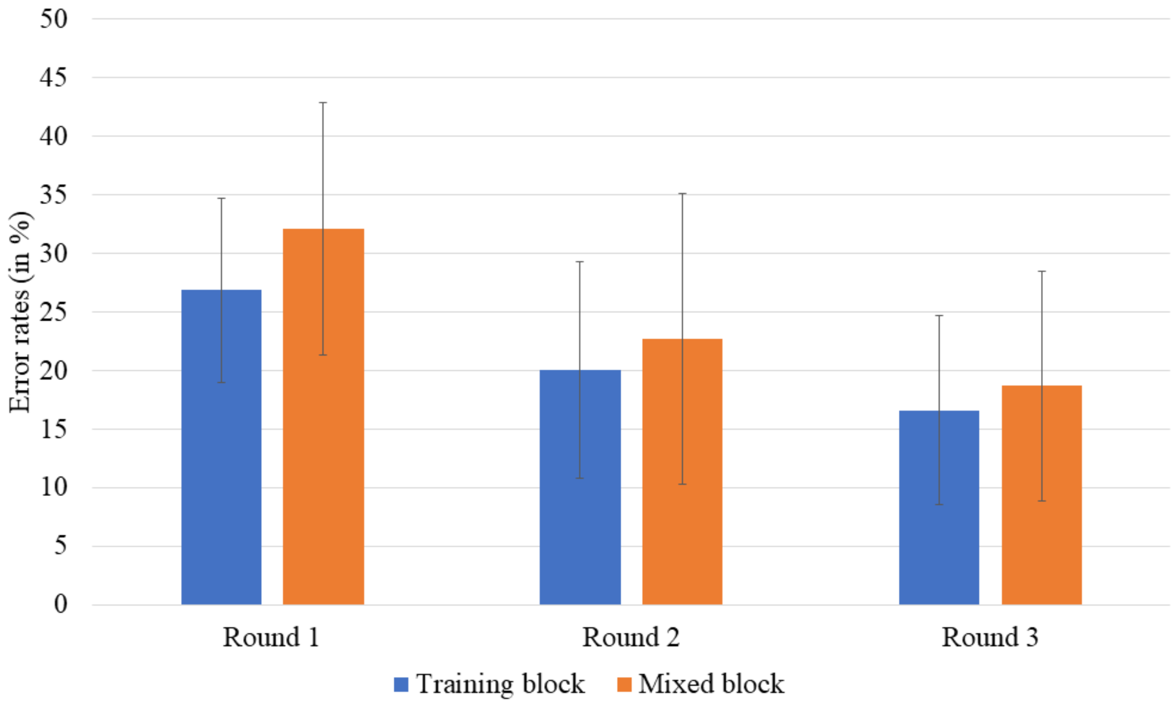

Fig 4. Mean error rates for training and mixed blocks across rounds.

3.2 Electrophysiological Results

Trough-to-peak measures were used to quantify the ERN and FRN. The peak of the ERN was identified per participant as the most negative deflection in the time window between 0ms and 100ms after the response at electrode location FCz. The trough was identified as the most positive point in the 100ms before the peak. The difference between the trough and the peak voltage created the ERN amplitude. The trough-to-peak measure is thus independent of the baseline. In a similar way the size of the FRN (feedback-locked) was also calculated with a trough-to-peak measure at electrode FCz. The peak was the most negative point between 200 and 300ms post feedback onset and the trough the most positive point 100ms prior to the peak. The difference created the FRN amplitude. Because the trough and peak latencies varied between participants, some effects might not be visible in the grand average, and therefore the mean amplitude values are presented separately (see Tables 2 and 3).

Due to the slow nature and the parietal location of the P300 the P300 was computed as the mean voltage level between 400 and 600ms post feedback onset in a cluster of electrodes around Pz (CPz, CP1, CP2, Pz, P3, P4). The time window was chosen based on visual inspection of the grand average. In a similar fashion the error positivity (PE) was computed as the mean voltage level between 200 and 300ms (see Gehring et al., 2011) after response in the electrode FCz.

Table 2. Mean trough-to-peak amplitudes in µV for response-locked data (ERN & CRN) within the sample of participants who made five or more errors in each block (n= 17). Standard deviations in parentheses.

