The effect of explicit phonetic instruction on the perception skills of Dutch L2 users of English

Perception of the English /æ/-/ɛ/ and word-final /t/-/d/ contrasts

Ellen van Drie

Master Thesis

Research Master Linguistics and Communication Sciences

Radboud Universiteit Nijmegen

Ellen van Drie

Supervisors: Dr. Mirjam Broersma & Emily Felker

Date: 10-02-2020

Word count: 18.072 (excluding tables, references and appendices)

Abstract

Very few studies have investigated how explicit instruction can affect L2 perception of difficult non-native contrasts. This thesis investigated how explicit phonetic instruction can be beneficial to Dutch listeners' perception abilities of the English /æ/-/ɛ/ and word-final /t/-/d/ contrast. Accuracy scores on an identification task were measured in pre- and post-test before and after participants received explicit phonetic instruction through a video. Videos offered instruction on either the /æ/-/ɛ/ or the word-final /t/-/d/ contrast, and included or excluded an explanation on the differences in duration of the (preceding) vowel, which is an important cue for native listeners to distinguish these contrasts. The results showed that listeners improved on the identification task after watching a video instruction. No difference was found between videos with and without duration explanations. Additionally, asymmetric results were found for word-final /d/-/t/: word-final /d/ was often miscategorised as a /t/, and /t/ less miscategorised as /d/. Finally, metalinguistic awareness was measured, and a positive correlation was found between metalinguistic awareness and perception abilities. Results also indicated that awareness is not a prerequisite for correct perception. In conclusion, explicit phonetic instruction can lead to positive improvements on perception ability in an identification task, regardless of whether listeners learned about the duration cue. Explicit instruction thus can be a helpful tool to teach listeners to perceive the difference between difficult sounds.

Table of Contents

APPENDICES

1. Introduction

1.1 What is this thesis about?

When learning a second language (L2), a learner faces many challenges: learning new vocabulary, new syntactic structures, but also learning to produce and perceive the phonology and phonetics of the L2 accurately. This thesis focuses on L2 perception specifically. Perception is not only difficult for those trying to learn the L2, but also for more experienced listeners who are no longer actively studying or learning the L2. In this thesis, these listeners are referred to as ‘L2 users’, which is more fitting than ‘L2 learners’. Furthermore, the term ‘L2 listeners’ is used here to cover both L2 learners and L2 users, and the term ‘non-native listeners’ is used to cover both L2 listeners and naïve listeners (i.e., those who have no experience with the language whatsoever).

Particular non-native phoneme contrasts are known to be challenging for L2 listeners. Especially when the native-language (L1) phonological inventory differs from the L2 (e.g., the L2 uses phonemes that do not occur in the L1) perception can be difficult (Best, 1994; Best & Tyler, 2007). A well-known example of this is the /r/-/l/ contrast, which is difficult for Japanese listeners (Aoyama, Flege, Guion, Akehana-Yamada, & Yamada, 2004). For Dutch listeners, the English phonetic distinction between /æ/ and /ɛ/ is difficult (Díaz, Mitterer, Broersma, & Sebastián-Gallés, 2012). Furthermore, perception can also be challenging when phonotactic constraints of the L1 differ from those of the L2 (e.g., familiar L1 sounds occur in unfamiliar positions in the L2) (Broersma, 2005). For instance, the distinction between word-final /t/-/d/ in English can be problematic for Dutch listeners, because they are not familiar with a voicing distinction in word-final position (Broersma, 2005). Especially tasks that involve word processing have been shown to be hindered by perception difficulties (Broersma & Cutler, 2008; Díaz et al., 2012).

The question under study here is what could help L2 users to overcome this kind of perception difficulties. When words contain difficult sounds, how can listeners learn to identify the right words? Explicit instruction has been found helpful in other domains of second language acquisition, such as grammar learning (Svalberg, 2007). Explicit instructions can direct attention and raise awareness

(Tomlin & Villa, 1994). Indeed, there is some empirical evidence that explicit instruction on perception can direct attention and drive learning effects (Chen & Pederson, 2017; Guion & Pederson, 2007; Pederson & Guion-Anderson, 2010). However, there is only little evidence whether it could be helpful for perception difficulties that arise with non-native phonetic contrasts (Hisagi & Strange, 2011; Porretta & Tucker, 2014), and even less evidence whether it is helpful for L2 listeners instead of naïve listeners (Hisagi & Strange, 2011; Kissling, 2012; Porretta & Tucker, 2014).

This thesis investigated whether explicit phonetic instruction can help improve the perception of non-native contrasts for L2 users. Specifically, whether explicit phonetic instruction helps Dutch listeners' L2 perception of the English /æ/-/ɛ/ and word-final /t/-/d/ sound contrasts. To do this, videos were recorded with explanations on the English /æ/-/ɛ/ and word-final /t/-/d/ contrasts. This explicit instruction was designed to direct a learner's attention to a difficult phonetic contrast (either /æ/-/ɛ/ or word-final /t/-/d/), and to inform learners that they could use a duration cue to distinguish the sounds. The main research questions therefore concerned whether explicit phonetic instruction helps Dutch listeners' L2 perception of the English /æ/-/ɛ/ and word-final /t/-/d/ sound contrasts, and secondly, whether instruction on the duration cue is helpful to the listeners. Additionally, the third and fourth research questions concerned whether /ɛ/ was easier to perceive accurately than /æ/ and /t/ easier to perceive accurately than /d/, and whether learners' metalinguistic awareness relates to their perception ability.

1.2 Mechanisms of L2 perception

To answer these questions, it is important to know why and how L2 perception can be difficult. In this chapter, mechanisms and models of L2 perception are discussed (Section 1.2.1), the two contrasts that this thesis focuses on are discussed in the context of these perception mechanisms (Section 1.2.2 & Section 1.2.3), and lastly, the possibility of 'asymmetric assimilation' is discussed (Section 1.2.4).

1.2.1 L1 perception versus L2 perception

In first language research, the development of language perception in young children has been thoroughly studied. Researchers found that children are born with a language general perception: an ability to differentiate between all sounds, but as they grow older, children gradually switch to a language specific perception (Polka & Werker, 1994; Werker & Curtin, 2005; Werker & Tees, 1984). This means that children learn to differentiate between the sounds that are specifically relevant to the language they are learning, and learn to ignore irrelevant differences, i.e., gradually losing their ability to differentiate between non-native sounds. Kuhl and Iverson (1995) called this the "Native Language Magnet effect": boundaries between sounds recede, as native categories are being established through language experience. Children gradually develop language-specific phonetic mental categories (Strange & Shafer, 2008).

Speech perception involves detecting subtle differences in the acoustic signal and accessing mental phonetic categories to identify which category the sound belongs to (Strange & Shafer, 2008). When an acoustic signal comes in, it is analysed for certain acoustic parameters (e.g. the duration and spectral characteristics of a sound) which are mapped to the phonetic categories (Strange & Shafer, 2008). This is how the brain processes speech: "Perception is, by definition, an internal mental (and physiological) process by which the perceiver recognizes incoming stimulus events as instances of mental categories" (Strange & Shafer, 2008, p. 159). This categorisation process is automatic, i.e., acoustic signal properties are automatically mapped to phonetic categories. This makes the process robust and efficient for native perception, but also hard to change, which is problematic for L2 speech perception (Strange & Shafer, 2008). To be able to perceive non-native phonetic contrasts, the language-specific automatic categorisation process needs to change.

Different models have been proposed on the perception of non-native sounds. The "Native Language Magnet" (NLM) theory by Kuhl and Iverson (1995) predicts that the difficulty of a certain foreign sound depends on how close it is to a native language magnet, i.e., how similar it is to a native phonetic category. The higher the proximity, the more likely assimilation will occur. Second, Flege's "Speech Learning Model" (SLM) (Flege, 1995, 2003, 2005) predicts that a new L2 sound is most likely to form a new phonetic category when it is perceived as dissimilar to any sound of the L1. When an L2 sound is perceived as similar to an existing L1 sound, assimilation takes place and no new phonetic category is formed. Third, the “Perceptual Assimilation Model” (PAM) by Best (1994) says that L2 learners are likely to assimilate non-native speech sounds to the L1 sound category that is perceptually closest (Best, 1994). Whether assimilation happens, depends on the proximity, the ‘goodness-of-fit’ between the L2 and L1 categories (Best, McRoberts, & Goodell, 2001).

The model that leads to the clearest predictions for non-native perception is the Perceptual Assimilation Model. Although the PAM was originally designed to explain non-native speech perception by naïve listeners, Best and Tyler (2007) extended the principles of PAM to be able to explain L2 speech perception for experienced listeners: the PAM-L2 model. The PAM and PAM-L2 predict that perception of a non-native contrast is most difficult when there are two phonetic categories in the non-native language and only one phonetic category in the native language, in the same phonetic space (Best, 1994; Broersma, 2005). In this ‘single category assimilation’, a listener judges two non-native sounds as equally good or equally poor tokens of one native phoneme (Best & Tyler, 2007). There can also be a ‘category goodness’ difference, which means that both non-native sounds are perceived as tokens of one native phoneme, but one non-native sound has a better ‘goodness of fit’ to the native phoneme than the other non-native sound (Best & Tyler, 2007). Thus, the PAM and PAM-L2 predict which sound contrasts are likely to be wrongly perceived (i.e., wrongly categorised), based on the similarities and dissimilarities between the native and the non-native language phoneme categories. In this thesis, the perception of two phonetic contrasts were investigated: the /æ/-/ɛ/ contrast and the word-final /t/-/d/ contrast.

1.2.2 L2 Dutch perception of the English /æ/-/ɛ/ contrast

In the phonetic space where English has /æ/ and /ɛ/, Dutch has only one vowel, which is written with the IPA character /ɛ/ (Booij, 1995). This Dutch /ɛ/ is lower than the English /ɛ/, which means that the phonetic realisation of Dutch /ɛ/ falls in between English /æ/ and /ɛ/ (Adank, Van Hout, & Smits, 2004; Broersma, 2012; Deterding, 1997). Therefore, the PAM (Best, 1994; Best & Tyler, 2007) predicts that the two non-native sounds (English /æ/ and /ɛ/) are assimilated to the one native category (Dutch /ɛ/).

In English, the realisation of /æ/ and /ɛ/ differ in two main ways. First, they differ in vowel quality (spectral frequency): there is a difference in the frequency of the second formant (Hillenbrand, Getty, Clark, & Wheeler, 1995). Second, they differ in phonetic duration: /æ/ typically has a longer duration than /ɛ/, in isolated syllables (Black, 1949; Hillenbrand et al., 1995) as well as in connected speech (Crystal & House, 1988; Van Santen, 1992).

The longer phonetic duration of /æ/ versus the shorter phonetic duration of /ɛ/ is actively used as a cue by native English listeners when they have to identify these vowels (Hillenbrand, Clark, & Houde, 2000). Hillenbrand et al. (2000) found this significant effect of duration for the identification of /æ/ versus /ɛ/ through manipulating the durations of /hVd/ syllables (from a database by Hillenbrand et al. (1995)). The typically long /æ/ was identified as /ɛ/ when the duration was shortened, and the typically short /ɛ/ was identified as /æ/ when the duration was lengthened. This effect was not found for all vowels in the experiment by Hillenbrand et al. (2000). They conclude that, when native English listeners have to discriminate different vowels, they do not rely on their duration, except for certain specific vowels like the /æ/-/ɛ/ contrast. Similar results were previously found by Ainsworth (1972) and Bennett (1968).

Dutch listeners may be unfamiliar with the /æ/-/ɛ/ distinction, but they are familiar with the duration cue in vowels. Dutch has long-short distinctions like: /aː/-/ɑ/ and /oː/-/ɔ/. These vowels differ in quality as well as in duration characteristics, but the formants are close together (Booij, 1995; Van der Feest & Swingley, 2011). Van der Feest and Swingley (2011) say that duration is used more extensively in Dutch than in English, and give three examples. First, the difference in length between long and short vowel length is typically larger in Dutch (Booij, 1995; Van der Feest & Swingley, 2011). Secondly, duration is also used in Dutch as a feature in phonological rules: diminutives are formed using -tje or -etje, depending on the length of the vowel in the previous syllable (Booij, 1995; Van der Feest & Swingley, 2011). Third, Dutch orthography marks this distinction between long and short vowels explicitly, which children learn in school as long and short equivalents of each other: e.g.

‘maan’ (moon, /a:/) and ‘man’ (man, /ɑ/) (Van der Feest & Swingley, 2011). Dutch L2 learners are therefore probably quite sensitive to durational differences.

Indeed, Van der Feest and Swingley (2011) found evidence for Dutch listeners’ sensitivity to vowel duration: when a long vowel was shortened or a short vowel was lengthened in Dutch words, Dutch listeners tended to identify the vowels as their long or short counterparts. In contrast, English listeners tended to recognise the right vowel in English words regardless of the duration manipulations (Van der Feest & Swingley, 2011).

Díaz et al. (2012) found that Dutch listeners were able to use both frequency and duration cues to categorise the /æ/-/ɛ/ contrast. Dutch listeners more frequently identified a vowel as /æ/ when a token was manipulated to have the longer typical duration of the English /æ/, compared to the tokens that were manipulated to have the shorter typical duration of the English /ɛ/ (Díaz et al., 2012). Even though Dutch listeners scored lower than English listeners, Dutch listeners were able to use vowel duration cues to discriminate non-native vowel contrast /æ/-/ɛ/ in the discrimination task (Díaz et al., 2012).

However, Díaz et al. (2012) found that the nature of a task, especially whether the task involves lexical access or not, influences the accuracy scores of non-native listeners. Listeners’ accuracy decreased in tasks that require lexical access, such as lexical decision or word identification tasks (Díaz et al., 2012). Spoken-word recognition always involves activation of multiple word candidates, which compete with each other to select the correct word (Cutler & Broersma, 2005; McQueen, 2004). Research has shown that this word recognition can be hindered by perception difficulties in at least three ways (Broersma, 2012): difficulty with minimal pairs, more activation of competitors, and phantom competition from near-words. Indeed, previous research has found that Dutch listeners experience these three kinds of word recognition difficulties with the English /æ/-/ɛ/ contrast in tasks that require lexical access.

First, Dutch listeners responded faster to ‘kettle’ when they had heard ‘cattle’ earlier in the list, and vice versa (Cutler & Otake, 2004). This repetition effect for minimal pairs in a lexical decision task means that the word ‘kettle’ was also activated when they heard the word ‘cattle’. Second, more competitors are activated when the onset of a word contains the /æ/-/ɛ/ contrast: ‘panda’ can temporarily activate ‘pencil’ for Dutch listeners (Weber & Cutler, 2004). The end of the word resolves the ambiguity, but word recognition may be slower because of the extra activated competitors.

Third, the phantom competition from near-words has been found for the /æ/-/ɛ/ contrast by Broersma and Cutler (2011) and Broersma (2012). Broersma and Cutler (2011) found that the non-words ‘daf’ and ‘lemp’ were more often accepted as real English words by Dutch listeners an auditory lexical decision task, than by English listeners. Dutch listeners have difficulties distinguishing ‘lamp’ from ‘lemp’ and ‘deaf’ from ‘daf’ (Broersma & Cutler, 2011). Broersma and Cutler (2011) also found that when a carrier word contains a near-word, like ‘DAFfodil’, the real word ‘deaf’ was strongly activated, even after the full carrier word had been fully heard. Near-words embedded in a carrier can increase the number of lexical competitors in word recognition, which are also resistant to deactivation. Additionally, a cross-modal priming experiment showed that near-words extracted from word or phrase contexts like ‘DAFfodil’ and ‘eviL EMPire’, activated ‘deaf’ and ‘lamp’ for the Dutch, but not for the English listeners. Dutch listeners experience competition from near-words when they are embedded in a carrier sequence, which remain activated for longer than real words. The activation of near-words is a potential problem for non-native word recognition.

Broersma (2012) investigated the recognition of partially onset-overlapping word pairs like ‘DAFFOdil’-‘DEFIcit’, as well as of minimal pairs like ‘flash’-‘flesh’. Results showed that mismatching words were activated for L2 listeners more than for L1 listeners (e.g., the onset ‘daffo’ activated both ‘daffodil’ and ‘deficit’ for L2 listeners). Mismatching competitors were efficiently inhibited for L1 listeners in lexical competition, but not for L2 listeners in most cases (e.g., ‘flesh’ was inhibited for English listeners when they heard ‘flash’, but inhibition and facilitation effects were found for Dutch listeners). Therefore, she concludes that the set of activated competitors may be much larger for non-native listeners, with activation of mismatching lexical competitors and less effective inhibition of those lexical competitors (Broersma, 2012). As multiple ambiguous contrasts can occur in a single word, this may cause the number of lexical competitors to become really large very quickly

(Broersma, 2012). When many competitors are activated at the same time, and there is a lack of effective deactivation of these competitors, this may slow down the retrieval of word meaning.

1.2.3 L2 Dutch perception of the English /t/-/d/ contrast

The second non-native contrast that is studied in this thesis is the English word-final /t/-/d/ contrast. In Dutch, all voicing contrasts of consonants are neutralised in syllable-final, prepausal position – word-final obstruents are always voiceless (Booij, 1995; Broersma, 2012). Because of this, Dutch listeners are not familiar with the word-final obstruent voicing distinction that occurs in English, for instance in the word-final /t/-/d/ contrast (Broersma, 2005, 2012). English and Dutch both have the alveolar stops /t/ and /d/, but Dutch only allows /t/ in word-final position, whereas English allows both /t/ and /d/ in word-final position.

The PAM (Best, 1994; Best & Tyler, 2007) does not have specific predictions for familiar contrasts in an unfamiliar position (Broersma, 2005). Because the PAM only looks at differences in phoneme inventories to model non-native perception, it fails to incorporate the potential effects of differences in phonotactic constraints (Broersma, 2005). A contrast can exist in the native language, but it can still be difficult to distinguish the phonemes if they occur in a particular position that is not possible in the listener's native language (Broersma, 2005). Thus, although the PAM (Best, 1994; Best & Tyler, 2007) had clear predictions for Dutch listeners' perception of the English /æ/-/ɛ/ contrast, it does not have specific predictions for Dutch listeners' perception of the English word-final /t/-/d/ contrast.

The realisations of /t/ and /d/ differ in both Dutch and English in voicing: the /t/ is voiceless and the /d/ is voiced. An explanation of the cues associated with voicing can be found in Broersma (2005) (e.g., voice onset time, vocal fold vibration, closure duration). In general, there is a high degree of similarity in articulation of the /t/ and /d/ in Dutch and English, and to some extent, the cues that listeners use for voicing distinctions overlap between English listeners and Dutch listeners (Broersma, 2005). Apart from voicing cues, listeners could hypothetically attend the release burst or duration of the preceding vowel, to distinguish word-final /t/ and /d/.

Typically, /t/ has a longer and more powerful release burst than /d/ (Broersma, 2005). However, in English, the release burst is not always present Byrd (1993), and Flege and Hillenbrand (1987) found that English listeners were able to recognise the voicing of stops regardless of the presence of a release burst. Additionally, Broersma (2005) found that Dutch listeners were as good as the English listeners at categorisation of final stops in nonwords when the release bursts were removed. This does not mean that Dutch listeners may not use the burst as a cue when it is present, but it means they are able to use other cues apart from the release burst to distinguish word-final /t/ and /d/ (Broersma, 2005).