Round	Block type	Amplitude error-related negativity		
		Error (ERN)	Correct (CRN)	Difference
1	Training	5.04 (2.42)	3.46 (1.45)	1.58
	Mixed	3.39 (2.01)	2.83 (1.58)	0.56
2	Training	4.8 (3.27)	2.65 (1.68)	2.15
	Mixed	3.92 (1.98)	2.66 (1.32)	1.26
3	Training	6.24 (3.74)	2.9 (1.83)	3.34
	Mixed	5.02 (4.30)	2.51 (1.49)	2.51

Table 3. Mean trough-to-peak amplitudes in µV for feedback-locked data (FRN) within the sample of participants who made five or more errors in each block (n= 17). Standard deviations in parentheses.

Round	Block type	Amplitude feedback-related negativity		
		Negative feedback	Positive feedback	Difference
1	Training	4.22 (1.78)	3.16 (2.16)	1.06
	Mixed	3.52 (1.86)	2.86 (2.01)	0.66
2	Training	3.7 (2.61)	2.96 (1.96)	0.74
	Mixed	3.07 (2.04)	2.59 (1.70)	0.48
3	Training	4.96 (3.17)	2.82 (2.11)	2.14
	Mixed	3.68 (2.06)	2.41 (1.87)	1.27

3.2.1 Response-locked

3.2.1.1 ERN - Rounds 1-2

Due to relatively low error rates in the final round the first analysis was performed on only the first two rounds. The analysis of the response-locked data revealed an ERN main effect (i.e. a difference between the ERN and CRN). The amplitude of the ERN was more negative than the amplitude of the CRN [F(1,23)= 25.36, p< .001, η_p^2= .524].

Block type modulated the amplitude of the ERPs with larger ERP amplitudes in training blocks [F(1,23)= 9.15, p= .006, η_p^2= .285]. There is a non-significant trend for the interaction between correctness and block type where the difference between ERN and CRN and thus the ERN effect appears to be larger in training blocks than in mixed blocks [F(1,23)= 3.78, p= .064, η_p^2= .141]. The interaction between correctness and round was not significant either, but it appears to trend towards the ERN effect being larger in the second round than in the first round [F(1,23)= 3.5, p= .074, η_p^2= .132] (see Figure 6). No difference with regards to the ERN could be found for different item types.

3.2.1.2 ERN - Rounds 1-3

The analysis on a subset of participants including round three revealed similar findings. The ERP main effect was again found to be significant with the ERN being generally larger than the CRN [F(1,18)= 24.05, p< .001, η_p^2= .572] (see Figure 5).

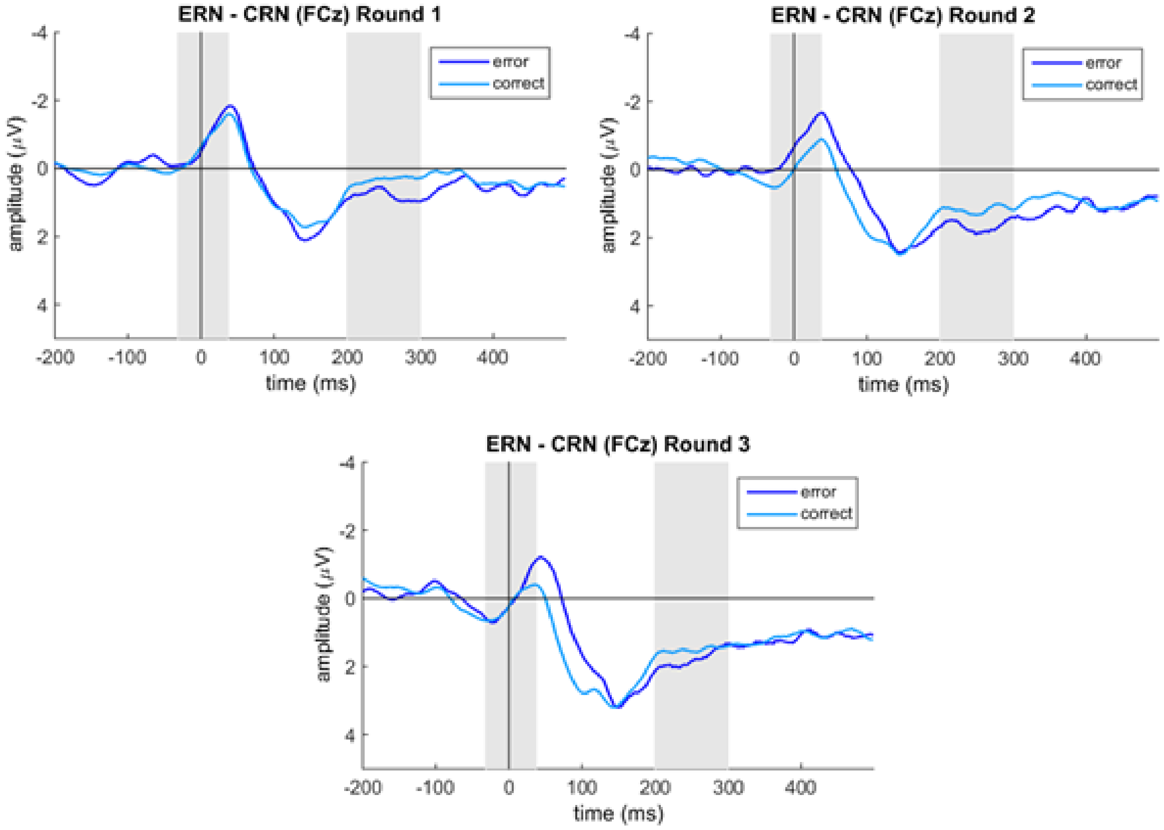

Fig 5. Response-locked data for error (ERN) and correct (CRN) responses across rounds at electrode location FCz. Grand averages for a subset of participants (n= 17). Shaded areas indicate the average latency of trough and peak (ERN) and the time window across which the PE was averaged.

Block type was significant as well [F(1,18)= 20.19, p< .001, η_p^2= .529]. The interaction between ERP and block type was significant with the ERN effect being larger in the training blocks than in the mixed blocks [F(1,18)= 8.93, p= .008, η_p^2= .331].

3.2.1.3 Pe - error positivity

For the first two rounds error positivity was generally larger for incorrect responses than for correct responses [F(1,23)= 5.28, p= .031, η_p^2= .187]. The amplitude for the ERP for both correct and incorrect responses was larger in the second round in comparison to the first round [F(1,23)= 5.931, p= .023, η_p^2= .205]. In the analyses including the third round no correctness main effect was found. The round effect, however, remained [F(2,36) = 4.06, p= .026, η_p^2= .184].

3.2.2 Feedback-locked

3.2.2.1 FRN - Round 1-2

The analysis of the FRN revealed a significant main effect of feedback [F(1,23)= 5.13, p= .033, η_p^2= .182]. Larger amplitudes were found for negative than for positive corrective feedback. There was a main effect of round, meaning that the amplitude for both correct and incorrect feedback decreased over rounds [F(1,23)= 4.39, p= .047, η_p^2= .160]. Block type was also found to modulate the amplitude of the electrophysiological responses for both kinds of feedback [F(1,23)= 5.86, p= .024, η_p^2= .203]. Larger amplitudes were found for training blocks than for mixed blocks.