The duration of the preceding vowel has been shown to be very important for English listeners in their perception of the voicing difference for word-final obstruents, like word-final /t/-/d/ (Broersma, 2005). The duration of sounds is dependent on a combination of factors besides inherent phonetic vowel length and phonetic lengthening of vowels preceding consonants, for example: speaking rate, emphatic stress, lexical stress and word- and phrase-final lengthening (Klatt, 1976). Even so, it is well established that in English the preceding vowel typically has a longer duration when the final consonant is voiced, and a shorter duration when the final consonant is voiceless (e.g. House, 1961; House & Fairbanks, 1953). For the /t/-/d/ contrast it means that the vowel before /d/ is typically longer than the vowel before /t/.

Raphael and colleagues (Raphael, 1972; Raphael, Dorman, Freeman, & Tobin, 1975; Raphael, Dorman, & Liberman, 1980) reported that duration may be a sufficient cue for the voicing of syllable-final stops and fricatives for English listeners. However, follow-up studies with natural speech instead of synthetic speech showed that preceding vowel duration is not a *necessary* cue to the distinction between voiced and voiceless stops, as Raphael and colleagues had claimed (Hillenbrand, Ingrisano, Smith, & Flege, 1984; Hogan & Rozsypal, 1980; Wardrip-Fruin, 1982). More recently, Broersma (2005) and Broersma (2008) found that English listeners are very reliant on the duration cue. Even when it was made unreliable, English listeners continued to use it for the voicing distinction in word-final obstruents (Broersma, 2005, 2008).

Dutch listeners are not familiar with voicing contrasts in word-final position, and thus have no experience with using the preceding vowel duration as a cue for the voicing of obstruents in word-final position. However, in Dutch, Slis and Cohen (1969) found a small difference in in the length of the preceding vowel between a voiced and a voiceless obstruent in word-medial position. The duration difference is larger in English: Peterson and Lehiste (1960) reported a difference of 96 milliseconds in vowel duration before voiced and voiceless stops, whereas Slis and Cohen (1969) reported an average difference of 30 milliseconds before stops. But even though the difference in preceding vowel duration is small in Dutch, it has been shown to be used as a cue for the voicing of intervocalic two-obstruent sequences, though not as the most important cue (Van den Berg, 1989).

Thus, hypothetically, the Dutch listeners could use their language experience with Dutch voicing to perceive the word-final voicing distinction in English (Broersma, 2005). Dutch listeners are familiar with the distinction between voiced and voiceless obstruents in word-initial and word-medial positions and voiceless obstruents in word-final position. They are also sensitive to vowel duration as a cue for phoneme distinctions, and use the preceding vowel duration cue for voicing distinctions before medial-word obstruents. Especially advanced Dutch learners of English are likely to be able to combine their native and non-native language experience, to learn to distinguish the word-final voicing contrasts (Broersma, 2005).

Indeed, Broersma (2005) found that Dutch listeners are as good as English listeners at categorising word-final obstruents like stops and fricatives in nonwords. Dutch listeners have also been found to be able to use vowel duration as a cue for English final fricative voicing distinctions (Broersma, 2005, 2008). However, Broersma (2010) found a robust difference between English and Dutch listeners in their usage of vowel duration - Dutch listeners use the duration cue *less* than English listeners as a cue for final fricative voicing. Even though Dutch listeners have language experience with duration as a cue, they could not apply it effectively to their English listening (Broersma, 2010).

However, although Dutch listeners are able to use the preceding vowel duration to identify a word-final fricative voicing distinction in nonwords, they stop using it when that cue is made unreliable, while English listeners continued to use it (Broersma, 2005, 2008). English listeners were found to rely heavily on the duration cue, while Dutch listeners were more flexible in their use of cues – they did not depend on the preceding vowel duration as much (Broersma, 2005, 2008). Although Dutch listeners did not use the cues in a native-like manner, they could distinguish the sounds as accurately as, or even more accurately than natives could (Broersma, 2005). Broersma (2005) concludes that a native-like use of perceptual cues is not necessary for accurate perception.

Although Dutch listeners have been shown to be able to categorise the word-final /t/-/d/ contrast very well in non-words (Broersma, 2005), Dutch listeners fail to do so when the words are only different from real words only in this voicing feature (Broersma & Cutler, 2008). The distinction may become more difficult to perceive when lexical access is involved, similar to the studies on the /æ/-/ɛ/ contrast discussed above.

As discussed above, non-native listeners can experience difficulty in perceiving minimal pairs, experience more activation of competitors and phantom competition from near-words, which can make the activation and competition process in L2 lexical processing more difficult (Broersma, 2012). Phantom near-word activation has been found for the word-final /t/-/d/ contrast: Broersma and Cutler (2008) found that hearing non-words like 'groof' and 'flide' facilitated the recognition of 'groove' and 'flight' for Dutch listeners. The spoken non-words activated the real words for Dutch listeners, while this was not the case for native English listeners (Broersma & Cutler, 2008). Not only isolated nonwords induced these phantom word activations, but also real speech samples with the same sequence (e.g., 'biG ROOF') had the same result (Broersma & Cutler, 2011). Broersma and Cutler (2008) also found that Dutch listeners accepted spoken non-words like 'groof' and 'flide' as real English words in an auditory lexical decision task, similar to (Broersma & Cutler, 2011)'s findings on the /æ/-/ɛ/ contrast.

In conclusion, the English /æ/-/ɛ/ and word-final /t/-/d/ contrasts are both difficult to perceive for Dutch L2 listeners, not necessarily in discrimination tasks, where they score very well (Broersma, 2005; Díaz et al., 2012), but especially in tasks that require lexical access (Broersma & Cutler, 2008; Díaz et al., 2012). Both contrasts are characterised by a duration difference of the (preceding) vowel, which can be used as a cue for identification of the right sound. Dutch listeners have been found to be able to use this duration cue in tasks that do not require lexical access: the discrimination of /æ/-/ɛ/ (Díaz et al., 2012), and the categorisation of word-final fricatives (Broersma, 2005, 2008, 2010). Still, Dutch listeners do not use the duration cue as native speakers do (Broersma, 2010). This thesis investigates Dutch listeners ability to distinguish the sounds in words – an identification task that requires lexical access. Perhaps directing attention to the contrasts and explaining the duration cue through explicit instruction can be helpful for Dutch listeners to improve on this task. The question of whether this explicit instruction is helpful in L2 perception is discussed in more detail in Section 2.3.

1.2.4 Asymmetry in word recognition

With the introduction of the PAM (Best, 1994; Best & Tyler, 2007) above, the possibility of the 'category goodness' difference was briefly mentioned. Non-native sounds can be perceived as tokens of one native phoneme, but one non-native sound can have a better 'goodness of fit' to the native phoneme than the other non-native sound. Could this be the case for Dutch listeners' perception of the /æ/-/ɛ/ and word-final /t/-/d/ contrast?

There is some evidence for an asymmetry in word recognition for word that include the /æ/-/ɛ/ contrast. For instance, Weber and Cutler (2004) found that Dutch listeners experienced interference of /ɛ/ when they hear a word with /æ/, but tended not to experience interference of /æ/ when they hear a word with /ɛ/. For instance, when Dutch natives were shown a picture of a pencil and a panda, and they heard 'pan…' (the first syllable of 'panda'), they still looked at the pencil as well as the panda. But when they heard 'pen…' (the first syllable of 'pencil'), they looked only at the pencil, and there was no interference of 'panda'. Therefore, Weber and Cutler (2004) conclude that /ɛ/ is 'dominant' over /æ/ in processing. Word recognition with this contrast is asymmetric. It seems that /ɛ/ and /æ/ do not assimilate with the Dutch /ɛ/ phonetic category with the same strength.

The lexical decision and word identification tasks in Díaz et al. (2012) (those that involved lexical access) showed 'asymmetric patterns of performance' as well, for both Dutch and English listeners. The /æ/-non-words such as 'lemp' were less often misidentified as real English words than the non-words with /æ/ as in the /ɛ/-non-word ‘dask’ (Díaz et al., 2012). Additionally, when participants listened to ‘cattle’ or ‘kettle’, the /ɛ/-word interpretation (‘kettle’) was relatively more frequently chosen (Díaz et al., 2012). Thus, Dutch listeners scored better on /ɛ/-words than on /æ/-words in a word identification task: /ɛ/-words were found to ‘win’ from /æ/-words that are very similar in the lexical competition process (Díaz et al., 2012). Words containing English /ɛ/ are activated quite strongly during lexical competition (Díaz et al., 2012). The speed of lexical access has also been found to be asymmetric: the words with the dominant sound were recognised faster in both Weber and Cutler (2004) and Cutler, Weber, and Otake (2006).

Thus, there may be a difference in ‘goodness of fit’ between the /æ/ and /ɛ/ and between /t/ and /d/ contrasts. For the /æ/-/ɛ/ contrast, /ɛ/ might be dominant over /æ/, as suggested by the literature discussed above. For the word-final /t/-/d/ contrast, listeners are familiar with word-final /t/, but not with word-final /d/. Thus, participants’ language experience may give them a bias towards the /t/ response; more often perceive a voiceless /t/ than a voiced /d/ in word-final position. However, not all studies on these contrasts show asymmetries. For instance, Broersma (2005) reported that Dutch listeners did not show a bias towards voiceless responses in final position.

If there is a dominance of /ɛ/ over /æ/ in word recognition, why would /ɛ/ be dominant? One explanation is perceived proximity: Dutch /ɛ/ is reported to be perceived as being closer to English /ɛ/ than to English /æ/ (Cutler et al., 2006). This explanation is in line with the possibility of ‘goodness of fit’ difference of the PAM. However, as mentioned above, the pronunciation of Dutch /ɛ/ is in between /ɛ/ and /æ/, so that English /ɛ/ is not closer to the Dutch /ɛ/ than English /æ/ (Adank et al., 2004; Broersma, 2012; Deterding, 1997). Therefore, it is uncertain whether the English /ɛ/ does have a better fit to the Dutch /ɛ/ category than the English /æ/, or is perceived as such. An alternative explanation is that it could be an effect of orthography: in Dutch, words written with <e> are pronounced similarly to English, whereas a written <a> represents different sounds (like /ɑ/ in *pad* and /a/ in *paard*) (Cutler et al., 2006). On a lexical level, English words that contain written <e> might be classified as having a front central vowel, whereas words with a written <a> might not, and therefore listening to /æ/ and /ɛ/ would both match to words containing <e> in the lexicon (Cutler et al., 2006). A third alternative explanation comes from Diaz et al (2012), who propose that the asymmetry may be due to the higher frequency of /ɛ/: /æ/ is less frequent than /ɛ/ in English, and /æ/ does not occur in Dutch (even though the English /ɛ/ does not really occur in Dutch either, as the Dutch /ɛ/ is not equivalent to the English /ɛ/, see above).

In this thesis, the identification task accuracy scores are analysed for asymmetry for the /æ/-/ɛ/ and the word-final /t/-/d/ contrast. The /ɛ/ words may be recognised more accurately than the /æ/ words, and the word with word-final /t/ may be recognised more accurately than the words with word-final /d/.

1.3 Investigating explicit phonetic instruction as a means to improve L2 perception

The previous section discussed the mechanisms of L2 perception and problems that can occur for Dutch listeners in their perception of the English /æ/-/ɛ/ and the word-final /t/-/d/ contrasts. This chapter explores explicit instruction as a means to improve L2 perception (Section 1.3.1) and theories on the usefulness and mechanisms of explicit instruction (Section 1.3.2). Additionally, previous findings on the effects of attention direction on learning L2 perception (Section 1.3.3) and the effects of explicit instruction on the L2 perception of specific non-native contrasts (Section 1.3.4) are discussed.

1.3.1 A case for explicit instruction

Most literature that investigates how learner's perception skills can be improved use identification or discrimination tasks with feedback. For instance, identification tasks with feedback were used to teach Japanese native speakers the English /l/-/r/ contrast (e.g. Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Iverson, Hazan, & Bannister, 2005), to teach Japanese native speakers English vowels (Iverson, Pinet, & Evans, 2012) and to teach Spanish native speakers the English /i/-/ɪ/ contrast, shifting the attention of Spanish learners to spectral cues instead of duration cues (Kondaurova &

Francis, 2010). A discrimination task with feedback was also found to be successful at teaching Japanese listeners English vowels (e.g., Grenon, Sheppard, & Archibald, 2018).

These methods use feedback to make listeners aware of their misperceptions, which then encourages them to use this information to figure out the difference between sounds by themselves. Often, multiple training sessions of the same training are spread over multiple weeks (e.g., Iverson et al. (2005) did 10 sessions over the course of 2-3 weeks, and Iverson et al. (2012) 8 sessions over the course of 1-2 weeks).

Francis and Nusbaum (2002) say that identification and discrimination tasks can help shift attention to relevant cues. However, learning to improve a score on an identification or a discrimination task may not be truly helpful outside the lab: Iverson and Evans (2009) argue that, even though a high variability phonetic training can make categorization processes more efficient, such trainings may not be sufficient to cause a reweighting of phonetic cues. Additionally, Kissling (2012) argues that these methods and outcomes are hard to translate to the L2 classroom. She cautions that the results of many laboratory studies cannot immediately be equated to the effects of formal instruction. Kissling (2012) says that more research should investigate the effects of less intensive, classroom-based phonetics instruction, which has potential to be included in language classes. Raising awareness through feedback is different from (oral or written) explicit phonetic instruction about a phonetic contrast, and may yield different results. Explicitly drawing learners' attention to particular features of the L2 may be more effective than hoping they will discover the relevant acoustic features themselves (Kissling, 2012).

1.3.2 Usefulness and mechanisms of explicit instruction

When a teacher points out a certain L2 phenomenon to a learner, this can initiate a learning process. There is a long-standing debate in second language acquisition (SLA) research about whether L2 learners need or do not need explicit instruction to reach ultimate attainment level, especially with regard to L2 grammar learning. Krashen (1982) is well-known for claiming that L2 (grammar) learning should come naturally, and explicit teaching is not effective. However, there is also evidence that explicit instruction (i.e., focus on form, teaching rules and correcting errors) does lead to better learning effects than implicit learning (DeKeyser, 1994). With regard to grammar learning, L2 learners are also known to voluntarily seek out meta-linguistic insights about language, to practise with the language consciously (Ellis, 2011). Ellis (2011) argues that explicit learning is necessary when trying to overcome the processing habits of the L1, and attention and transfer from L1 stand in the way of natural, communicative learning of L2.

Indeed, explicit instruction could be useful, not only for L2 grammar, but for L2 perception as well. Explicit instructions could help learners identify exactly those characteristics of the L2 sound system that are difficult for them (Lacabex & Gallardo del Puerto, 2014, p. 502). Pederson and Guion-Anderson (2010, p. EL54) say that “it is widely believed that attention plays an important role in phonetic learning”. Reorienting of attention is required for accurate perception of at least some non-native phonemes, and explicit instruction could help learners to do that – to reorient their attention to the relevant cues in the auditory input (Pederson & Guion-Anderson, 2010). Learners need to shift their attention to the relevant acoustic cues, to focus on the specific parts of the acoustic signal that are phonologically relevant for categorisation (Francis, Baldwin, & Nusbaum, 2000). This orienting of attention to the relevant cues can lead to reweighting different cues or reweighting of different dimensions, as Francis et al. (2000) and Francis and Nusbaum (2002) call them in their Attention to Dimension model. Learning to categorise new sounds is a process of reweighting of the existing dimensions (i.e., features, cues), emphasizing or de-emphasizing different sensory dimensions (Francis & Nusbaum, 2002). Training can help to shift listeners’ attention to informative, relevant acoustic cues, instead of misleading ones (Francis et al., 2000).

To make sense of the different terms ‘attention’, ‘awareness’ and ‘noticing’, which are used in the literature mainly for grammar learning, but also relevant for perception learning, these cognitive constructs are explained here in more detail. In a review on language awareness and language learning by Svalberg (2007), and originally in Tomlin and Villa (1994), ‘attention’ is defined as a combination of ‘alertness’, ‘orientation’ and ‘detection’. Firstly, ‘alertness’ is about the readiness of a learner to process information or learn, influenced by the learner’s motivation and the learning environment

(Tomlin & Villa, 1994). Second, ‘orientation’ is about the focus of the learner’s attention at that moment - which the learner is not necessarily aware of (Tomlin & Villa, 1994). A learner can also have a higher sensitivity to a specific feature of some incoming stimulus when their orientation is directed a certain way (Tomlin & Villa, 1994).

Third, ‘detection’ is the most important part of attention: it is the cognitive registration or selection of a specific bit of information, on a specific moment of acquisition (Tomlin & Villa, 1994). Tomlin and Villa (1994) say that alertness and orientation can enhance detection – alertness and orientation are not required, but together or separately, they can increase the chance of detection. The capacity of attention is limited – it is impossible to pay attention to everything (Tomlin & Villa, 1994).

Tomlin and Villa (1994) also discuss how ‘awareness’ relates to attention. Awareness is another cognitive construct: when learners receive explicit instruction, they are made cognitively aware of some specific thing. Tomlin and Villa (1994) awareness requires attention, but attention does not require awareness: alertness, orientation and detection can take place without the learner being aware of it. Tomlin and Villa (1994) say learning can take place through attention with or without awareness, which is parallel to DeKeyser (1994)’s characterisation of implicit versus explicit learning.

Explicit instruction is a way of raising awareness, which supports attention. Explicit instruction can enhance alertness of the learner and move the orientation towards a certain linguistic phenomenon (Tomlin & Villa, 1994). Detection can also occur independently from awareness, alertness and orientation, but awareness can set the circumstances right for detection to take place (Tomlin & Villa, 1994).

In addition to the concepts attention and awareness, the concept of ‘noticing’ (Schmidt, 1990) is a related, well-known term, although mostly used for grammar learning. Noticing can be defined as ‘attention combined with awareness’ (Svalberg, 2007), or ‘detection within selective attention’ (Tomlin & Villa, 1994). Schmidt (1990) stated that acquisition of a certain aspect of L2 can only take place when it is first noticed: the Strong Noticing Hypothesis – implicit learning should be impossible. However, this view is highly debated. Tomlin and Villa (1994) argue against the strong hypothesis:

they say awareness is not a necessity for learning. In their review, Svalberg (2007) assumes that both attention and awareness can *facilitate* learning, but it does not mean that language learning cannot occur without it.

In conclusion, explicit instruction could help raise awareness, help alertness and orientation, and increase the chances of detection (Tomlin & Villa, 1994). Explicit phonetic instruction could potentially help directing learner's attention to specific non-native contrasts and phonetic cues; it could impact what listeners focus on, or are sensitive to, while listening (i.e., it influences listeners' 'orientation' which can facilitate 'detection'). The literature also suggests that a listener does not have to be aware of the difference between sounds to be able to detect the difference between them.

1.3.3 The effect of explicit instruction on attention direction and learning L2 perception

Can explicit instruction direct attention, and direct learning? There are a couple recent studies that have empirically investigated whether explicit instructions can direct attention in non-native perception. The studies by Guion and Pederson (2007), Pederson and Guion-Anderson (2010) and Chen and Pederson (2017) suggest that subjects have higher sensitivity for an aspect of listening that their attention was directed towards by instructions, and not for other aspects.