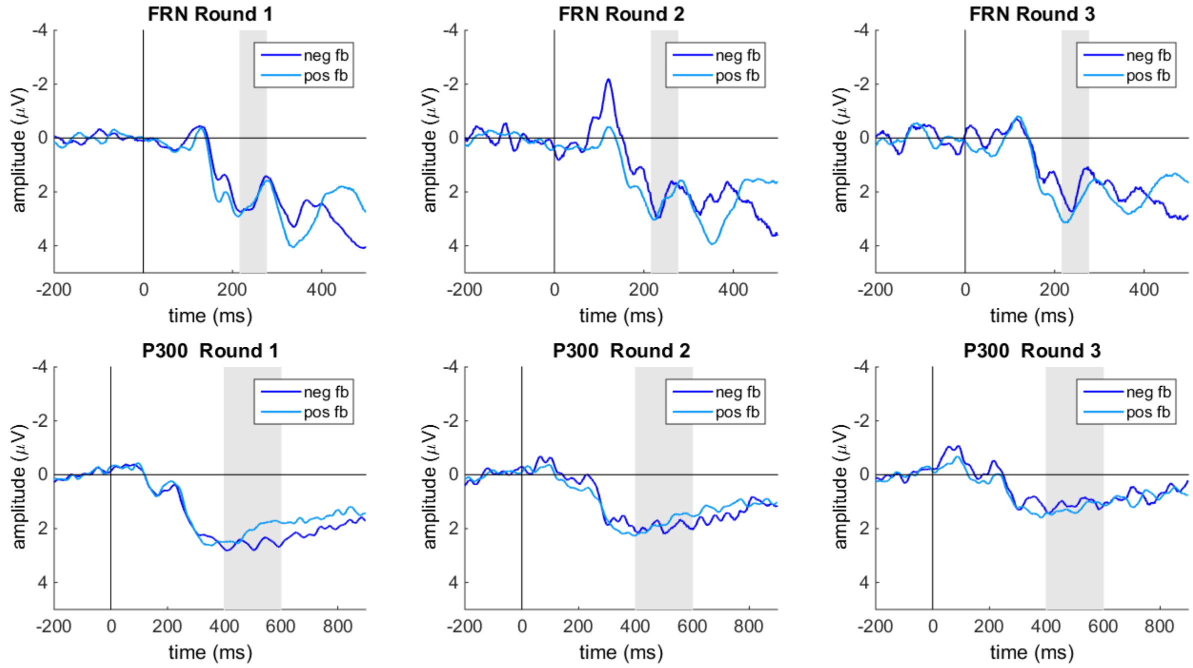

Fig 6. Feedback-locked data for positive and negative feedback across rounds at electrode location FCz. Grand averages for a subset of participants (n= 17). Shaded areas indicate the average latency of trough and peak (FRN) and the time window across which the P300 was averaged.

3.2.2.2 FRN - Rounds 1-3

The analysis of a smaller subset of participants including round 3 (see Figure 6) revealed that the main effects of feedback, [F(1,18)= 6.41, p= .021, η_p^2= .263] round, [F(2,36)= 3.91, p= .029, η_p^2= .179] and block type [F(1,18)= 6.91, p= .017, η_p^2= .277] persisted. An additional analysis of only the mixed blocks, to investigate any differences between item types, revealed no significant item type effect, but an interaction between feedback type and blocks [F(2,32)= 11.8, p< .001, η_p^2= .424].

The analysis of the P300 revealed a significant feedback main effect [F(1,23)= 18.84, p< .001, η_p^2= .450] (see Figure 6). The average amplitude of the P300 is larger for negative feedback than for positive feedback. Over rounds, the amplitude of the P300 decreases for all kinds of feedback [F(1,23)= 12.76, p= .002, η_p^2= .357]. No other significant main effects were found. The analysis of the smaller subset of participant including round 3 revealed the same main effects for feedback [F(1,18)= 21.69, p< .001, η_p^2= .546] and round [F(2,36)= 10.63, p< .001, η_p^2= .371].

4. Discussion

The main purpose of this study was to investigate the mechanisms involved in phonological learning. We examined electrophysiological activity during a novel accent learning task in native Dutch speakers. By analyzing the response-locked (ERN & Pe) and feedback-locked (FRN & P300) activity to correct vs. incorrect responses and positive vs. negative feedback across a total of six blocks, we wanted to investigate the inner workings of feedback-guided phonological learning and find evidence for external feedback processing as well as the development of internal error monitoring. The behavioral results clearly indicate that participants managed to adapt to the novel accent as seen by the decrease in error rates throughout the experiment. It is noteworthy that the participants displayed very rapid and good learning and were already better than chance in the very first block. In total their error rates dropped from 26.87% in the first block to 18.68% in the sixth block. The rapid accent adaptation observed in this study had also been demonstrated in other studies (Clarke & Garrett, 2004, Maye et al., 2008). However, the present study differed significantly in the methodology. Words were presented in isolation rather than embedded in a sentence, which could provide additional information with regards to the phonology of the target word. On the one hand this could have made learning more difficult, but on the other hand participants received very explicit corrective feedback about what the correct choice was, which could have made this task potentially easier.

The reaction time data support the idea that participants did not have much difficulty with the task, as their reaction times decreased over blocks. While they were always asked to respond as fast as possible and while a time limit was in place, encouraging fast reaction times, the actual reaction times were well below the time limit (round 1: 1300ms, round 2: 1200ms, round 3: 1100ms). The average reaction times dropped from 931ms in block 1 to 785ms in block 6.

Out of 31 participants 24 explicitly reported that they noticed a pattern in how certain sounds were produced and tried to choose the correct word based on this pattern. However, it is possible that participants used other strategies as well. Despite our exclusion of people from various regions, which might have similar vowel shifts in their accents, it is possible that our participants, who were not from these regions were still familiar with them, especially considering how just extremely limited (Clarke & Garrett, 2004) and even passive exposure (Maye et al., 2008) to accents are enough to lead to successful accent adaptation. And any familiarity with accents containing similar vowel shifts to the ones used in this experiment could have benefited the participant and sped up learning. This would mean that participants would not guess randomly but rather make educated guesses based on their prior knowledge of accents, making their performance better than chance from the very beginning.

However, prior knowledge of various accents and the use of different strategies to learn a novel accent is likely a part of natural accent adaptation and thus not a concern for this study. Most importantly, the present study eliminated alternative sources of information, such as lexical knowledge and contextual constraints, which could bring about learning that is not based on external feedback. Therefore, all these data evidently show that participants learned to understand the accent and that they used the feedback to do so.

4.1 Internal monitoring

We hypothesized that based on external feedback participants would develop internal error monitoring that guided their behavior. To investigate internal monitoring, we examined the ERN and the Pe. The results of our study clearly show larger electrophysiological responses to incorrect responses (ERN) than to correct responses (CRN), that is it shows an ERN effect. This robust effect was found consistently, for the analysis of blocks one to four as well as for the analysis on a smaller subset of participants including all six blocks. Similar findings have previously been found in another study concerning phoneme discrimination (Sebastian-Gallés et al., 2006). However, this is the first evidence of an ERN effect in a phonological learning task that involved remapping accented vowels to the correct phoneme representations.

Contrary to our hypothesis, the ERN effect did not change over blocks. If the difference between the ERN and CRN had interacted with rounds and thus increased over time it would have reflected an increase in the degree of internal monitoring as a result of more successful learning (Bellebaum & Daum, 2008;

Bultena et al., 2017). The analysis on the first two rounds seems to indicate a trend that goes in this direction, but it was not significant. The absence of this interaction and the presence of the main effect is surprising because it suggests that internal error monitoring for the novel accent is already in place at the very beginning of the experiment and that it does not change over time.

The accent was created specifically for this experiment and thus no internal error monitoring specific to this accent could have been present prior to learning. Therefore, the ERN effect had to develop at some point during the experiment. It appears that the phonological learning happened rapidly within the first block and so did the development of internal error monitoring. However, the learning might have happened too quickly to be detected by the ERN. To be able to compute the ERN a certain number of errors is required but maybe the internal monitoring developed already after only a very small number of trials, which made it impossible to detect a change in the ERN effect.

Furthermore, participants were already exposed to the novel accent before the actual main task. They had received six practice trials, one per vowel shift, including corrective feedback, which could have already been a learning opportunity. Considering how remarkably fast people have been shown to adapt to novel accents (Clarke & Garrett 2004) it appears realistic that learning could have been partially aided by the practice items.

Moreover, the novel accent contained one general vowel shift rule: all vowels were shifted downward. Normal accents usually do not show such regularity for all or several vowels. This regularity and the experimental setting might have helped some participants in learning the accent so quickly. However, this seems unlikely because to extract this general rule participants would have needed specific knowledge about F1-F2 vowel space and vowel production, that they most likely did not have. It is more likely that participants learned to associate the corresponding vowels in each vowel shift separately. Nevertheless, even under those circumstances it would have been possible to learn this quickly because the corrective feedback that participants received during this study was very direct and explicit, unlike in our everyday interactions.

Another issue in our experiment was that the strategy that “it is never the word you hear” (e.g. /lʏst/ written as “lust”; Target: “list”; Distractor: “lust”) was still reported despite our efforts to mask the relationship. On the one hand, this could be problematic because using such strategy does not reflect phonological learning. On the other hand, it is very likely that humans use several different strategies to optimize their learning. Moreover, the results show that most of the learning must have taken place in the first block, which was a training block and did not contain any items where such strategy could have been applied. Ideally, analyses on the first and second half of the first block would have been performed to reveal potential changes in the ERN effect but unfortunately too few participants made enough errors in each halves of the block to perform such an analysis.

Moreover, while the largest development might have happened in the very beginning it is possible that the internal monitoring kept increasing over blocks at a small rate. Numerically the ERN effect seems to increase over time, but maybe more power would be necessary to detect this interaction due to the small effect size of the additional increase over blocks. Testing more participants or making the task slightly more difficult, so the drop-out rate would not be as high, could be possible solutions to counteract this problem.