In Guion and Pederson (2007), two groups of native English listeners received different instructions: one group was instructed to pay attention to different sounds in Hindi stimuli (focus on form), whereas the second group was instructed to pay attention to the sound-meaning correspondence in the same Hindi stimuli (focus on meaning). Essentially, one group thought the aim of the experiment was to learn the sounds, whereas the other group thought the aim was to learn the meanings of the words. When both groups on discrimination ability of the non-native phonetic contrasts, the results showed that the sound-attending group performed better at the discrimination task than the meaning-attending group. So even though they had had the same exposure to the stimuli, the direction of attention was key to an improvement in discrimination of the non-native sounds.

Pederson and Guion-Anderson (2010) found a similar effect of attention direction. Two groups of English listeners were trained on their perception of Hindi words. One group was instructed to pay attention to the vowels during training, while the other group was instructed to pay attention to the consonants. While both groups practised with the same words in identification tasks with feedback, the groups scored differently on a discrimination task. The consonant-attending group had improved in discrimination ability of the consonants, while the vowel-attending group had not.

Furthermore, Chen and Pederson (2017) presented a preliminary study that showed similar patterns. In their study, native Mandarin speakers did an identification training in which one group was instructed to identify different consonants in Quanzhou Southern Min stimuli, and the other group to identify the different tones in the same stimuli. They found that the two groups performed differently on the discrimination task on tones and the discrimination task on consonants: the consonant-attending group did not improve on discriminating tones, and the tone-attending group did not improve on discriminating consonants.

Thus, in these studies, attention direction was found to be directly related to the learning effects of the listeners. These studies indicate that instructions or trainings can lead to different attention direction: focus on form versus focus on meaning (Guion & Pederson, 2007), focus on vowels versus focus on consonants (Pederson & Guion-Anderson, 2010) and focus on tones versus focus on phonemes (Chen & Pederson, 2017).

1.3.4 The effect of explicit instruction on learning to perceive L2 sound contrasts

There is very little research on how explicit phonetic instruction by a teacher might impact L2 perception skills – studies are more often focused on production, not perception (Kissling, 2012). Here, three studies are discussed that investigate how attention direction can be used to improve the perception of specific non-native sound contrasts: Hisagi and Strange (2011), Porretta and Tucker (2014) and Kissling (2012).

Hisagi and Strange (2011) tested native English listeners on their discrimination abilities regarding three Japanese contrasts that differ in duration: vowel length, consonant length and syllable number/length. For one group (study 1), the three contrasts were presented in blocks, each block preceded by detailed written instructions for each contrast, directing their attention to the relevant duration cues. A second group (study 2) were presented with the same stimuli in random order (not in blocks), without any instruction about the nature of the contrasts. The group that was given instructions on what to listen for performed significantly better than the group who were given no instructions.

Secondly, Porretta and Tucker (2014) found that native English listeners were able to improve on a discrimination task and a forced-choice identification task on the Finnish phonemic distinction for consonant length, with only an explanation that this distinction exists. Listeners' attention was drawn to the importance of a particular acoustic feature (i.e., the duration cue), with only a very basic instruction. Because of this attention, they experienced a heightened sensitivity to this particular duration cue, facilitating detection of the duration difference.

These studies suggest that attention-directing instruction can facilitate L2 perception learning for a particular non-native contrast, and listeners can profit from attention direction to particular relevant phonetic cues. However, Hisagi and Strange (2011) and Porretta and Tucker (2014) both tested only naïve listeners. Porretta and Tucker (2014) state that *at the most novice level* the information about a particular non-native contrast (as well as attention to phonetic detail, and possibly also previous language experience) can facilitate the detection of cues that are important for phonological distinctions in the L2. They say that to understand the role of explicit instruction on the processing of a phonetic contrast, it is also 'interesting and necessary' to compare these attention-orienting studies on novice learners to advanced L2-learners (Porretta & Tucker, 2014).

Kissling (2012) is, to my knowledge, the only source of data on the effect of explicit instruction on particular phonetic contrasts with more advanced learners. She found that her instruction for L1 Spanish learners of L2 English led to overall significant improvement in English sound perception (although this was not significant for all sounds individually). This instruction included explanation of grapheme-phoneme correspondences; explanation of the point, place, and manner of articulation (with an animated diagram of the vocal tract); explanation of differences in the articulation of analogous Spanish/English sounds and the phonological environments in which the sounds are produced in each language; identification activities which required learners to identify Spanish and English sounds or identify the manner of articulation; and multiple choice comprehension checks with feedback (Kissling, 2012). More evidence is necessary to show how explicit phonetic instruction could facilitate L2 perception of specific non-native phonetic contrasts for advanced learners, and whether directing attention to specific cues is helpful to them.

1.4 Research Questions & Hypotheses

The /æ/-/ɛ/ and the English word-final /t/-/d/ contrasts can be challenging to Dutch listeners, because of differences in the vowel inventory (Best, 1994; Best & Tyler, 2007) and differences in phonotactic constraints (Broersma, 2005). Especially tasks that involve lexical access have been shown to be a challenge (Broersma & Cutler, 2008; Díaz et al., 2012). And even though Dutch listeners are able to use the duration cue in the recognition of these sounds, it remains difficult to use it as fully as native speakers do (Broersma, 2010). There is a gap in the literature on how explicit phonetic instruction could be helpful to advanced listeners to improve their L2 perception of difficult phonetic contrasts. Although there is some evidence that directing attention directs language learning in general (Guion & Pederson, 2007; Pederson & Guion-Anderson, 2010), there is very little evidence on how explicit instruction on phonetic contrasts may help non-native perception (Hisagi & Strange, 2011; Porretta & Tucker, 2014), and even less evidence on how such instruction may affect advanced learners (Kissling, 2012). The present study aimed to fill this gap, using short video-instructions to teach advanced Dutch users of English about either the English /æ/-/ɛ/ contrast or the word-final /t/-/d/ contrast. The following research questions and hypotheses were formulated:

RQ 1) Does a short explicit phonetic instruction improve the perception skills of Dutch listeners of English, for the English contrasts /æ/-/ɛ/ and word-final /t/-/d/?

Explicit instruction can raise awareness, focusing learners' attention: enhancing alertness and orientation, and increase the likelihood of detection (Ellis, 2011; Tomlin & Villa, 1994). Kissling (2012) has shown that explicit instruction can lead to an improvement in the perception of a non-native contrast. Therefore it is hypothesised that the instruction may lead to improvement of their perception scores.

RQ 2) If the explicit phonetic instruction helps improve the English perception skills of Dutch listeners, how much of this effect is due to the presence of instruction on the duration cue?

The /æ/-/ɛ/ and /t/-/d/ contrasts are both characterised by a difference in the duration of the (preceding) vowel, which native speakers tend to rely on to distinguish the sounds. It might help to tell Dutch listeners to learn explicitly about the duration cue. It is hypothesised that listeners may profit from this explicit knowledge on the duration cue, direct their attention to this cue and use it to recognise the contrasts more accurately. Dutch listeners have been shown to be able to use the duration cue before for word-final fricatives (Broersma, 2005, 2008), even though they used it less by than English listeners (Broersma, 2010).

RQ 3) Is there evidence for asymmetric assimilation for the English /æ/-/ɛ/ and word-final /t/-/d/ contrasts for Dutch listeners?

There might be a 'goodness of fit' difference between /ɛ/ and /æ/, where one member of a contrast is dominant over the other in word recognition (Cutler et al., 2006; Díaz et al., 2012; Weber & Cutler, 2004). However, not all studies find an asymmetry in word recognition (Broersma, 2005). From the literature, it is therefore not clear whether any asymmetrical patterns in the accuracy on /ɛ/ versus /æ/ or /t/ versus /d/ should be visible in the identification task data.

RQ 4) Is there a correlation between Dutch L2 users of English metalinguistic awareness on the pronunciation differences between /æ/-/ɛ/ or word-final /t/-/d/ minimal pairs and the perception skills of these Dutch L2 users of English regarding the same minimal pairs?

To my knowledge, metalinguistic awareness has not been measured in this way before. Given that awareness helps detection (Tomlin & Villa, 1994), it is hypothesised that listeners who are able to indicate whether word pairs are homophones or minimal pairs also perform well at the perception task. Therefore, participant's performance on metalinguistic awareness and their performance on the perception task may be correlated. At the same time, Tomlin and Villa (1994) say that detection can take place without awareness, so participants may not have to be able to have the metalinguistic awareness to be able to detect the sounds.

2 Methodology

2.1 Participants

Sixty-four native Dutch participants (mean age: 22.3, SD: 2.89, range: 18-31, 47 female, 17 male) were tested in a sound-attenuating booth. They all gave written consent before participating, and were rewarded for their participation in the experiment with either university credits or payment in vouchers. In general, the English proficiency for this age group in the Netherlands is expected to be quite high. The participants started learning English in school from, on average, 10.67 years of age (SD: 1.19, range 8-14). They had received on average 7.47 years (SD: 1.58, range: 4-12) of English-language focused education. Participants were asked to rate their proficiency for English speaking, listening, writing and reading on a 6-point Likert scale from 0-5 (zero: no ability – five: perfect ability). Participants' self-rated listening proficiency ranged from two to five, with a mean of 3.59 (SD: 0.85). A 'mean proficiency' score was calculated for each participant, by averaging the ratings of speaking, listening, writing and reading. These scores resulted in a 'mean proficiency' average of 3.30 (SD: 0.68, range 1.5-4.75) for all participants. Participants reported listening to English (which was defined as "time you spend watching movies and TV shows in English, attending English-taught lectures, participating in English conversations, etc, in a typical week") on average 12.37 hours a week, but with a lot of variation (SD: 12.47, range 0-80 hours).

2.2 Design

The experiment followed a pre-test – training – post-test design. Participants were assigned to one of four different conditions, between-subjects, resulting in 16 participants per condition. The conditions differed in the training phase: a different phonetic instruction video for each condition. The videos explained either the /æ/-/ɛ/ contrast or the word-final /t/-/d/ contrast, with or without an explanation of the (preceding) vowel duration cue; see Table 1. The independent variables were thus: on the one hand, which contrast the video focused on (/æ/-/ɛ/ contrast or word-final /t/-/d/ contrast), and on the other hand, whether the video contained an explanation of the (preceding) vowel duration cue or not (indicated by ‘dur’ versus ‘nocue’).

***Table 1*. Video conditions: ae-dur, ae-nocue, td-dur and td-nocue**

Video conditions	Instruction on /æ/-/ɛ/ contrast	Instruction on word-final /t/-/d/ contrast
Including duration cue	ae-dur	td-dur
Excluding duration cue	ae-nocue	td-nocue

Before and after participants received instruction, they were tested on their perceptive ability of both /æ/-/ɛ/ and word-final /t/-/d/. A two-alternative forced choice (2-AFC) identification task was used for this purpose. Secondly, participants also did an offline ‘metalinguistic awareness’ task, which aimed to measure whether participants were consciously aware that /æ/-/ɛ/ and word-final /t/-/d/ minimal pairs are supposed to sound different (see Section 2.4 for procedures and the specific ordering of the tests).

Within-subjects, every participant was tested on both /æ/-/ɛ/ minimal pairs and word-final /t/-/d/ minimal pairs in both tasks, regardless of which video they were shown. This way, the items that the video was not about (e.g. the /æ/-/ɛ/ minimal pairs when a video on word-final /t/-/d/ was shown) were used as a control. The dependent variables were thus: accuracy scores on the 2-AFC identification task in pre- and post-test for /æ/-/ɛ/ items and word-final /t/-/d/ items, and accuracy scores on the metalinguistic awareness task in pre- and post-test.

2.3 Materials

2.3.1 Instruction videos

Four videos (ae-dur, ae-nocue, td-dur and td-nocue) were recorded and edited at Radboud University’s ‘Kennisclips’ studio, which specializes in the production of short informative videos that can be used for online learning in university courses. The four videos that were produced for this experiment contained explicit phonetic instruction on either the /æ/-/ɛ/ contrast or the word-final /t/-/d/ contrast, including or excluding an explanation on the duration cue, and they offer exposure to (exaggerated) minimal pairs.

A female native speaker of American English was recorded in front of a green screen as she read the scripts from an autocue. The full scripts for all four videos can be found in Appendix A. The scripts for the duration videos were parallel to the scripts for the no-cue videos, except that the sentences about duration were replaced by filler sentences. Each video was about 2 minutes long.

Visual aids were added for extra clarity and visual learning. As shown in Figures 1.1 and 1.2, the capital letters A and E (or T and D) always appeared one after another on screen the moment they were mentioned. The first time the word ‘vowel’ was mentioned, which occurred only in the td-dur video, the translation of the word was shown (i.e., Vowel = Klinker, a e i o u), because it is likely that the word is not familiar to all Dutch listeners.

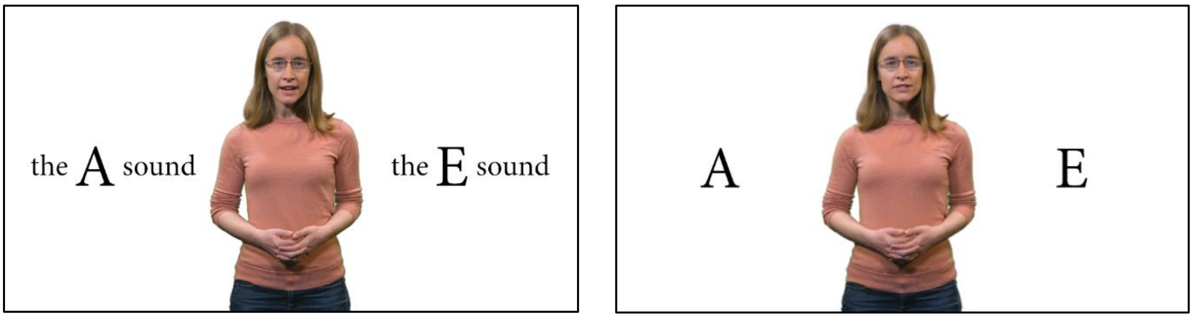

Figure 1.1 (left) and 1.2 (right). **Screenshots of one of the /æ/-/ɛ/ explanation videos. Capital letters appeared the moment the sounds were pronounced.**

The minimal pairs ‘jam-gem’ and ‘pan-pen’ were used to explain the difference between /æ/ and /ɛ/ in the ae-videos. The minimal pairs ‘not-nod’ and ‘bit-bid’ were used to explain the difference between word-final /t/ and /d/ in the td-videos. These minimal pairs were used exclusively for the instruction, and did not appear as stimuli in the pre- and post-tests. Visuals of the word pairs appeared on screen as the speaker pronounced the difference between the minimal pairs; see Figure 2.1 and 2.2 below.

***Figure 2.1 (left) and Figure 2.2 (right)*. Screenshots of an ae-video (left) and of a td- video (right).**

To explain the duration difference, the speaker exaggerated the durational difference between two members of a minimal pair (e.g. ‘jaaam-gem’). Previous research has shown that gestures may be beneficial to learning vowel length (see e.g. Hirata, Kelly, Huang, & Manansala, 2014). Therefore, the exaggerations were accompanied by a gesture to visually support the difference in duration. The gesture used in the present study was inspired by the gesture used by Van Maastricht, Hoetjes, and Van Drie (2019): the speaker starts with her palms together, then moves them apart. For the short vowel, her hands move apart only a little, while for the longer vowel, her hands move more widely apart, compare Figure 3.1 and 3.2 below.

***Figure 3.1 (left) and 3.2 (right).* Screenshots from the ae-dur video: gestures indicating vowel duration during exaggerated pronunciation of short ‘pen’ (small gesture, exaggerated short pronunciation) versus long ‘pan’ (big gesture, exaggerated long pronunciation).**

The difference between the sounds was exaggerated in the no-cue videos as well, not on duration, but on vowel quality (ae-nocue) or the burst of air for the /t/ (td-nocue). In all videos, it was mentioned that /æ/ and /ɛ/ differ in the color or quality of their sound, and the /t/ comes with a puff of air, while the /d/ does not. These other cues were briefly mentioned to not make it seem like duration was the main or only difference between the sounds.

In the ae-nocue and td-nocue videos, the sentences stating the differences in duration were omitted and replaced by filler sentences. For instance, the conclusion of the ae-dur video was: “Just remember: the A sounds longer while the E sounds shorter” (with visuals as in Figure 4, below), while for the ae-nocue video, the filler sentence was: “It will become easier to hear the difference between A and E, the more you practice listening.” (with visuals like Figure 1.2).

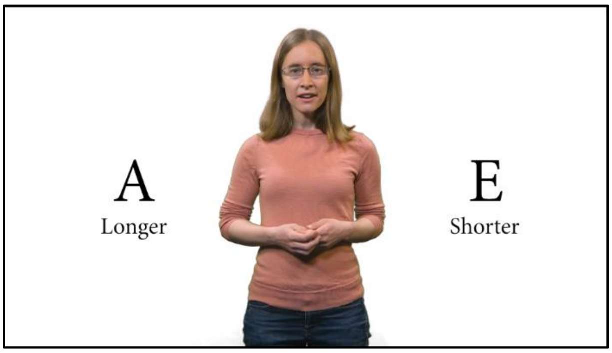

Figure 4. **Conclusion of the ae-dur video. The td-dur video was similar but with a T and D.**

2.3.2 Stimuli

For the 2-AFC identification pre-test and post-test, 40 critical minimal pairs were selected: 20 unique /æ/-/ɛ/ minimal pairs and 20 unique word-final /t/-/d/ minimal pairs. Only words that Dutch students were likely to know were included, and Dutch-English cognates were excluded as far as possible. A full word list can be found in Appendix B.

All minimal pairs were recorded by the same speaker who gave the phonetic instruction in the videos, a female speaker of American English. This way, the examples in the videos matched with her pronunciation of the stimuli. Using only one speaker increased the chances of finding a learning effect within the present study. The 80 unique words of the 40 critical minimal pairs were recorded three times in a sound-attenuating booth, in a random order, to avoid any systematic patterns occurring in prosody. They were segmented using Audacity 2.2.2 (Audacity-Team, 2018). Subsequently, one of the three recordings of each word was selected for the 2-AFC identification task. Exaggerated intonation or pronunciation was avoided (e.g., no audible schwa after word-final ‘d’). Thus, for each word, the most neutral recording was selected.

In the metalinguistic awareness task, the same 40 critical minimal pairs were used as in the 2-AFC identification task. In this task, participants were asked whether two words were pronounced the ‘same’ or ‘different’. To offer more variety in the task than only the /æ/-/ɛ/ and /t/-/d/ critical minimal pairs, 30 easier minimal pairs were added (e.g. ‘pray-play’, ‘came-game’) for which it was more obvious that the correct answer was ‘different’. Additionally, 30 homophones were added (e.g. ‘piece-peace’ and ‘soul-sole’), for which the correct answer was ‘same’. In total, the metalinguistic awareness task had 20 /æ/-/ɛ/ minimal pairs, 20 word-final /t/-/d/ minimal pairs, 30 filler minimal pairs and 30 filler homophones, so that the answers ‘same’ and ‘different’ were predicted to be well balanced. For a full list of the words, see Appendix B.