Generally, only very few errors were made, and some participants had to be excluded from the analysis. Some of the excluded participants, however, were actually the ones that learned the fastest and thus did not make enough errors to be included in the analysis. The electrophysiological data of these fast learners would be very insightful but with this design it was simply not possible to analyze.

One way to detect the development of internal monitoring and avoiding exclusion of many participants would be to slow down the learning process by making the task more difficult. Furthermore, introducing more uncertainty to a task leads to larger CRNs and thus to initially smaller ERN effect (Pailing & Segalowitz, 2004). In fact, contrary to our hypothesis that our task would not introduce any uncertainty because the vowel shifts were all rule-based, the pattern for the CRN actually suggests that participants felt some uncertainty with regards to the correct choice. Usually the CRN is flatter and does not peak as it does in our results. The pattern we found has previously been linked to uncertainty during the task (e.g. Bultena et al., 2017). Moreover, in the present study we had two types of blocks: training blocks, only containing training items and mixed blocks, containing training and test items. While the two item types did not differ with regards to the electrophysiological results one could argue that the mixed blocks added more uncertainty in general with regards to the correct response because they contained two types of item types. Within the analysis of block 1-4 the interaction between correctness of response and block type was not significant but trending. Within the analysis on all blocks this interaction was significant. The ERN effect was smaller in mixed blocks than in training blocks which would support the idea that the ERN effect is smaller with increased uncertainty about the correct response.

The results of the Pe analyses support the findings of the ERN analyses to some extent. At least for the first and second round the Pe was also consistently larger for incorrect responses than for correct responses. Moreover, there was no interaction with rounds, that is the difference between the electrophysiological response for the incorrect vs. correct response did not differ across rounds.

Taken together, the results from the response-locked data reveal patterns that indicate internal error monitoring during the phonological learning task. While no development of the internal monitoring could be demonstrated, further research employing a different, potentially more difficult task, other ways to measure internal monitoring, or simply testing more participants could very well find such development. Nonetheless, we can deduce that this development must have happened. Moreover, this study provides the first evidence for internal error monitoring in phonological learning and as such partially supports the LED model with regards to how we learn novel accents, that is, that we use internal monitoring to monitor our behavior in phonological learning.

4.2 Feedback processing

We hypothesized that participants would develop internal error monitoring by processing corrective feedback after their response. Feedback processing was investigated by looking at two feedback-locked components, which have been shown to be linked to feedback processing: the FRN and the P300. For the FRN specifically we hypothesized that it would decrease over time since external feedback and its processing would be redundant after successful internal error monitoring (see Gehring et al., 2011).

The results of the feedback-locked ERPs support the idea that the feedback was processed by the participants. The amplitude of the FRN and the average amplitude of the P300 were both larger for negative feedback than for positive feedback. This pattern is expected for feedback processing during a learning task since negative feedback is usually more surprising and informative and thus elicits larger ERPs. This finding was robust for both ERPs and even for the smaller subset of participants where round 3 was included into the analysis. It is evidence for feedback processing in phonological learning and strongly implies that we rely on external feedback to learn to understand a novel accent. However, no decrease in the FRN effect over time could be found. These results suggest that feedback processing happened in all rounds equally and thus that feedback remained relevant throughout the entire experiment. Despite internal error monitoring as shown by our ERN results and thus the feedback being redundant participants still processed it because they attributed relevance to it. This finding falls in line with other studies, which argue that feedback often stays relevant in tasks with a certain degree of uncertainty (Bultena et al., 2017; Heldmann et al., 2008). The error rates show that even in the last block participants did not perform perfectly and thus it is realistic to assume that participants perceived some uncertainty with regard to the correct choice.

However, this finding could in part be because of a selection bias. The participants who learned the quickest and who made the fewest mistakes had to be excluded from the ERP analyses because of a minimum requirement of errors. In contrast, the participants who made more errors throughout the experiment were included in the analyses. One could argue that the participants who kept making more errors throughout the experiment made more errors because they were less certain with regards to the correct choice and thus needed and processed the feedback more strongly. The participants who learned very quickly might not have processed the feedback as much. Unfortunately, because of the limitations of ERP analyses this study could not test this. However, if more trials had been included and/or the time pressure had been higher it would have been possible to include all participants and get a better picture of phonological learning.

In the present study the FRN effect cannot easily be seen on the grand average plots. But the amplitudes for the FRN were computed for each participant individually and thus the figure might not truly reveal the effect due to its small effect size. The grand average does reveal another small negativity just before 200ms, but its latency and time course do not resemble the FRN that has been demonstrated in previous studies (e.g., Bultena et al., 2017). The FRN usually peaks between 200 and 300ms (see Gehring, 2011), therefore this time window was chosen to extract the peak for a trough to peak measure. The trough was defined as the most positive point on the 100ms preceding the negativity.

The downside of the trough-to-peak measure is that the peak will vary per participant and in this case, it could have been anywhere between 200 and 300ms. When averaging across participants to create the grand average this information gets lost. Another problem with trough-to-peak measures is that the differences in amplitude can also arise due to differences in the trough rather than the peak (San Martin, 2012). Moreover, it has been pointed out before that it is difficult to analyze the FRN because measuring the FRN with different measures can lead to vastly different results (e.g., Bellebaum, Kobza, Thiele, & Daum, 2010). However, trough-to-peak measures have been utilized successfully in several studies (Bultena et al., 2017; Hajcak et al., 2007; Hauser et al., 2014; Holroyd, & Coles, 2002) and are thus an established way of quantifying the FRN. The fact that the FRN results also strongly overlap with the results of the feedback related P300, further supports this point.

5. Conclusions

The variability of pronunciation between speakers is incredibly large and it is especially exaggerated in accents and dialects (Nygaard & Pisoni, 1998). Therefore, the ability to adapt to the speech of different speakers is crucial for effortless communication. Not surprisingly research has shown that we can adapt, and that we do so very quickly (Maye et al., 2008; Clarke & Garrett, 2004).

One proposed mechanism that could explain accent adaptation is lexically guided perceptual learning (Norris, McQueen, & Cutler, 2003). It proposes that the perception of speech sounds is retuned as a result of lexical knowledge. For example, the nonword “wetch” could be identified as “witch” based on prior lexical knowledge. As a result, the perception of the vowel /ɛ/ could be retuned to optimize comprehension. However, without any context supporting that “witch” is the intended word, “watch” would also be a plausible option. In such cases of lexical ambiguity, lexical knowledge alone cannot explain accent adaptation. Similarly, in our study both the target and distractor were highly frequent words and thus plausible choices. Therefore, we suggest that in our study learning was guided by feedback and based on errors instead.

The results of the present study demonstrate behavioral improvement in feedback-guided phonological learning, and the electrophysiological findings show signs of internal error monitoring and feedback processing throughout the task. This evidence supports the idea that the underlying mechanism of accent adaptation could be error-based, feedback-guided phonological learning as suggested by the LED model (Broersma, 2015; Broersma, 2018). This mechanism can explain phonological learning where lexically guided perceptual learning fails and therefore it is a relevant finding for our general understanding of speech perception. However, it is possible that usually both mechanisms are working together allowing us to adapt to novel accents after just a few utterances (Clarke & Garrett, 2004).

The present study shows that internal monitoring was already present in the beginning of our experiment and contrary to our expectations the degree of internal monitoring, indicated by the difference between the ERN and CRN, did not change over time. Hence the development of internal monitoring happened too quickly to be shown. The presence of internal error monitoring to a novel accent suggests that participants learned the accent and from the presence of feedback processing we can deduce that it was contributing to that development. Unlike we expected, feedback, as evidenced by the FRN and P300, was present throughout the entire experiment. The finding suggests that feedback remained relevant to the participants despite their ability to internally monitor themselves.

All in all, this study clearly demonstrates our remarkable ability to adapt to novel accents very quickly. It is the first evidence for feedback processing and internal error monitoring in phonological learning and thus broadens our understanding of the mechanisms involved in speech perception and accent adaptation. Essentially, we can learn accents by receiving feedback and by learning from our mistakes.

References

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 520.

Bellebaum, C., & Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *European Journal of Neuroscience, 27*, 1823-1835.

Bellebaum, C., Kobza, S., Thiele, S., and Daum, I. (2010). It was not my fault: eventrelated brain potentials in active and observational learning from feedback. *Cereb. Cortex 20*, 2874–2883. doi: 10.1093/cercor/bhq038.

Broersma, M., (2015). We learn from our mistakes - or do we? Towards more efficient use of talking and listening experience in a second language. *Netherlands Organisation for Scientific Research (NWO)*, 276-89-006.