The words that were selected for each of these four categories were matched for word length; see Table 2 below. A one-way analysis of variance showed that the four groups of stimuli did not differ significantly from each other in word length ($F(3, 196) = 1.27$, $p = 0.29$).

Table 2. **Descriptives of word length for the four groups of stimuli.**

word length (no. of letters)	**critical items in ‘/æ/-/ɛ/’ group (n = 40)**	**critical items in ‘word-final /t/-/d/’ group (n = 40)**	**fillers in ‘easy minimal pairs’ group (n = 60)**	**fillers in ‘homophones’ group (n = 60)**
Mean	4.33	4.65	4.42	4.27
SD	2,07	0.64	0.76	0.78
Range	3-7	3-6	3-7	3-7

The words used in the four groups were also matched for Zipf SUBTLEX-US word frequency scores (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). A one-way analysis of variance showed that there was no significant difference for word frequency between groups ($F(196, 3) = 0.22$, $p = 0.88$). A summary of the frequency scores for the different words is presented in Table 3.

Table 3. **Descriptives of the Zipf SUBTLEX-US word-frequency scores for the four groups of stimuli.**

Zipf SUBTLEX-US word frequency	**critical items in ‘/æ/-/ɛ/’ group (n = 40)**	**critical items in ‘word-final /t/-/d/’ group (n = 40)**	**fillers in ‘easy minimal pairs’ group (n = 60)**	**fillers in ‘homophones’ group (n = 60)**
Mean	4.74	4.59	4.71	4.70
SD	1.22	0.72	0.48	0.94
Range	2.33-7.13	2.85-6.60	3.27-6.60	2.88-6.73

2.4 Procedure

The experiment took place at the Centre for Language Studies at Radboud University Nijmegen. Participants were invited to take a seat in a sound-attenuated booth, where they were provided with an information document and a consent form. After reading and signing those forms, participants were given a short oral instruction by the experimenter about the approximate duration of the experiment, 30-45 minutes, and usage of the button-box. At the start of the experiment, participants were asked to respond as accurately and quickly as possible. Written explanations, which included examples, were shown on screen before each task. The experimenter could see and hear the participant through a video connection, but most of the participants were not aware of this. The experiment was programmed using the open-source software PsychoPy (Peirce et al., 2019). An overview of the different pre-tests, training and post-tests is shown in Figure 5.

Figure 5. **Order of the main elements in the experiment: pretests, instruction and posttests.**

Participants started the experiment with the metalinguistic awareness pre-test. In this task, word pairs were shown on screen one pair at a time. For each word pair, participants were asked the question: “Do these words sound the same or different?”. Participants answered by pressing ‘same’ or ‘different’ on the button-box. Participants never received feedback on their performance. Word pairs were displayed in a random order, and the position of the word pairs (i.e., left or right on screen) was counterbalanced across participants. Every word pair was shown once, resulting in 100 trials for this task (i.e., 40 critical minimal pairs, 30 filler homophones and 30 filler minimal pairs).

After the metalinguistic awareness task, participants completed the 2-AFC identification pre-test, for which participants were instructed to put on the headphones. In this task, participants were shown the /æ/-/ɛ/ and /t/-/d/ critical minimal pairs in a random order. A minimal pair appeared on screen (e.g. ‘leg-lag’), then, after one second in which the participants could read the words, a recording of one of the words was played (e.g. ‘lag’). Participants then answered the question “Which word did you hear?" by pressing 'left' for the word on the left side of the screen, or 'right' for the word on the right side of the screen. Each member of a minimal pair was played over the headphones once, resulting in 80 trials total (i.e., two times 40 critical minimal pairs). The /æ/-/ɛ/ and /t/-/d/ critical minimal pairs were always displayed in the same order within-subjects, but it was counterbalanced between-subjects. For example, Participant 1 saw all /æ/ words on the left and all /ɛ/ words on the right, and Participant 2 saw the /ɛ/ words on the left and the /æ/ words on the right. Because of this, participants may have associated a button with a particular sound. Secondly, it may have helped participants to recognise which word was the /æ/ word and which one was the /ɛ/ word, when the spelling does not contain a transparent 'a' and 'e' (e.g. 'sad-said').

After these pre-tests, participants could take a short break if they wanted to. Then, they watched one of the four videos. After watching the video, participants were told they were going to do the first two tasks again, and they were asked to try to apply what they had learnt from the video. The 2-AFC identification post-test was placed right after the video, to make sure that the information of the video was still fresh in participants' minds. Subsequently, they took off their headphones and completed the metalinguistic awareness post-test. Both post-tests were identical to the pre-tests, with a new randomised order of the word pairs. The instructions for the post-tests were a little shorter, as participants were already familiar with the tasks.

After two pre-tests, video instruction and two post-tests, participants finished the experiment with a vocabulary task and a questionnaire. They were asked "Do you know this word?" for all 80 words of the 40 critical minimal pairs in the vocabulary task, to which they responded by pressing the 'yes' or 'no' button on the button-box. Then, participants completed the questionnaire on paper, which included demographic questions, questions about language background, as well as evaluation questions (see Appendix C).

2.5 Analyses

The data was cleaned using PyCharm (Python 3), and subsequent statistical analyses were done using R statistical software (R-Core-Team, 2019) and the package lme4 (Bates, Mächler, Bolker, & Walker, 2015). The R code that was written for the analyses can be found in Appendix G for reference. Binomial linear mixed effects models were used to analyse the accuracy scores on the identification task. Each model consisted of various fixed effects (main effects and interactions) and random effects, to account for by-participant and by-item variation. These random effects thus account for idiosyncratic variation for different items, and idiosyncratic variation due to individual differences between participants - they allow for different 'baselines' for each item and each participant (Winter, 2013). Using random effects resolves the non-independence of the repeated-measures design (Winter, 2013). In addition to the binomial mixed effects models, correlation analyses were carried out to answer the last research question.

3 Results

In this chapter, the results are presented in four parts, each concerned with a different research question. For the sake of readability, ‘*ae*-items’ is used to mean the critical minimal pairs with the /æ/-/ɛ/ contrast and ‘*td*-items’ to mean the critical minimal pairs with the word-final /t/-/d/ contrast. The term ‘*ae*-videos’ is used for the instruction videos about the /æ/-/ɛ/ contrast (i.e., both ae-dur and ae-nocue videos) and similarly, the term ‘*td*-videos’ is used for the instruction videos about the word-final /t/-/d/ contrast (i.e., both td-dur and td-nocue videos).

In the vocabulary test, participants reported knowing the meaning of on average 92,9% (SD: 25,7%) of the critical *ae*- and *td*-items (range 70%-100%). Forty-eight words of the 80 words were reported as ‘known’ by all participants. There were 32 words that not everybody knew, for which the scores can be found in Appendix D. Although not knowing a word could slow down a response, we assume that the 2-AFC identification task was not affected that much by lexical knowledge (i.e., to make the decision whether a /t/ or /d/ was heard, participants did not have to know the word to be able to click on a *t*-item or a *d*-item). Because the word knowledge scores in general were so high, it was decided not to exclude any words from the 2-AFC identification task or metalinguistic awareness task for the analyses.

The questionnaire question “How well do you think you understood the explanation in the video?” was answered by participants on a 6-point Likert scale with values from 0-5, for which the mean score was 4.48 (SD: 0.67, range 3-5). Fifty-eight of 64 participants indicated they understood the videos either very well (rating of 4 out of 5, n = 21) or perfectly (rating of 5 out of 5, n = 37). These results indicate that the videos were generally understandable and clear for the participants, which is a baseline requirement for the videos to be useful.

3.1 Effects of phonetic instruction versus unrelated phonetic instruction

RQ 1) Does a short explicit phonetic instruction improve the perception skills of Dutch listeners of English, for the English contrasts /æ/-/ɛ/ and word-final /t/-/d/?

The effect of the instruction videos are discussed for the *ae*-items (Section 5.1.1) and the *td*-items (Section 5.1.2). For this analysis, the accuracy scores of the 2-AFC identification task were calculated for the pre- and post-tests: button-box responses were scored as correct or incorrect, so that each trial received an accuracy score of zero or one. The accuracy scores on the identification task were compared for experimental versus control item, see Table 4. When the contrast of the item matches the contrast explained in the video (e.g., *ae*-item + *ae*-video condition), they are experimental items. When the contrast of an item does not match with the contrast explained in the video (e.g., *ae*-item + *td*-video condition), they are control items.

Table 4. **Accuracy on the identification task is compared for experimental items versus control items. Experimental items are those items that match with the instruction video. Control items are those that do not match with the instruction video.**

Condition	*ae*-items	*td*-items
ae-video (= ae-dur and ae-nocue, n = 32)	experimental items	control items
td-video (= td-dur and td-nocue, n = 32)	control items	experimental items

3.1.1 Effect of video instruction on the perception of the /æ/-/ɛ/ contrast

Do participants show a higher increase in accuracy on the *ae* items after instruction on the *ae* contrast, compared to the accuracy score on the same items after instruction on the *td* contrast? A binomial linear mixed effects model was performed in R (R-Core-Team, 2019) using lme4 (Bates et al., 2015). The binary variables ‘video contrast’ (ae-video or td-video) and ‘time’ (pre-test or post-test) were entered into the model as fixed effects, as well as their interaction (time * video contrast) into the model as fixed effects. We also had intercepts for ‘participant’ and ‘item’ (the correct word per trial) as random effects, as follows:

*glmer(accuracy ~ time + video_contrast + time*video_contrast*
+ (1|participant) + (1|correct_word)

The results are presented in Table 5 below. There was a significant main effect for time (p = 0.000063) and a significant interaction between time and video contrast (p = 0.0058).

***Table 5.* Binomial linear mixed effects model output for ae items. There is a significant increase in accuracy from pre- to post-test, as well as a significant interaction between time and video contrast.**

Predictor	β	Sd. Error	z-value	p-value
Intercept *(ae-video, pre-test)*	1.56427	0.18285	8.555	< 2e-16 ***
Time *(post)*	0.43609	0.10895	4.002	6.27e-05 ***
Video contrast *(td)*	-0.01297	0.21389	-0.061	0.95165
Time * Video contrast *(post, td)*	-0.41439	0.15008	-2.761	0.00576 **

The significant main effect of time means that, overall, there was a significant improvement on *ae*-items from pre-test to post-test. The significant interaction between time and video contrast means that the group who watched an *ae*-video improved significantly more (from pre- to post-test) on the experimental *ae*-items, than the group who watched a *td*-video on the control *ae*-items. This is visible in Figure 6: both groups scored similarly in the pre-test, but after video instruction, only the group who watched an *ae*-video improved their perception of the /æ/-/ɛ/ contrast.

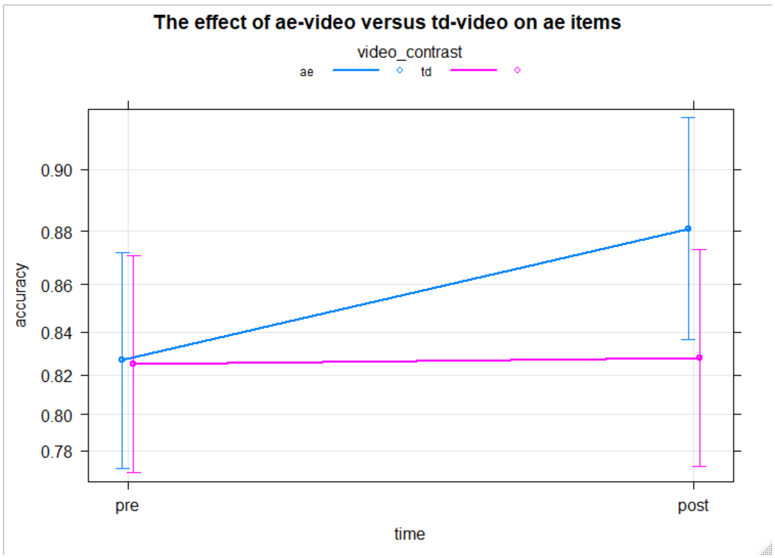

***Figure 6.* Effect of video instruction on accuracy of *ae*-items over time. The blue line displays the accuracy scores on the *ae*-items for participant group who watched an *ae*-video (experimental items, n = 32). The pink line displays the mean accuracy scores on the *ae*-items for the participant group who watched a td-video (control items, n = 32). (Error bars represent the standard error of the mean.)**

The English proficiency of the participants may have influenced the identification task results. As mentioned above, participants gave their self-rated English proficiency for speaking, reading, writing and listening on a 6-point Likert scale, in the questionnaire. Additionally, participants also estimated how many hours they actively listened to English in a typical week.

Firstly, the self-rated listening proficiency was entered into the model as an extra fixed effect and in the interaction. However, the three-way interaction (time x video contrast x listening proficiency) failed to converge. So instead, the listening proficiency was only added as a fixed effect, as follows:

*glmer(accuracy ~ time + video_contrast + listening_proficiency + time*video_contrast + (1|participant) + (1|correct_word)*

The results are shown in Table 6. There was a significant main effect of listening proficiency (p = 0.0029), but this did not change the interaction between time and video contrast (p = 0.0057).

***Table 6.* Binomial linear mixed effects model output for *ae*-items, with listening proficiency as a fixed effect. The self-rated listening proficiency had a significant effect on accuracy of the *ae*-items.**

Predictor	**β**	**Sd. Error**	**z-value**	**p-value**
Intercept *(ae pre)*	0.3929	0.4273	0.919	0.35785
Time *(post)*	0.4368	0.1091	4.005	6.21e-05 ***
Video contrast *(td)*	-0.01031	0.2040	-0.505	0.61331
Listening proficiency	0.3385	0.1135	2.983	0.00286 **
Time * Video contrast *(post, td)*	-0.4151	0.1501	-2.765	0.00569 **

Secondly, we calculated the mean score of their self-rated proficiency in speaking, reading, writing and listening taken together, and added this ‘mean proficiency’ score to the model as a fixed effect. Again, a three-way interaction again failed to converge. Instead, the mean proficiency was only added as a fixed effect:

*glmer(accuracy ~ time + video_contrast + mean_proficiency + time*video_contrast + (1|participant) + (1|correct_word)*

The results are presented in Table 7, below. According to this model, mean proficiency ratings had a significant effect on the accuracy scores of ae items (p = 0.00050), and the interaction between time and video contrast remained significant as well (p = 0.0056).

***Table 7.* Binomial linear mixed effects model output for ae items, with mean proficiency as a fixed effect. The self-rated listening proficiency had a significant effect on accuracy of the ae items.**

Predictor	**β**	**Sd. Error**	**z-value**	**p-value**
Intercept *(ae pre)*	0.02705	0.47110	0.057	0.95421
Time *(post)*	0.43726	0.10914	4.006	6.16e-05 ***
Video contrast *(td)*	-0.13297	0.20132	-0.660	0.50895
Mean proficiency	0.48379	0.13899	3.481	0.00050 ***
Time * Video contrast *(post, td)*	-0.41562	0.15015	-2.768	0.00564 **

Third, the estimations participants gave of their listening experience in a typical week was also added into the model as a fixed effect. Again, a three-way interaction again failed to converge. Instead, the listening hours per week were added as a fixed effect only:

*glmer(accuracy ~ time + video_contrast + listening_hpw + time*video_contrast + (1|participant) + (1|correct_word)*

The results are presented in Table 8 below. The number of hours people said they listened to English in a typical week had a significant effect on the accuracy on the ae items (p = 0.0050), and the interaction between time and video contrast remained significant (p = 0.0058).

***Table 8.* Binomial linear mixed effects model output for accuracy scores for *ae*-items in the identification task, with listening hours per week as a fixed effect. The self-rated listening proficiency had a significant effect on accuracy of the *ae*-items.**

Predictor	**β**	**Sd. Error**	**z-value**	**p-value**
Intercept *(ae pre)*	1.329729	0.194452	6.838	8.01e-12 ***
Time *(post)*	0.435877	0.108949	4.001	6.31e-05 ***
Video contrast *(td)*	-0.099015	0.206083	-0.480	0.63090
Listening hours per week	0.022580	0.008048	2.805	0.00502 **
Time * Video contrast *(post, td)*	-0.414178	0.150107	-2.759	0.00579 **

The three models discussed above found that three different proficiency measures (self-rated listening proficiency, mean self-rated proficiency or estimated hours of listening to English per week) were all significant predictors of the accuracy scores of participants in the identification task. Which proficiency measure is the best predictor? The AIC scores for the three models are indicative of how well the models ‘fit’ with the data. The lower the AIC, the better the model fit. The AIC scores for the three models were as follows:

(1) Model including self-rated listening proficiency: AIC of 4645.0
(2) Model including mean self-rated proficiency: AIC of 4642.2
(3) Model including estimated hours of listening to English per week: AIC of 4645.9

At first glance, the AIC scores of the three models are very similar. ANOVA comparisons indicated that the model including the mean self-rated listening proficiency (Model 2) had a significantly better fit than the model including the self-rated listening proficiency (Model 1) ($X^2(7) = 2.88$, $p < 2.2e\text{-}16$). No significant differences were found for other ANOVA comparisons between the three models. Surprisingly, even though the difference in AIC scores between Model 2 and Model 1 was smaller than between Model 2 and Model 3, no significant difference was found in an ANOVA comparison between Model 2 and 3. In conclusion, even though the mean of the self-rated proficiency scores was found to be a slightly better predictor than the self-rated listening score, the three productivity measures are quite equally good predictors.

3.1.2 Effect of video instruction on the perception of the word-final /t/-/d/ contrast

Now we turn to the analysis of the accuracy scores on the *td*-items in the 2-AFC identification task. Again, the scores of the group who watched a *td*-video was compared to the group who watched an *ae*-video. Do participants show a bigger improvement in accuracy for the *td*-items after watching a *td*-video, compared to when they watched an *ae*-video?

The binary variables ‘time’ (pre- or post-test) and ‘video contrast’ (*td*-video or *ae*-video) and their interaction were entered as fixed effects into a binomial linear mixed effects model, with ‘participant’ and ‘item’ (correct word) as random effects, as follows:

*glmer(accuracy ~ time + video_contrast + time*video_contrast*
+ (1|participant) + (1|correct_word)

The results are presented in Table 9 below. There was a significant main effect for time ($p = 0.046$) and a significant interaction between time and video type ($p = 0.0077$). The significant interaction indicates that *td*-videos were significantly more effective than *ae*-videos at improving identification accuracy scores on *td*-items.