Broersma, M., (2018). Vowel soup: How second language learners learn to recognize speech sounds. *Netherlands Organisation for Scientific Research (NWO)*, 016-154-363.

Bultena, S., Danielmeier, C., Bekkering, H., & Lemhöfer, K. (2017). Electrophysiological correlates of error monitoring and feedback processing in second language learning. *Frontiers in Human Neuroscience*, *11*, 29.

Dambacher, M., & Hübner, R. (2015). Time pressure affects the efficiency of perceptual processing in decisions under conflict. *Psychological Research*, *79*, 83-94.

Davidson, D., & Indefrey, P. (2011). Error-related activity and correlates of grammatical plasticity. *Frontiers in Psychology*, *2*, 219.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*, 222.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009.

Ernst, B., & Steinhauser, M. (2012). Feedback-related brain activity predicts learning from feedback in multiple-choice testing. *Cognitive, Affective, & Behavioral Neuroscience*, *12*, 323-336.

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*, 385-390.

Gehring, W. J., Liu, Y., Orr, J. M., & Carp, J. (2012). The error-related negativity (ERN/Ne). In S. J. Luck & E. Kappenman (Eds.), *Oxford handbook of event-related potential component*s (pp. 231–291). New York, NY: Oxford University Press.

Gibbon, D. (1998). Intonation in German. *Intonation systems: A survey of twenty languages*, 78-95.

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2007). It's worse than you thought: The feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology*, *44*, 905-912.

Hammer, A., Heldmann, M., & Münte, T. F. (2013). Errorless and errorful learning of face-name associations: an electrophysiological study. *Biological Psychology*, *92*, 169-178.

Hauser, T. U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., & Brem, S. (2014). The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. *Neuroimage*, *84*, 159-168.

Heldmann, M., Rüsseler, J., & Münte, T. F. (2008). Internal and external information in error processing. *BMC Neuroscience*, *9*, 33.

Herrmann, M. J., Römmler, J., Ehlis, A. C., Heidrich, A., & Fallgatter, A. J. (2004). Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). *Cognitive Brain Research*, *20*, 294-299.

Hewig, J., Coles, M. G., Trippe, R. H., Hecht, H., & Miltner, W. H. (2011). Dissociation of Pe and ERN/Ne in the conscious recognition of an error. *Psychophysiology*, *48*, 1390-1396.

Hohnsbein, J., Falkenstein, M., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. I. Simple and choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, *78*, 438-446.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*, 679.

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, *21*, 233-256.

Luft, C. D. B. (2014). Learning from feedback: the neural mechanisms of feedback processing facilitating better performance. *Behavioural Brain Research*, *261*, 356-368.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*, 543-562.

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience & Biobehavioral Reviews*, *28*, 441-448.

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

O'Connell, R. G., Dockree, P. M., Bellgrove, M. A., Kelly, S. P., Hester, R., Garavan, H., Robertson, I.H., & Foxe, J. J. (2007). The role of cingulate cortex in the detection of errors with and without awareness: a high-density electrical mapping study. *European Journal of Neuroscience*, *25*, 2571-2579.

Olvet, D. M., & Hajcak, G. (2009). The stability of error-related brain activity with increasing trials. *Psychophysiology*, *46*, 957-961.

Pailing, P. E., & Segalowitz, S. J. (2004). The effects of uncertainty in error monitoring on associated ERPs. *Brain and Cognition*, *56*, 215-233.

Polich, J. (2012). Neuropsychology of P300. In S. J. Luck & E. Kappenman (Eds.), *Oxford handbook of event-related potential component*s (pp. 159–188). New York, NY: Oxford University Press.

San Martín, R. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Frontiers in Human Neuroscience*, *6*, 304.

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

Sheen, Y., & Ellis, R. (2011). Corrective feedback in language teaching. *Handbook of Research in Second Language Teaching and Learning*, *2*, 593-610.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, *38*, 253-265.

Warton, D. I., & Hui, F. K. (2011). The arcsine is asinine: the analysis of proportions in ecology. *Ecology*, *92*, 3-10.

Wessel, J. R., Danielmeier, C., & Ullsperger, M. (2011). Error awareness revisited: accumulation of multimodal evidence from central and autonomic nervous systems. *Journal of Cognitive Neuroscience*, *23*, 3021-3036.