Table 9. **Binomial linear mixed effects model results for accuracy scores on the identification task on td items. There is a significant increase in accuracy from pre- to post-test, as well as a significant interaction between time and video contrast (instruction on the ae contrast versus the td contrast).**

Predictor	**β**	**Sd. Error**	**z-value**	**p-value**
Intercept *(ae pre)*	2.7056	0.2650	10.208	< 2e-16 ***
Time *(post)*	0.2814	0.1411	1.994	0.0462 *
Video contrast *(td)*	-0.1706	0.2930	-0.582	0.5604
Time * Video contrast *(post, td)*	0.5400	0.2027	2.665	0.0077 **

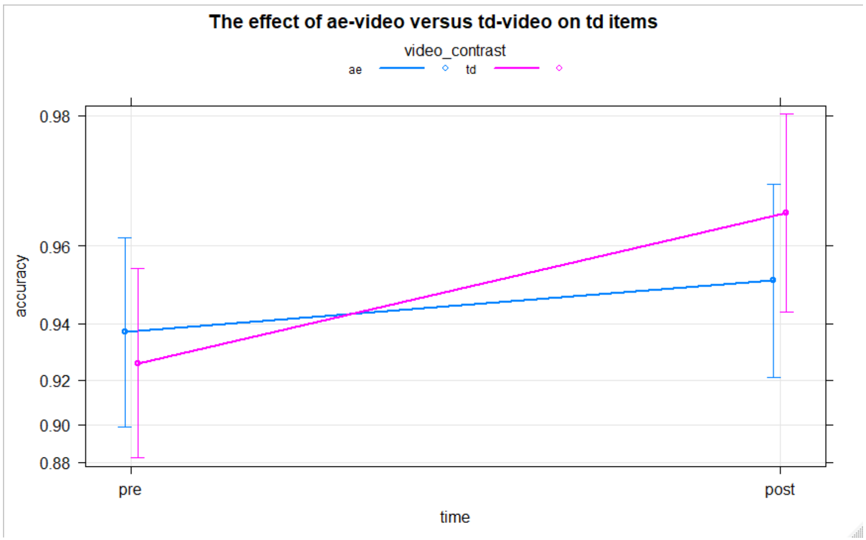

Figure 7. **Effect of video instruction on accuracy of *td* items over time. The pink line displays the mean accuracy scores on the *td* items for the participant group who watched a td-video (experimental items, n = 32). The blue line displays the accuracy scores on the *td* items for the participant group who watched an ae-video (control items, n = 32). (Error bars represent the standard error of the mean.)**

Figure 7 shows the results for both groups in a graph. This graph suggests that the *td*-video group started with slightly lower scores than the *ae*-video group at pre-test, but at post-test, the *td*-video group scored higher than the *ae*-video group. The graph also suggests that the group who watched an *ae*-video also improved on *td*-items. Indeed, the effect of time was significant for the group who watched an *ae*-video: a binomial linear mixed effects model for the data on *ae*-video *td*-items, with time as fixed effect and participant and item as random effects yielded a significant effect for time (β = 0.28, Sd. Error = 0.14, p = 0.047). This could be due to practise effects - the recognition of *td*-items may become easier as a result of doing the task.

These results indicate that participants were able to improve their identification accuracy of *td*-items after watching a *td*-video, more than after watching an *ae*-video. Those who saw an *ae*-video (control group) also improved on the *td*-items, but the *td*-video was more effective improving participants' perception of the word-final /t/-/d/ contrast.

The influence of English proficiency was analysed for the *td*-items as well. Firstly, the self-rated 'listening proficiency' was entered into the model, alongside 'time' and 'video contrast'. The model failed to converge with a three-way interaction, so the proficiency was only added as a fixed effect.

glmer(accuracy ~ time + video_contrast + listening_proficiency
*+ time*video_contrast + (1|participant) + (1|correct_word)*

The results are displayed in Table 10 below: self-rated listening proficiency had a significant effect on the accuracy of the identification task for *td*-items (p = 0.0045). The interaction between time and video contrast remained significant (p = 0.0079).

***Table 10.* Binomial linear mixed effects model output for *td*-items, with listening proficiency as a fixed effect. The self-rated listening proficiency had a significant effect on accuracy of the *td*-items.**

Predictor	**β**	**St. Error**	**z-value**	**p-value**
Intercept *(ae pre)*	1.1536	0.5973	1.931	0.05346 .
Time *(post)*	0.2817	0.1413	1.994	0.04619 *
Video contrast *(td)*	-0.2967	0.2792	-1.063	0.28788
Listening proficiency	0.4486	0.1579	2.842	0.00449 **
Time * Video contrast *(post, td)*	0.5390	0.2028	2.658	0.00786 **

Secondly, the mean proficiency measure (averaged across self-rated writing, listening, speaking and reading) was added in the same way as the listening proficiency. Again, a three-way interaction (time x video contrast x mean proficiency) failed to converge, so mean proficiency was added as a fixed effect only, as follows:

*glmer(accuracy ~ time + video_contrast + mean_proficiency + time*video_contrast*
+ (1|participant) + (1|correct_word)

***Table 11.* Binomial linear mixed effects model output for *td*-items, with mean (self-rated) proficiency as a fixed effect. The mean proficiency had a significant effect on accuracy of the *td*-items.**

Predictor	β	St. Error	z-value	p-value
Intercept *(ae pre)*	0.9103	0.6634	1.372	0.16998
Time *(post)*	0.2820	0.1414	1.995	0.04609 *
Video contrast *(td)*	-0.3221	0.2789	-1.155	0.24817
Mean proficiency	0.5649	0.1948	2.900	0.00373 **
Time * Video contrast *(post, td)*	0.5380	0.2028	2.653	0.00798 **

Table 11 displays the results for this binomial mixed effects model. It shows that the mean self-rated proficiency had a significant effect on the accuracy of the identification task for *td*-items (p = 0.0037), and the interaction between time and video contrast is also still significant (p = 0.0080).

Third, participants' estimation of their active listening to English in hours in a typical week (hpw) were also added to the model in the same way. The model failed to converge with a three-way interaction (time x video contrast x listening hpw), so the listening hours per week were added as a fixed effect only:

*glmer(accuracy ~ time + video_contrast + listening_hpw + time*video_contrast + (1|participant) + (1|correct_word)*

Table 12 below shows that there was a significant effect of listening hours per week on the accuracy scores (p = 0.047), and the interaction between time and video contrast remained significant as well (p = 0.0077).

***Table 12.* Binomial linear mixed effects model output for *td*-items, with listening hours per week (as reported by participants) as a fixed effect. The listening exposure had a significant effect on accuracy of the *td*-items.**

Predictor	β	St. Error	z-value	p-value
Intercept *(ae pre)*	2.47503	0.28358	8.728	< 2e-16 ***
Time *(post)*	0.28131	0.14115	1.993	0.04627 *
Video contrast *(td)*	-0.26529	0.28845	-0.920	0.35772
Listening hours per week	0.02247	0.01133	1.984	0.04730 *
Time * Video contrast *(post, td)*	0.54039	0.20271	2.666	0.00768 **

The three different proficiency measures (self-rated listening proficiency, mean self-rated proficiency or estimated hours of listening to English per week) were all significant predictors of the accuracy scores of *td* items in the identification task. The AIC scores for the three models were as follows:

(1) Model including self-rated listening proficiency: AIC of 2849.2
(2) Model including mean self-rated proficiency: AIC of 2848.9
(3) Model including estimated hours of listening to English per week: AIC of 2852.9

The AIC scores for these models are very similar. ANOVA comparisons indicated that the model including mean self-rated proficiency (Model 2) had a significantly better fit than the model that included the self-rated listening proficiency (Model 1) ($X^2(7) = 0.2416$, $p < 2.2e-16$). Surprisingly, even though the difference in AIC scores between Model 2 and Model 3 is greater than between Model 1 and 2, no significant difference was found comparing mean self-rated proficiency (Model 2) and estimated hours of English listening per week (Model 3). Thus, although the mean of the self-rated proficiency scores was found to be a slightly better predictor than the self-rated listening score, the three productivity measures are quite equally good predictors.

3.2 Effects of explanation type (duration or no-cue) on accuracy

From the results described in Section 5.1, it seems that an instruction video on a specific phonetic contrast was helpful to participants so that they performed better on the post-test than on the pre-test on this contrast. The follow-up question is, how much of this learning effect is due to explaining the duration cue, versus mere attention toward the contrast in question? Directing attention to this specific cue could hypothetically make recognition easier for listeners. Does a video in which the duration cue is explained help participants identify the right sound, more than a video in which this duration cue is not explained?

RQ 2) If the explicit phonetic instruction helps improve the English perception skills of Dutch listeners, how much of this effect is due to the presence of instruction on the duration cue?

3.2.1 *The effect of attention to the duration cue on perception of ae items*

To answer this research question for the *ae*-items, the accuracy scores for *ae*-items on the 2-AFC identification task were compared for the group who watched an ae-dur video (in which the duration cue was explained) and the group who watched an ae-nocue video (in which the duration cue was not explained). A binomial linear mixed model was used, in which the fixed effects were 'time' (pre- versus post-test) and 'video type' (dur versus nocue) and the interaction between these two. The random effects were 'participant' and 'item'.

*glmer(accuracy ~ time + video_type + time*video_type*

+ (1|participant) + (1|correct_word)

The outcomes of the model are shown in Table 13. There was a significant main effect of time (p = 0.026) but no significant interaction effect between time and video type (p = 0.49).

Table 13. **Binomial linear mixed effects model results for *ae*-items. There is a significant increase in accuracy from pre- to post-test, but no significant interaction between time and video type (instruction with or without duration cue).**

Predictor	**β**	**St. Error**	**z-value**	**p-value**
Intercept *(pre, dur)*	1.7967	0.2320	7.746	9.51e-15 ***
Time *(post)*	0.3591	0.1614	2.225	0.0261 *
Video type *(nocue)*	-0.4252	0.2805	-1.516	0.1296
Time * Video type *(post, nocue)*	0.1498	0.2191	0.683	0.4943

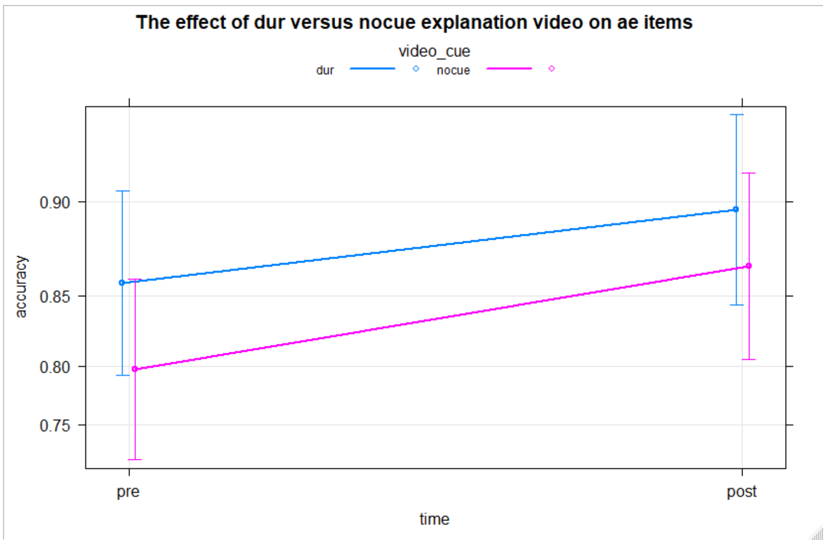

Figure 8. **The effect of the video explanation (*ae*-video with duration cue or *ae*-video without the duration cue) on the accuracy scores on the perception task, for the *ae*-items. (Error bars represent the standard error of the mean.)**

The data is visualised in Figure 8. The participants who watched the video with a duration cue improved their accuracy scores on the perception task, but not more than the participants that watched the video without the duration cue. Figure 8 also suggests that the no-cue group scored lower on the identification task in both pre- and post-test than the duration group. However, this difference between the groups is not significant (no main effect of video type, p = 0.13, see Table 13). The significant main effect for time and the absence of an interaction between time and video type means that both groups improved on the *ae*-items after watching an *ae*-video, regardless of whether the explanation on the duration cue was included in the video or not.

3.2.2 The effect of attention to the duration cue on perception of word-final td items

The same analysis was done for the *td*-items: the accuracy scores on *td*-items were compared for the participants who watched the td-dur video versus those who watched the td-nocue video. Is the duration cue useful for listeners to pay attention to? Again, a binomial linear mixed model was used to analyse the data in R. The fixed effects were 'time' and 'video type' and their interaction, and the random effects were 'participant' and 'item'. The results of this model are shown in Table 14 below. The model shows a significant main effect of time (p = 0.000022) but, similar to the outcome for the ae items, no significant interaction between time and video type (p = 0.65).

Table 14. **Binomial linear mixed effects model results for *td*-items. There is a significant increase in accuracy from pre- to post-test, but no significant interaction between time and video type (instruction with or without duration cue).**

Predictor	**β**	**St. Error**	**z-value**	**p-value**
Intercept *(pre, dur)*	2.7353	0.3776	7.244	4.37e-13 ***
Time *(post)*	0.9028	0.2128	4.243	2.20e-05 ***
Video type *(nocue)*	-0.2373	0.4581	-0.518	0.604
Time * Video type *(post, nocue)*	-0.1304	0.2912	-0.448	0.654

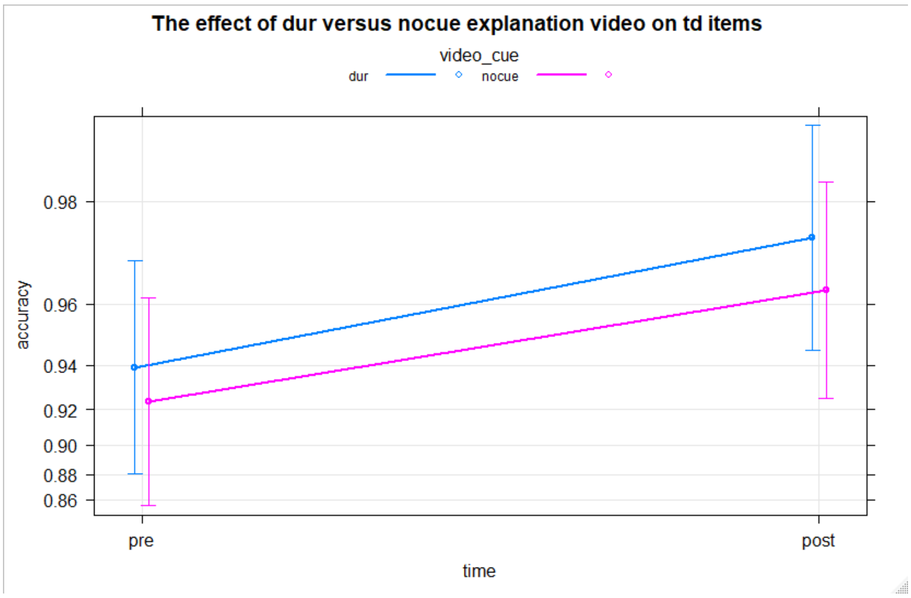

Figure 9. **The effects of the td-dur video, which included duration cue explanation (blue line) versus the effect of the td-nocue video which did not include the duration cue explanation (pink line). (Error bars represent the standard error of the mean.)**

The results for the *td*-items, split up for video type, are shown in Figure 9. It suggests that the no-cue group scored a little lower again overall than the duration group, but there was no significant main effect for video cue (p = 0.60). From these analyses, it can be concluded that there was no significant difference between the td-dur and td-nocue groups. The participants to whom the duration cue had been explained scored similarly on the identification task over time as the participants who had not been instructed about the duration cue.

3.3 Asymmetry between recognition of the different phonemes

The third research question asked:

RQ 3) Is there evidence for asymmetric assimilation for the English /æ/-/ɛ/ and word-final /t/-/d/ contrasts for Dutch listeners?

The accuracy scores on the 2-AFC identification task were analysed for differences between accuracy on /æ/ versus /ɛ/ and for accuracy on word-final /t/ versus /d/. To assess the accuracy scores for each phoneme, all trials were coded for ‘right answer type’, which had the value ‘a’ or ‘e’ or ‘t’ or ‘d’. Because analyses in Section 5.2 showed that the duration and nocue groups scored similarly, the analysis does not differentiate between duration and nocue videos (i.e. using the groups *ae*-videos, ae-dur and ae-nocue together, and td-videos, td-dur and td-nocue together).

3.3.1 Difficulty of /æ/ versus /ɛ/ items

There were 20 trials with the /æ/ sound, and 20 trials with the /ɛ/ sound in the pre-test of the identification task, and the same amount in the post-test. From the boxplots in Figure 10.1 below, the /æ/ words seem to be of similar difficulty as the /ɛ/ words. Indeed, a binomial linear mixed effects model with ‘right answer type’ as a fixed effect, and ‘participant’ and ‘item’ as random effects, showed no significant main effect of ‘right answer type’ on accuracy scores (β = 0.27, Sd. Error = 0.21, p = 0.20).

accuracy ~ right_answer_type + (1 | participant) + (1 | correct_word)

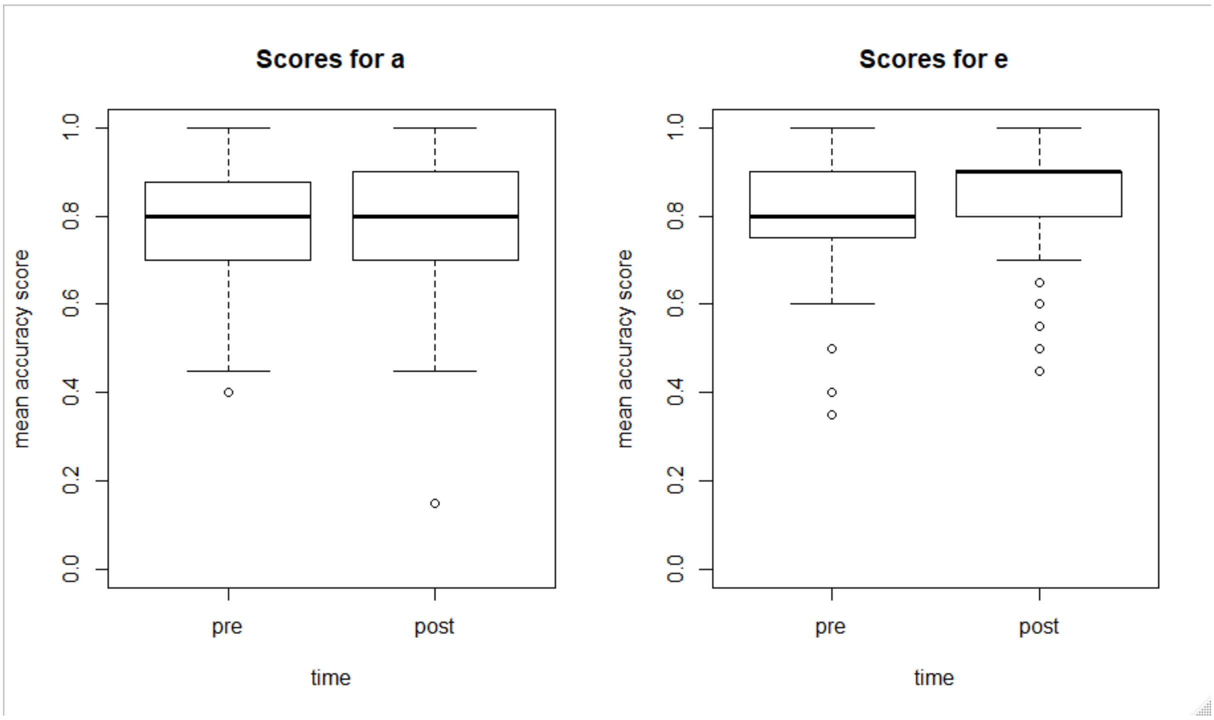

***Figure 10.1 (left) and 10.2 (right)*. Boxplots of the mean accuracy scores for those trials where the /æ/ word was the right answer and those trials where the /ɛ/ word was the right answer.**

Figure 10.2 seems to indicate that participants improved on identification of /ɛ/ words, while participants did not improve on identification of /æ/ words. However, the effect of time is more informative when video contrast (*ae*-video or *td*-video) is taken into account. The graphs in Figure 11 below show the results for /æ/ and /ɛ/ by video contrast. From this figure, it seems that participants who saw the *td*-videos (upper row, Figure 11) seem to have slightly more difficulty identifying /æ/ words. For participants who watched *ae*-videos (lower row, Figure 11), it seems that the accuracy on /æ/ words and on /ɛ/ words both improved in the post-test, with a slightly steeper improvement for the /ɛ/ words. Figure 11 suggests that there could be significant interactions between time, video contrast and right answer type.

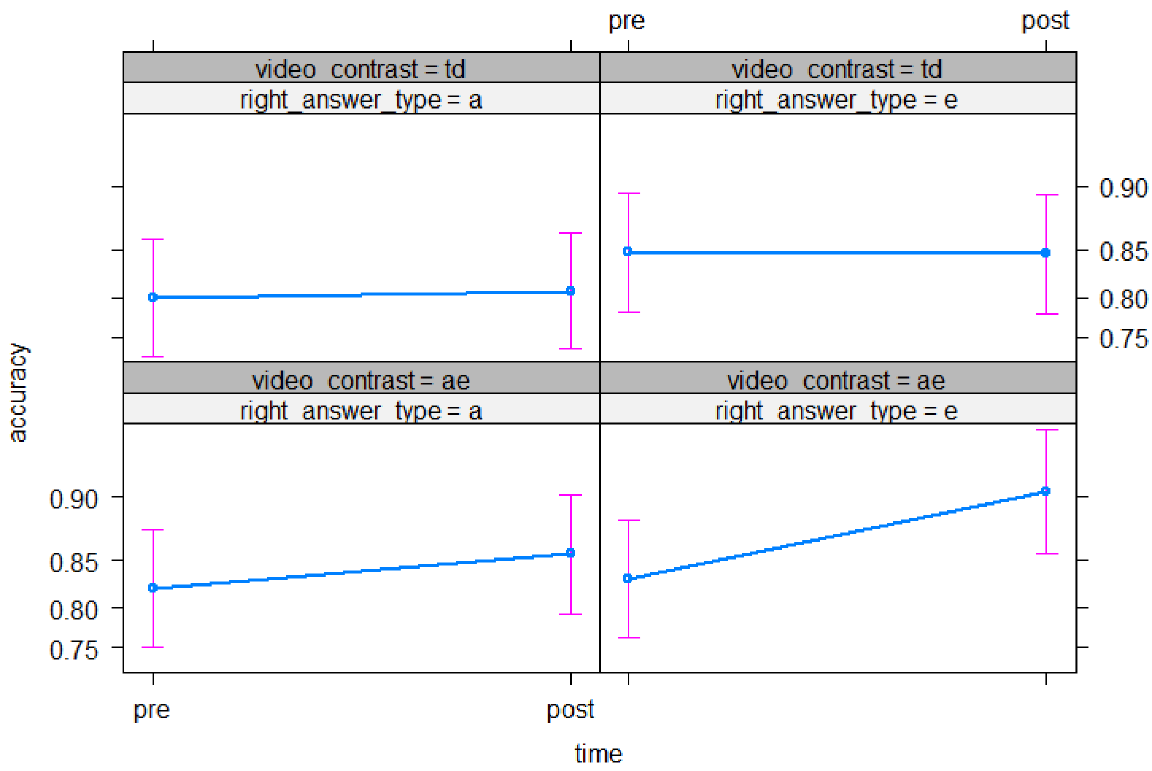

Figure 11. **Line graphs of the accuracy scores for trials in which the right answer was the /æ/ word (left column) and for trials for which the right answer was the /ɛ/ word (right column). The effect of time is visible (pre- vs post-test) for videos that explained the /t/-/d/ contrast (upper row) and for videos that explained the ae contrast (lower row). (Error bars represent the standard error of the mean.)**

Another binomial linear mixed effects model was calculated, this time with the fixed effects 'time' (pre or post), 'right answer type' (/æ/ or /ɛ/ word), 'video contrast' (*ae*-video or *td*-video), and their interactions, as well as random effects for 'participant' and 'item'.

accuracy ~ time + right_answer_type + video_contrast
*+ time * right_answer_type * video_contrast*
+ (1 | participant) + (1 | correct_word)

The results are shown in Table 15 below, which shows that none of the main effects or interactions were significant. The interaction between 'time' and 'right answer type' was approaching significance ($p= 0.079$). (The interaction between 'video contrast' and 'time' was not significant for *ae*-items here, while it was significant previously in Section 5.1.1.)

Table 15. **Binomial linear mixed effects model results for *ae* items: no significant main effects or interactions between time, right answer type (/æ/ or /ɛ/) and video contrast (*ae*-video or *td*-video).**

Predictor	**β**	**St. Error**	**z-value**	**p-value**
Intercept *(a, ae, pre)*	1.52383	0.21905	6.957	3.49e-12 ***
Time *(post)*	0.25758	0.14858	1.734	0.0830 .
Right answer type *(e)*	0.07357	0.24645	0.299	0.7653
Video contrast *(td)*	-0.1374	0.23544	-0.583	0.5597
Time *(post)* * Right answer type *(e)*	0.38426	0.21898	1.755	0.0793 .
Time *(post)* * Video contrast *(td)*	-0.20669	0.20522	-1.007	0.3139
Right answer type *(e)* * Video contrast *(td)*	0.26057	0.20585	1.266	0.2056
Time *(post)* * Right answer type *(e)* * Video contrast *(td)*	-0.44677	0.30132	-1.483	0.1382

In conclusion, this analysis shows that, although the tendencies in Figure 15 are generally in line with the expectations, these tendencies are not statistically significant. The data do not show that /ɛ/ words are easier than /æ/ words, or that /æ/ items are more often confused with /ɛ/ than the other way around.

3.3.2 Difficulty of /t/ versus /d/ items

For the *td*-items, there were 20 /t/ trials and 20 /d/ trials, for pre- and post-test. Figure 12 below seems to indicate that the accuracy for /t/ words is generally higher than the scores for /d/ words.

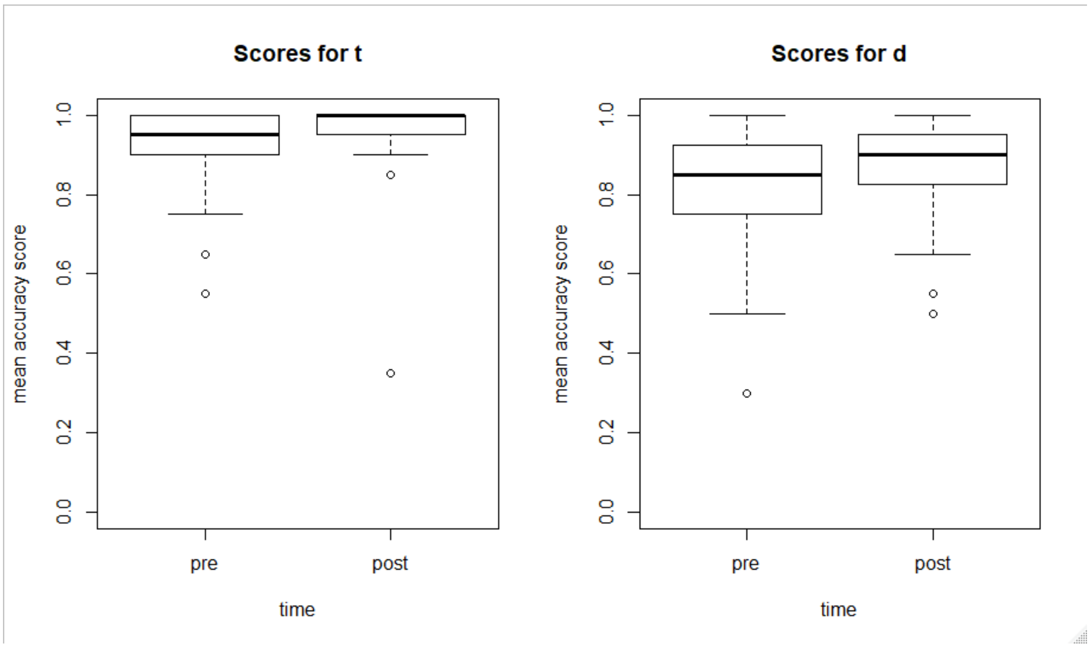

Figure 12. **Boxplots to compare participants' performance on trials for which the /t/ word was the correct answer and trials for which the /d/ word was the correct answer.**

A binomial linear mixed effects model was used to analyse this effect statistically. We entered 'right answer type' as a fixed effect, and 'participant' and 'item' as random effects. The results showed a significant main effect of 'right answer type' on accuracy scores (β = 1.18, Sd. Error = 0.27, p = 0.000015).

accuracy ~ right_answer_type + (1 | participant) + (1 | correct_word)

This is in line with the hypothesis: /d/ items are significantly more difficult than /t/ items. When /d/ is pronounced at the end of a word, it is interpreted by Dutch listeners as a /t/ significantly more often than the other way around. Again, the effect of 'time' in the boxplots in Figure 12 above are not very informative without taking into account the effect of video contrast. The graphs in Figure 13 below show the effect of the video for video contrast (*ae*-video or *td*-video) and right answer type (t or d).

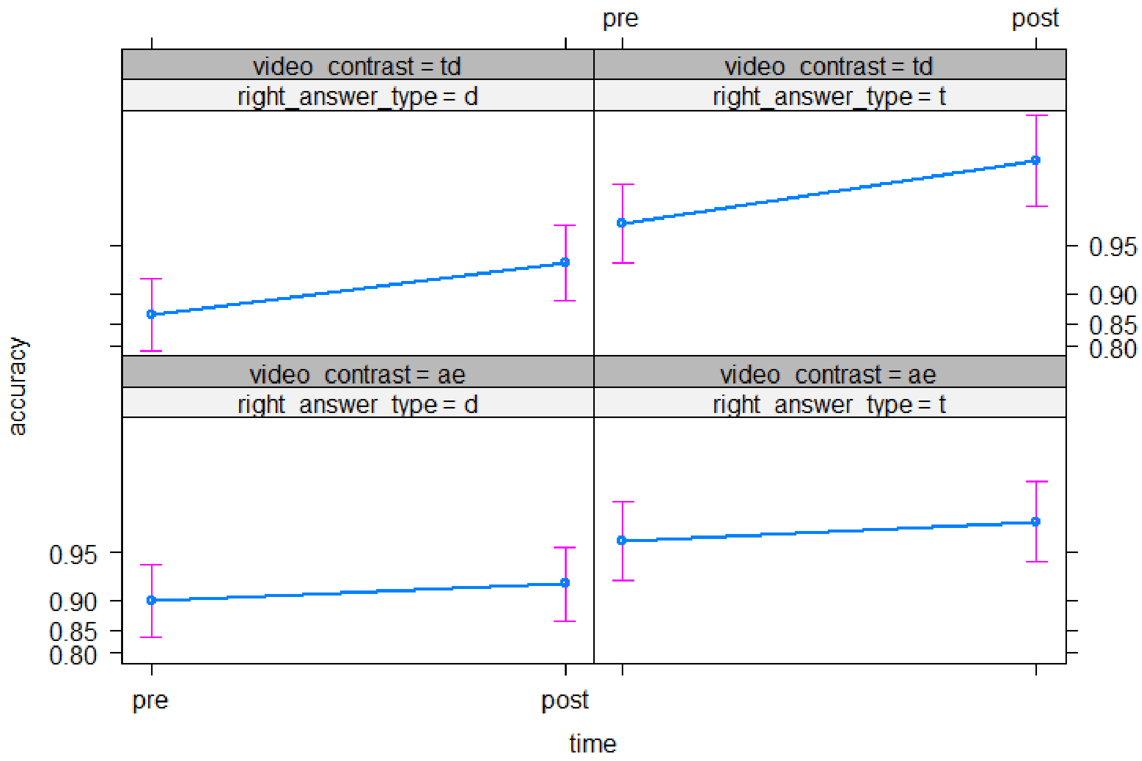

Figure 13. **Line graphs of the accuracy scores for trials in which the right answer was the /d/ word (left column) and for trials for which the right answer was the /t/ word (right column). The effect of time is visible (pre- vs post-test) for td-videos (upper row) and for ae-videos (lower row). (Error bars represent the standard error of the mean.)**

Figure 13 seems to indicate that participants scored better on /t/ items than on /d/ items regardless of which video contrast they received instruction on (main effect of right answer type). Both groups seem to show improvement over time (main effect of time), and the upper right graph shows the best scores: td-video and /t/ items (interaction between video contrast and right answer type).

To find out whether these tendencies are significant, a second binomial linear mixed effects model involved the fixed effects ‘time’ (pre or post), ‘right answer type’ (/t/ or /d/ word) and ‘video contrast’ (ae-video or td-video), and their interactions. Again, ‘participant’ and ‘item’ were added as random effects.

accuracy ~ time + right_answer_type + video_contrast
*+ time * right_answer_type * video_contrast*
+ (1 | participant) + (1 | correct_word)

Table 16 shows that the main effect of 'right answer type' was significant (p = 0.0043). However, there was no significant main effect for time (p = 0.12), and there was no significant interaction between right answer type and video contrast, although it is approaching significance (p = 0.081). The results do show a significant interaction between time and video contrast (p = 0.030) − the td-video was more effective at improving participants' performance over time than the ae-videos could, which was already discussed in Section 5.1.2.

***Table 16* . Mixed effects model output for td items: effects of right answer type, time and video contrast.**

Predictor	**β**	**St. Error**	**z-value**	**p-value**
Intercept *(d, ae, pre)*	2.18973	0.28434	7.701	1.35e-14 ***
Time *(post)*	0.26590	0.17372	1.531	0.12585
Right answer type *(t)*	0.92114	0.32241	2.857	0.00428 **
Video contrast *(td)*	-0.31853	0.30683	-1.038	0.29921
Time *(post)* * Right answer type *(t)*	0.03298	0.29403	0.112	0.91069
Time *(post)* * Video contrast *(td)*	0.52928	0.24451	2.165	0.03041 *
Right answer type *(t)* * Video contrast *(td)*	0.48952	0.28007	1.748	0.08050 .
Time *(post)* * Right answer type *(t)* * Video contrast *(td)*	0.12916	0.44584	0.290	0.77204

In conclusion, the accuracy scores on the 2-AFC identification task do show an asymmetry for the /t/-/d/ contrast. The significant main effect for 'right answer type' (t or d) indicates that the words with word-final /t/ were easier recognised than those with word-final /d/, the /d/ words were more often confused with a /t/ word.

3.4 Metalinguistic awareness and perception

In the metalinguistic awareness task, participants were asked whether word pairs sound the same (homophones) or different (minimal pairs). If participants experience difficulty distinguishing between the /æ/-/ɛ/ and word-final /t/-/d/ contrasts, they might think that the /æ/-/ɛ/ and /t/-/d/ minimal pairs are pronounced the same - they could also be unaware that there is *supposed* to be a difference in pronunciation between these word pairs.

Being aware that these words are supposed to sound different might help with identifying them. If participants categorise these minimal pairs as 'same' in pronunciation, their lack of awareness might correlate with a lower performance on the identification pre-test. If participants categorise these minimal pairs correctly as 'different' in pronunciation, they might also score higher on the identification pre-test.

Dutch listeners reported that they found it quite difficult to decide whether word pairs sounded exactly the same or not. Many participants based their judgements on their own pronunciation (i.e., they whispered the words before making a decision). The participants scored highly accurate on the filler minimal pairs, with a mean score of 0.92 (SD 0.27, range 60-100 percent correct), which means that participants understood the task. For the true homophones, the mean score was 0.80 (SD 0.40, range 30-100 percent correct). The results for the /æ/-/ɛ/ word pairs are presented in Figure 14.1 and the /t/-/d/ word pairs in Figure 14.2.

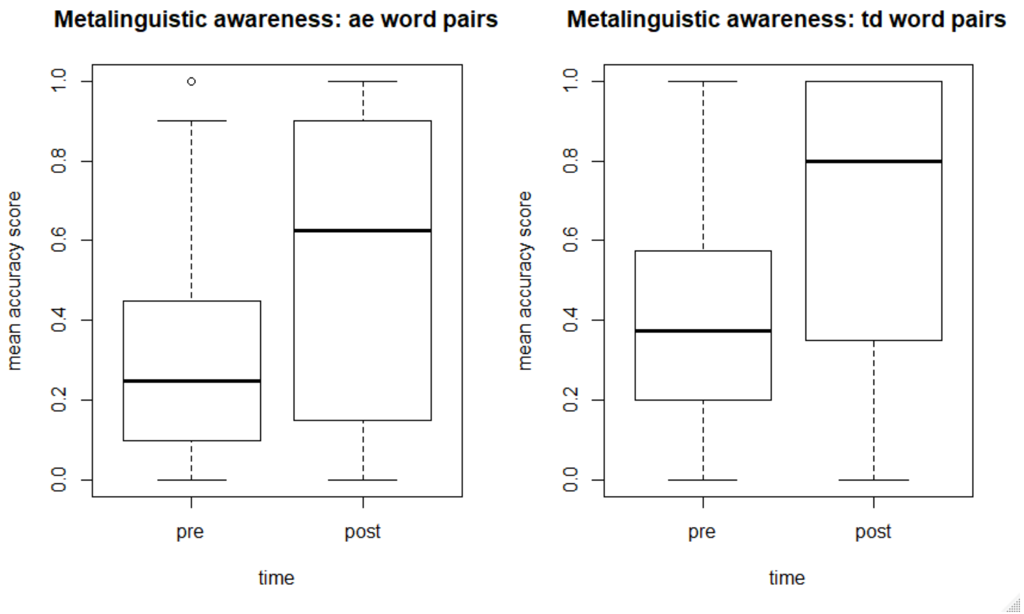

Figure 14.1 and 14.2. **The accuracy scores on the metalinguistic awareness task, for the ae word pairs (left) and the td word pairs (right), in pre- and post-test.**

It is surprising to see that the accuracy scores were low in the pre-test for both /æ/-/ɛ/ word pairs and /t/-/d/ word pairs. Apparently, most participants actively believed that /æ/-/ɛ/ word pairs and /t/-/d/ word pairs were pronounced the same, generally unaware of the subtle phonetic differences. Additionally, Figure 14 shows that participants performed better on the post-test than on the pre-test, for both /æ/-/ɛ/ word pairs and /t/-/d/ word pairs. For the /æ/-/ɛ/ word pairs, a binomial linear mixed effects model with ‘time’ as fixed effect and ‘participant’ and ‘item’ as random effects showed a significant effect of time ($\beta = 1.33$, Sd. Error = 0.10, $p = 2e-16$). For the /t/-/d/ word pairs, the same binomial linear effects model also resulted in a significant effect of time ($\beta = 1.56$, Sd. Error = 0.11, $p < 2e-16$).

From the boxplots in Figure 14, a difference between the awareness scores and the identification task scores can already be observed: the scores for metalinguistic awareness range from 0 to 100% correct, in both pre- and post-test, for both contrasts. For the 2-AFC identification task, nobody scored lower than 40%-50%, around chance level. For the metalinguistic awareness task, this was not the case, there is a wider range in scores. Of course, 50% accuracy is chance level for a binary answer task, so in both tasks, there were participants with no knowledge. We will analyse the relation between the two in more detail using correlation analyses, below.

In the post-test of the metalinguistic awareness task in Figure 14, the accuracy scores still range from 0-100%, which is surprising. Even though the 2-AFC identification task essentially implies that there is a difference between /æ/ and /ɛ/ words and /t/ and /d/ words, not all participants are aware by the end of the experiment that all /æ/-/ɛ/ word pairs and /t/-/d/ word pairs are pronounced differently. Interaction analyses between time (pre- and post-test) and video contrast type (*ae*-video or *td*-video) for these metalinguistic awareness scores are outside the scope of this thesis (but see Appendix F for the boxplots).

In the fourth research question, we asked:

RQ 4) Is there a correlation between Dutch L2 users of English metalinguistic awareness on the pronunciation differences between /æ/-/ɛ/ or word-final /t/-/d/ minimal pairs and the perception skills of these Dutch L2 users of English regarding the same minimal pairs?

To answer this question, participants’ pre-test score of the metalinguistic awareness task and their pre-test score on the identification task were analysed for correlation. The results for the /æ/-/ɛ/ word pairs are discussed in Section 5.4.1 and the results for /t/-/d/ word pairs in Section 5.4.2.

3.4.1 Correlation between awareness and identification task: the /æ/-/ɛ/ word pairs

How does awareness of the difference in pronunciation between ae word pairs relate to perception ability of that contrast? A correlation analysis was performed on the accuracy scores for the metalinguistic awareness task (pre-test) and the accuracy scores for the 2-AFC identification task. To do this, mean scores were calculated for every participant for (1) the /æ/-/ɛ/ word pairs on the metalinguistic awareness pre-test, and (2) the /æ/-/ɛ/ items in the 2-AFC identification pre-test. Figure 15, below, visualises the distribution of these two means in a scatterplot, in which every dot represents a participant (n = 64).

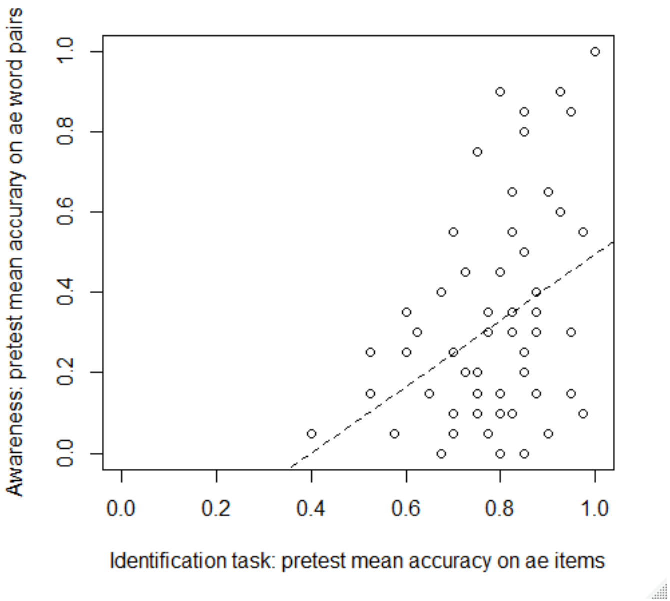

***Figure 15*. Scatterplot of the mean scores for /æ/-/ɛ/ word pairs of the metalinguistic awareness pre-test (y-axis) versus the mean scores for *ae*-items of the 2-AFC identification pre-test (x-axis) for every participant. The dotted line is the regression line of the two variables.**

Figure 15 shows that participants who have a lower mean score on the *ae*-items in the identification task (x-axis) also tend to score lower on metalinguistc awareness for /æ/-/ɛ/ word pairs (y-axis). However, participants who score high on the identification task can still score low on metalinguistic awareness (i.e., there are many dots in the lower right corner). It seems that metalinguistic awareness is not a prerequisite to score high on the identification task: many participants are not aware of the difference in pronunciation for the /æ/-/ɛ/ word pairs, but are able to distinguish them very well in the indentification task.

A Pearson product-moment correlation coefficient was computed to assess the relationship between the mean scores per participant on the identification task pre-test and the mean scores per participant on the metalinguistic awareness task pre-test. A medium significant correlation between the variables was found: r(62) = 0.38, p = 0.0018 (with quite a large range of confidence: 95% CI 0.15 - 0.57).

3.4.2 Correlation between awareness and identification task: the word-final /t/-/d/ items

A similar graph was computed for the td contrast: the mean scores on the td word pairs of the metalinguistic awareness pre-test, and the mean scores on the td items of the 2-AFC identification pre-test. The result is shown in Figure 16, where each dot represents a participant (n = 64).

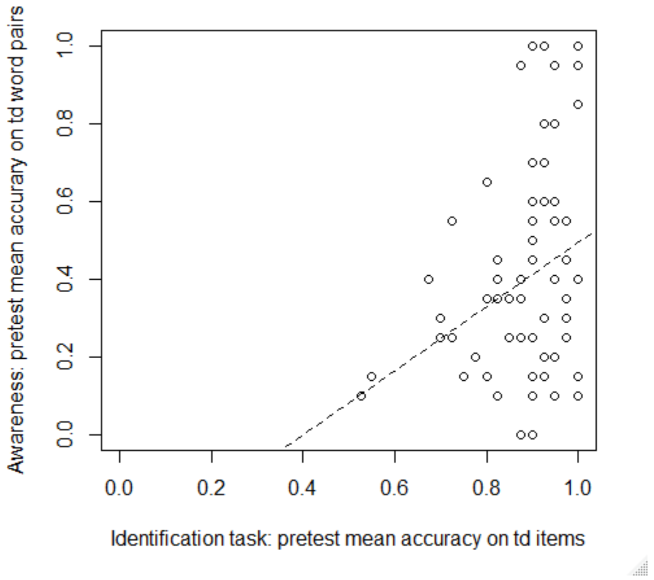

***Figure 16*. Scatterplot of the mean scores for td word pairs of the metalinguistic awareness pre-test (y-axis) versus the mean scores for td items of the 2-AFC identification pre-test (x-axis) for every participant. The dotted line is the regression line of the two variables.**

Figure 16 looks quite similar to Figure 15, although the scores on the /t/-/d/ contrast tended to be higher overall than the scores on the /æ/-/ɛ/ contrast. Figure 16 shows that participants who have a lower mean score on the *td*-items in the identification task (x-axis) also tend to score lower on metalinguistic awareness for /t/-/d/ word pairs (y-axis). Again, like for the /æ/-/ɛ/ contrast, participants who score high on the identification task can still score low on metalinguistic awareness.

Metalinguistic awareness is not a prerequisite to score high on the identification task: many participants are not aware of the difference in pronunciation for the /t/-/d/ word pairs, but score highly accurate on *td*-item identification. Those who thought word-final /t/ and /d/ sound the same could still perceive the difference. But awareness does seem to help: if a participant scores high on the metalinguistic awareness task, they tend to score high on the identification task as well.

A Pearson product-moment correlation coefficient was computed to assess the relationship between the identification scores and the metalinguistic knowledge in the pre-tests for the items involving the word-final /t/-/d/ contrast. A medium significant correlation between the variables was found: $r(62) = 0.31$, $p = 0.012$ (with quite a large range of confidence: 95% CI 0.071 - 0.52).

4 Discussion

In the literature, there is very little evidence on how a short explicit phonetic instruction affects L2 perception. This thesis aimed to fill this gap. How can explicit phonetic instruction on the English phonetic contrasts /æ/-/ɛ/ and word-final /t/-/d/ help Dutch advanced learners of English to improve their perception of these two non-native contrasts? All participants watched one of four instruction videos: they received instruction on the /æ/-/ɛ/ contrast or on the word-final /t/-/d/ contrast, and received explanation on the duration cue or not. Pre- and post-tests measured metalinguistic awareness and perception ability (using a 2-AFC identification task) on the two contrasts before and after participants watched a video. In this section we discuss the results, their interpretation, limitations and suggestions for further research, for each research question in turn.

4.1 RQ 1 – Explicit phonetic instruction

The first research question was formulated as follows: does a short explicit phonetic instruction improve the perception skills of Dutch listeners of English, for the English contrasts /æ/-/ɛ/ and word-final /t/-/d/? In line with the hypothesis, the results showed that explicit phonetic instruction helped learners to perform better on the 2-AFC identification task. The group who had watched an *ae*-video improved significantly on the *ae*-items, significantly more compared to the group who had watched a *td*-video. Similarly, those who had watched a *td*-video improved significantly on the *td*-items, significantly more compared to the group who had watched an *ae*-video. These results indicate that explicit instruction can lead to an improvement in the perception of a non-native contrast, which is in line with previous findings by Kissling (2012). Directing attention to a non-native contrast through videos may have increased participants' alertness and orientation on that contrast, which helped them to detect the sounds more accurately (Ellis, 2011; Tomlin & Villa, 1994).

The proficiency measures from the questionnaire (self-rated proficiency in listening, overall self-rated proficiency and exposure to English in daily life) had a significant effect on the participants' identification accuracy. Those who had given themselves a higher proficiency rate or indicated a greater listening experience with English performed significantly better on the identification task. Due to time limitations, no objective proficiency measure was used. Although self-rated proficiency measures are subjective, they provide holistic judgements and have been found to be quite reliable (Luk & Bialystok, 2013), future research could include an objective standardised proficiency measure, in combination with self-rated proficiency, as advised by Luk and Bialystok (2013).

Follow-up studies could investigate, for instance, whether the learning effect is retained over a longer time, whether the learning effect can be replicated with different phonetic contrasts of different languages, whether instruction on perception could affect production, and whether the learning effect is corelated to short-term memory. The present study intended to facilitate a learning effect, keeping variability low and keeping pronunciation realisations comparable between stimuli and video explanation, using only one speaker to record all stimuli as well as the video. Future research could investigate whether the learning effects using explicit phonetic instruction can also be found for a task involving multiple speakers for the stimuli and/or the video instruction. This higher variability in the realisations of the contrasts could improve the generalisability of listeners' perception skills (see e.g. Perrachione, Lee, Ha, & Wong, 2011). Lastly, in real-world contexts, explicit instruction can be used in an L2 classroom. The effect of this explicit phonetic instruction on listening proficiency would be interesting to measure in a classroom setting, even though this is a less controlled environment.

4.2 RQ 2 – The duration cue

The /æ/ typically has a longer duration than /ɛ/, and the vowel preceding word-final /d/ typically has longer duration than the vowel before word-final /t/. Native speakers have been shown to rely on this duration cue to distinguish these sounds, whereas Dutch listeners do not rely on it (Broersma, 2005, 2008) and use it less (at least for word-final fricatives) (Broersma, 2010). The second research question asked: if the explicit phonetic instruction helps improve the English perception skills of Dutch listeners, how much of this effect is due to the presence of explanation of the duration cue? It was hypothesised that the videos that included an explanation on the duration cue would help listeners improve their accuracy on the 2-AFC identification task, more than an instruction video that did not include this explanation on the duration cue.

The results showed that participants that watched a duration video had no advantage over those that watched a no-cue video in the 2-AFC identification task. Both the videos that included an explanation on the duration cue and the videos that did not include this explanation led to similar improvements in the identification task, for both the /ae-/ɛ/ and the /t/-/d/ contrast. Both the duration and no-cue videos were successful at improving participants' perception of the contrasts. Telling Dutch listeners explicitly about the duration cue did not lead to different results – listeners did not seem to need this explicit knowledge about the duration cue.

There are several possible explanations for these findings. Possibly, the differences in vowel duration in the naturally pronounced stimuli were too subtle for Dutch speakers to pick up on. The present study did not artificially control for vowel length or vowel quality. The items in the identification task were naturally produced, not phonetically controlled. Additionally, it is also possible that participants used other cues that were briefly mentioned in both duration and no-cue videos, like the /t/ release burst or the subtle difference in vowel quality for /æ/-/ɛ/. It is also possible that directing listeners' attention to duration may not have been enough for listeners to be able to apply this knowledge immediately. The post-test immediately followed the video, so it is possible that listeners may have needed more time, or more practise, to be able to effectively use this cue.

Both videos with and without explanation on the duration cue were successful at directing attention to the contrasts, and it is possible that this was all participants needed to take away from the videos to improve on the identification task. All videos exposed participants to exaggerated minimal pairs and challenged participants to listen carefully – their attention to the contrast may have been enough. This could be similar to the findings in (Guion & Pederson, 2007), who found that directing participants' attention to sound in general was enough for them to improve on a discrimination task.

In conclusion, the results implicate that explaining the duration cue was not essential to improve learners' perception ability. More research is needed to uncover which of the explanations discussed above is most likely. Future research could also measure whether Dutch listeners rely more on the duration cue after watching a video with duration cue explanation, as this was not measured in this study.

4.3 RQ 3 – Asymmetry

To answer the third research question, we looked into a possible asymmetry in recognition of the /æ/ versus /ε/ and word-final /t/ versus /d/ sounds. Previous studies have reported an asymmetry in the recognition of words containing /ε/ versus words containing /æ/ (Cutler et al., 2006; Díaz et al., 2012; Weber & Cutler, 2004). This asymmetry could be due to a ‘goodness of fit’ difference between /ε/ and /æ/, so that /ε/ may be a better fit to the L1 category than /æ/ - /ε/ could be ‘dominant’ over /æ/ in word recognition (Best & Tyler, 2007). However, as not all studies reported this asymmetry (Broersma, 2005)’, it was not certain whether any asymmetries would be found for the /æ/-/ε/ and word-final /t/-/d/ contrast.

Results indicated no significant difference in the accuracy of /ε/-items versus /æ/-items on the identification task, although the descriptives showed a tendency in line with the hypothesis that /ε/-items may be dominant over /æ/-items. Additionally, the analysis for the /t/-/d/ contrast did show an asymmetry: there was a significant difference in accuracy between /t/-items and /d/-items. The word-final /t/ was recognised correctly significantly more often than word-final /d/. Because the /d/ sound does not occur word-finally in Dutch, a word-final /d/ could logically be more difficult to recognise.

A possible explanation of the lack of asymmetry for the /æ/-/ε/ and the precence of such asymmetry for the /t/-/d/ contrast is that the dominance of /ε/ over /æ/ may be more subtle than the dominance of /t/ over /d/. It is possible that the asymmetry was there for /æ/-/ε/, but the identification task was not sensitive enough to catch it. Another explanation is that the asymmetry for the /æ/-/ε/ may be less likely, if it is due to ‘goodness of fit’, because the Dutch /ε/ is actually in-between English /ε/ and /æ/. So the English /ε/ could be perceived as close to /ε/ as it is to /æ/. But the asymmetry could be caused by other factors, see Section 1.2.4. However, it is still possible that asymmetries are not actually there at all, and there might be a publication bias for positive findings on these asymmetries (Broersma, 2005). Continued research on this topic is needed to show whether the asymmetry for the /t/-/d/ contrast reported here can be replicated, and whether the asymmetry for /æ/-/ɛ/ is there or not.

4.4 RQ 4 – Metalinguistic awareness

The fourth research question explored the metalinguistic awareness of the participants on the /æ/-/ɛ/ contrast and the word-final /t/-/d/ contrast. Is there a correlation between Dutch L2 users of English metalinguistic awareness on the pronunciation differences between /æ/-/ɛ/ or word-final /t/-/d/ minimal pairs and the perception skills of these Dutch L2 users of English regarding the same minimal pairs? How does awareness relate to perception ability? Theory suggests that if listeners are already aware of the difference in pronunciation between the contrasts, this awareness could benefit detection (Ellis, 2011; Tomlin & Villa, 1994). However, detection can also take place without awareness (Tomlin & Villa, 1994).

A correlation analysis of the mean accuracy scores for awareness and pre-test identification accuracy showed a medium correlation for both the /æ/-/ɛ/ contrast and the word-final /t/-/d/ contrast. As expected, participants that showed a high awareness (i.e., those who knew that words that differ in /æ/-/ɛ/ or word-final /t/-/d/ sound different) also scored high on the identification task on these contrasts. However, many participants scored low on metalinguistic awareness, but performed very well on the identification task. Thus, metalinguistic awareness was not strictly necessary for successful L2 perception. L2 learners do not necessarily need to be aware that phonetic differences are there to be able to hear them. Therefore, these findings are in line with the theory by Tomlin and Villa (1994): detection can take place without awareness.

To my knowledge, this study is the first to investigate this relationship between metalinguistic awareness and its relation to perception ability empirically. The results reported here still leave some questions unanswered, which are opportunities for further research. The results indicate a similar relationship between awareness and perception ability for both the /æ/-/ɛ/ and the /t/-/d/ contrast.

Future research could investigate whether this pattern is visible for other language combinations and phonetic contrasts. Future research could explore different methods to measure metalinguistic awareness. Additionally, future research should explore the relationship between metalinguistic awareness and explicit instruction further (i.e. even though explicit instruction may raise awareness, a lot of variability was found in the accuracy scores of the metalinguistic awareness task, even after watching the videos and doing the identification tasks).

4.5 The big picture

The perception of non-native phonetic contrasts is difficult when the L1 phonological inventory differs from the L2 as predicted by the PAM-L2 (for Dutch listeners: English /æ/-/ɛ/ contrast) (Best, 1994; Best & Tyler, 2007). It is also difficult when phonotactic constraints of the L1 differ from those of the L2 (for Dutch listeners: English word-final /t/-/d/ contrast) (Broersma, 2005). To overcome these difficulties with L2 perception, the findings of the present study indicate that videos with explicit phonetic instruction helped Dutch L2 listeners to better distinguish between /æ/-/ɛ/ and word-final /t/-/d/ minimal pairs. This finding is in line with previous evidence that explicit instruction on perception can direct attention and drive learning effects (Chen & Pederson, 2017; Guion & Pederson, 2007; Pederson & Guion-Anderson, 2010). Little research had previously investigated the effects of explicit phonetic instruction on non-native phonetic contrasts, especially for experienced L2 users instead of naïve listeners, though their results were promising (Hisagi & Strange, 2011; Kissling, 2012; Porretta & Tucker, 2014). This thesis makes a contribution to this existing research, showing that explicit instruction can help L2 users to overcome the perception difficulties. The findings of this thesis shed more light on the mechanisms of non-native listeners' perception of non-native sounds and how their perception skills can be improved by explicit instruction. Future research may show the usefulness of explicit phonetic instruction for L2 listening outside the lab in the L2 classroom.

5 Conclusion

This thesis found evidence that explicit phonetic instruction on the English /æ/-/ɛ/ and word-final /t/-/d/ contrasts can improve Dutch listeners' ability to identify the right word in a 2-AFC identification task. The explanation on the duration cue, which native listeners use to distinguish these non-native contrasts, was not essential to the learning effect. Likely, the instructions with or without duration cue were both successful at directing participants' attention to the phonetic contrast, which was enough to result in a learning effect. Additionally, the present study found that the word-final /d/ was often miscategorised as a /t/, more often than word-final /t/ was miscategorised as /d/. This may indicate there is an asymmetry for this contrast in word processing; however, no asymmetry was found for the /æ/-/ɛ/ contrast. Lastly, the results showed that metalinguistic awareness can positively affect perception abilities, but that it is not a prerequisite for correct perception.

References

Adank, P., Van Hout, R., & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical Society of America, 116*(3), 1729-1738. doi:10.1121/1.1779271

Ainsworth, W. A. (1972). Duration as a cue in the recognition of synthetic vowels. *The Journal of the Acoustical Society of America, 51*(2B), 648-651. doi:10.1121/1.1912889

Aoyama, K., Flege, J. E., Guion, S. G., Akehana-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics, 32*, 233-250. doi:10.1016/S0095-4470(03)00036-6

Audacity-Team. (2018). Audacity®. Audio editor and recorder. Available from: http://audacityteam.org/ (Version 2.2.2).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01

Bennett, D. C. (1968). Spectral form and duration as cues in the recognition of English and German vowels. *Language and Speech*, *11*, 65-85.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America, 109*(2), 775-794. doi:10.1121/1.1332378

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

Black, J. W. (1949). Natural frequency, duration and intensity of vowels in reading. *Journal of Speech, Language, and Hearing Research*, *14*, 216-221.

Booij, G. (1995). *The phonology of Dutch*. Oxford: Oxford University Press.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299-2310. doi:10.1121/1.418276

Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *The Journal of the Acoustical Society of America, 117*(6), 3890-3901. doi:10.1121/1.1906060

Broersma, M. (2008). Flexible cue use in nonnative phonetic categorization. *The Journal of the Acoustical Society of America, 124*(2), 712-715. doi:10.1121/1.2940578

Broersma, M. (2010). Perception of final fricative voicing: native and nonnative listeners' use of vowel duration. *The Journal of the Acoustical Society of America, 127*(3), 1636-1644. doi:10.1121/1.3292996

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes, 27*, 1205-1224. doi:10.1080/01690965.2012.660170

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System, 36*(1), 22-34. doi:10.1016/j.system.2007.11.003

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *The Quarterly Journal of Experimental Psychology, 64*(1), 74-95. doi:10.1080/17470218.2010.499174

Byrd, D. (1993). 54 000 American stops. *UCLAWorking Papers, 83*, 97-115.

Chen, Y., & Pederson, E. (2017). *Directing Attention During Perceptual Training: A Preliminary Study of Phonetic Learning in Southern Min by Mandarin Speakers*. Paper presented at the Interspeech 2017.

Crystal, T. H., & House, A. S. (1988). The duration of American-English vowels: an overview. *Journal of Phonetics, 16*, 263-284.

Cutler, A., & Broersma, M. (2005). Phonetic precision in listening. In W. J. Hardcastle & J. M. Beck (Eds.), *A Figure of Speech: A Festschrift for John Laver* (pp. 63-91). Mahwah, NJ: Erlbaum.

Cutler, A., & Otake, T. (2004). *Pseudo-homophony in non-native listening*. Paper presented at the The 147th Meeting of the Acoustical Society of America, New York, New York.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*(2), 269-284. doi:10.1016/j.wocn.2005.06.002

DeKeyser, R. M. (1994). How implicit can adult second language learning be? In J. H. Hulstijn & R. Schmidt (Eds.), *Consciousness in second language learning, AILA Review* (Vol. 11, pp. 83-96).

Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association, 27*, 47-55.

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences, 22*(6), 680-689. doi:10.1016/j.lindif.2012.05.005

Ellis, N. C. (2011). Implicit and explicit SLA and their interface. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning: conditions, processes and knowledge in SLA and bilingualism*. Washington, DC: Georgetown University Press.

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*. Timonium, MD: York Press.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production* (pp. 319-355). Berlin-New York: Mouton de Gruyter.

Flege, J. E. (2005). *Origins and development of the Speech Learning Model*. Keynote Lecture at 1st ASA Workshop on L2 Speech Learning. Simon Fraser University. Vancouver, BC.

Flege, J. E., & Hillenbrand, J. (1987). A differential effect of release bursts on stop voicing judgments of native French and English listeners. *Journal of Phonetics*, *15*, 203-208.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics*, *62*(8), 1668-1680. doi:10.3758/BF03212164

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 349-366. doi:10.1037/0096-1523.28.2.349

Grenon, I., Sheppard, C., & Archibald, J. (2018). *Discrimination training for learning sound contrasts*. Paper presented at the ISAPh 2018 International Symposium on Applied Phonetics, 19-21 September 2018, Aizuwakamatsu, Japan.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O. S. Bohn & M. Munro (Eds.), *Language experience in second language speech learning* (pp. 57-77). Amsterdam: John Benjamins.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Acoustical Society of America, 108*(6), 3013-3022. doi:10.1121/1.1323463

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97*(5), 3099-3111. doi:10.1121/1.411872

Hillenbrand, J. M., Ingrisano, D. R., Smith, B. L., & Flege, J. E. (1984). Perception of the voiced-voiceless contrast in syllable-final stops. *Journal of the Acousical Society of America*, *76*(1), 18-26. doi:10.1121/1.391094

Hirata, Y., Kelly, S. D., Huang, J., & Manansala, M. (2014). Effects of hand gestures on auditory learning of second-language vowel length contrasts. *Journal of Speech, Language, and Hearing, 57*(6), 2090-2101. doi:10.1044/2014_JSLHR-S-14-0049

Hisagi, M., & Strange, W. (2011). Perception of Japanese temporally-cued contrasts by American English listeners. *Language and Speech, 54*(2), 241–264. doi:10.1177/0023830910397499

Hogan, J. T., & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *J Acoust Soc Am, 67*(5), 1764-1771. doi:10.1121/1.384304

House, A. S. (1961). On vowel duration in English. *The Journal of the Acoustical Society of America, 33*(9), 1174-1178. doi:10.1121/1.1908941

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America, 25*(1), 105-113. doi:10.1121/1.1906982

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America, 126*(2), 866-877. doi:10.1121/1.3148196

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: a comparison of methods for teaching English /r/-/l/ to Japanese adults. *J Acoust Soc Am, 118*(5), 3267-3278. doi:10.1121/1.2062307

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics, 33*(1), 145-160. doi:10.1017/S0142716411000300

Kissling, E. M. (2012). *The effect of phonetics instruction on adult learners'perception and production of L2 sounds.* (Doctor of Philosophy in Linguistics Dissertation). Georgetown University, Washington, DC.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *The Journal of the Acoustical Society of America, 59*(5). doi:10.1121/1.380986

Kondaurova, M. V., & Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. *Journal of Phonetics, 38*(4), 569-587. doi:10.1016/j.wocn.2010.08.003

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

Kuhl, P. K., & Iverson, P. (1995). Linguistic Experience and the "perceptual magnet effect". *Speech perception and linguistic experience: Issues in cross-language research*, 121-154.

Lacabex, E. G., & Gallardo del Puerto, F. (2014). Two phonetic-training procedures for young learners: investigating instructional effects on perceptual awareness. *Canadian Modern Language Review, 70*(4), 500-531. doi:10.3138/cmlr.2324

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology, 25*(5), 605-621. doi:10.1080/20445911.2013.795574

McQueen, J. M. (2004). Speech perception. In K. Lamberts & R. Goldstone (Eds.), *The handbook of cognition* (pp. 255-275). London: Sage.

Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *The Journal of the Acoustical Society of America, 127*(2), EL54-59. doi:10.1121/1.3292286

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., . . . Lindeløv, J. K. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195-203. doi:10.3758/s13428-018-01193-y

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America, 130*(1), 461-472. doi:10.1121/1.3593366

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acousical Society of America, 32*, 693-703.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 421-435. doi:10.1037/0096-1523.20.2.421

Porretta, V. J., & Tucker, B. V. (2014). Perception of non-native consonant length contrast: The role of attention in phonetic processing. *Second Language Research, 31*(2), 239-265. doi:10.1177/0267658314559573

R-Core-Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *The Journal of the Acoustical Society of America, 51*(4B), 1296-1303. doi:10.1121/1.1912974

Raphael, L. J., Dorman, M., Freeman, F., & Tobin, C. (1975). Vowel and nasal duration as cues to voicing in word-final stop consonants: spectrographic and perceptual studies. *Journal of Speech, Language, and Hearing Research, 18*, 389-400.

Raphael, L. J., Dorman, M., & Liberman, A. M. (1980). On defining the vowel duration that cues voicing in final position. *Language and Speech, 23*.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. doi:10.1093/applin/11.2.129

Slis, I. H., & Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction I. *Language and Speech, 12*, 80-102.

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 (Vol. 36, pp. 153-191). Amsterdam-Philadelphia: John Benjamins Publishing Company.

Svalberg, A. M. L. (2007). Language awareness and language learning. *Language Teaching, 40*(4), 287-308. doi:10.1017/s0261444807004491

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and slecond language acquisition. *Studies in Second Language Acquisition*, *16*, 183-203.

Van den Berg, R. J. H. (1989). Perception of voicing in Dutch two obstruent sequences: Covariation of voicing cues. *Speech Communication*, *8*, 17-25.

Van der Feest, S. V., & Swingley, D. (2011). Dutch and English listeners' interpretation of vowel duration. *The Journal of the Acoustical Society of America, 129*(3), EL57-EL63. doi:10.1121/1.3532050

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*, 1176-1190.

Van Maastricht, L., Hoetjes, M., & Van Drie, E. (2019). *Do gestures during training facilitate L2 lexical stress acquisition by Dutch learners of Spanish?* Paper presented at the International Conference on Auditory-Visual Speech Processing (AVSP2019), Melbourne, Australia.

Van Santen, J. P. H. (1992). Contextual effects on vowel duration. *Speech Communication, 11*, 513-546.

Wardrip-Fruin, C. (1982). On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants. *The Journal of the Acoustical Society of America, 71*(1), 187-195. doi:10.1121/1.387346

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*, 1-25. doi:10.1016/S0749-596X(03)00105-0

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197-234. doi:10.1207/s15473341lld0102_4

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganisation suring the first year of life. *Infant Behaviour and Development, 7*, 49-63.

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499.* Retrieved from [http://arxiv.org/pdf/1308.5499.pdf]

Appendix A − Video Scripts

1. ae-dur: *ae*-video with duration explanation

Hi! I'm Emily, and I'm a native speaker of English. In this video, I'm going to teach you about the difference between two sounds in English: the A sound and the E sound. They may sound similar, but these two sounds make an important distinction in English. For example, the difference between A and E distinguishes words like pan and pen, and jam and gem. Do you think it's hard to hear? The A sound is usually spelled with the letter a as in map, while the E sound is usually spelled with the letter e as in desk. The sounds A and E differ in the color, or quality, of their sound, but that's very subtle. What really helps to hear the difference is paying attention to how long the sound is. Listen closely to these examples, in which I exaggerate the difference: [gestures] Pen. Paaan. Gem. Jaaam. Now I'm going to pronounce the words more normally. If you listen carefully, you'll hear that the A sound is longer than the E sound. Try to hear the difference between pan, pen, pan, pen, pan, pen. Can you hear the difference in another word pair? Listen to jam, gem, jam, gem, jam, gem. In short, just remember: the A sounds longer while the E sounds shorter. I hope that helps you!

2. ae-nocue: *ae*-video without duration explanation

Hi! I'm Emily, and I'm a native speaker of English. In this video, I'm going to teach you about the difference between two sounds in English: the A sound and the E sound. They may sound similar, but these two sounds make an important distinction in English. For example, the difference between A and E distinguishes words like pan and pen, and jam and gem. Do you think it's hard to hear? The A sound is usually spelled with the letter a as in map, while the E sound is usually spelled with the letter e as in desk. The sounds A and E differ in the color, or quality, of their sound, but that's very subtle. Native speakers can hear the difference between the A sound and the E sound very easily, but for people who speak English as a second language, it can be difficult. Listen closely to these examples, in which I exaggerate the difference: Pen. Pan. Gem. Jam. Now I'm going to pronounce the words more normally. Try to hear the difference between pan, pen, pan, pen, pan, pen. Can you hear the difference in another word pair? Listen to jam, gem, jam, gem, jam, gem. It will become easier to hear the difference between A and E, the more you practice listening. I hope that helps you!

3. td-dur: *td*-video with duration explanation

Hi! I'm Emily, and I'm a native speaker of English. In this video, I'm going to teach you about the difference between two sounds in English: the T sound and D sound at the end of a word. They may sound similar, but these two sounds make an important distinction at the end of a word in English. For example, the difference between T and D distinguishes words like not and nod, and bit and bid. Do you think it's hard to hear? If a word ends with t or t-e , the sound is always T as in sit; if a word ends with d or d-e, the sound is nearly always D as in did. The T sound comes with a little puff of air, while the D sound does not, but that's very subtle. What really helps to hear the difference is paying attention to how long the vowel before it sounds. Listen closely to these examples, in which I exaggerate the difference: [gestures] Not. Noood. Bit. Biiiiid. Now I'm going to pronounce the words more normally. If you listen carefully, you'll hear that the vowel before the D sound is longer than the vowel before the T sound. Try to hear the difference between bit, bid, bit, bid, bit, bid. Can you hear the difference in another word pair? Listen to not, nod, not, nod, not, nod. In short, just remember: if the vowel is longer, you're usually hearing a D; if the vowel is shorter, you're usually hearing a T. I hope that helps you!

4. td-nocue: *td*-video without duration explanation

Hi! I'm Emily, and I'm a native speaker of English. In this video, I'm going to teach you about the difference between two sounds in English: the T sound and D sound at the end of a word. They may sound similar, but these two sounds make an important distinction at the end of a word in English. For example, the difference between T and D distinguishes words like not and nod, and bit and bid. Do you think it's hard to hear? If a word ends with t or t-e, the sound is always T as in sit; if a word ends with d or d-e, the sound is nearly always D as in did. The T sound comes with a little puff of air, while the D sound does not, but that's very subtle. Native speakers can hear the difference between the T sound and the D sound at the end of a word very easily, but for people who speak English as a second language, it can be difficult. Listen closely to these examples, in which I exaggerate the difference: Not. Nod. Bit. Bid. Now I'm going to pronounce the words more normally. Try to hear the difference between bit, bid, bit, bid, bit, bid. Can you hear the difference in another word pair? Listen to not, nod, not, nod, not, nod. It will become easier to hear the difference between T and D at the end of a word, the more you practice listening. I hope that helps you!

Appendix B − List of stimuli

1. Critical minimal pairs /ae/-/e/ (20)

Word	ZipfSUBTLEX-US	DomPOS	Length	Word	ZipfSUBTLEX-US	DomPOS	Length	MeanZipf
and	7,126116	Conjunction	3	end	5,424094	Noun	3	6,275105
bad	5,735957	Adjective	3	bed	5,271565	Noun	3	5,503761
bag	4,972804	Noun	3	beg	4,706974	Verb	3	4,839889
bat	4,314263	Noun	3	bet	5,234537	Verb	3	4,7744
cattle	4,121138	Noun	6	kettle	3,450197	Noun	6	3,785667
dad	5,704647	Noun	3	dead	5,651651	Adjective	4	5,678149
expanse	2,371015	Noun	7	expense	3,872759	Noun	7	3,121887
flash	4,18615	Noun	5	flesh	4,343373	Noun	5	4,264761
gas	4,830659	Noun	3	guess	5,65646	Verb	5	5,24356
had	6,223663	Verb	3	head	5,569398	Noun	4	5,89653
lag	3,172648	Noun	3	leg	4,751679	Noun	3	3,962163
land	4,944564	Noun	4	lend	4,06708	Verb	4	4,505822
man	6,265581	Noun	3	men	5,570862	Noun	3	5,918221
mansion	3,810348	Noun	7	mention	4,77385	Verb	7	4,292099
mantle	3,337157	Noun	6	mental	4,293135	Adjective	6	3,815146
radish	2,796984	Noun	6	reddish	2,333227	Adjective	7	2,565105
sad	4,80144	Adjective	3	said	6,044128	Verb	4	5,422784
sand	4,307194	Noun	4	send	5,254203	Verb	4	4,780699
than	5,867945	Conjunction	4	then	6,172459	Adverb	4	6,020202
track	4,745764	Noun	5	trek	3,653562	Name	4	4,199663
MEAN	4,681802		4,2		4,80476		4,45	**4,743281**
SD	1,23952		1,472556		0,978139		1,431782	**1,01073**
MAX	7,126116		7		6,172459		7	**7,126116**
MIN	2,371015		3		2,333227		3	**2,333227**

2. Critical minimal pairs /t/-/d/ (20)

Word	**ZipfSUBTLEX-US**	**DomPOS**	**Length**	**Word**	**ZipfSUBTLEX-US**	**DomPOS**	**Length**	**MeanZipf**
beat	5,119009	Verb	4	bead	3,055262	Noun	4	4,087136
bright	4,647094	Adjective	6	bride	4,383853	Noun	5	4,515473
built	4,614053	Verb	5	build	4,681532	Verb	5	4,647793
cart	3,956476	Noun	4	card	4,931121	Noun	4	4,443799
fate	4,430453	Noun	4	fade	3,749716	Verb	4	4,090084
feet	5,081274	Noun	4	feed	4,626891	Verb	4	4,854082
float	3,873898	Verb	5	flowed	2,848137	Verb	6	3,361017
got	6,518774	Verb	3	god	5,955177	Name	3	6,236976
great	5,913685	Adjective	5	grade	4,466185	Noun	5	5,189935
greet	3,721586	Verb	5	greed	3,682769	Noun	5	3,702178
height	3,970353	Noun	6	hide	4,842674	Verb	4	4,406513
hurt	5,390997	Verb	4	heard	5,587973	Verb	5	5,489485
right	6,602377	Adverb	5	ride	5,130998	Noun	4	5,866687
seat	4,895952	Noun	4	seed	3,879545	Noun	4	4,387749
wrote	3,5599	Verb	5	road	3,7566	Noun	4	3,65825
sight	4,655822	Noun	5	side	5,302473	Noun	4	4,979148
slight	4,007837	Adjective	6	slide	4,250876	Verb	5	4,129357
spent	4,84511	Verb	5	spend	4,969259	Verb	5	4,907184
threat	4,31714	Noun	6	thread	3,713438	Noun	6	4,015289
white	5,233594	Adjective	5	wide	4,37641	Adjective	4	4,805002
MEAN	4,767769		4,8		4,409544		4,5	**4,588657**
SD	0,864078		0,833509		0,809393		0,760886	**0,732563**
MAX	6,602377		6		5,955177		6	**6,602377**
MIN	3,5599		3		2,848137		3	**2,848137**

3. Filler minimal pairs (30)

Word	**ZipfSUBTLEX-US**	**DomPOS**	**Length**	**Word**	**ZipfSUBTLEX-US**	**DomPOS**	**Length**	**MeanZipf**
like	6,601354	Preposition	4	look	6,288836	Verb	4	6,445095
true	5,403164	Adjective	4	through	5,739411	Preposition	7	5,571287
came	5,665684	Verb	4	game	5,368365	Noun	4	5,517025
boat	4,980788	Noun	4	both	5,469746	Determiner	4	5,225267
forgot	4,994781	Verb	6	forget	5,442007	Verb	6	5,218394
lesson	4,508	Noun	6	listen	5,735645	Verb	6	5,121822
left	5,684672	Verb	4	lift	4,532882	Verb	4	5,108777
play	5,549081	Verb	4	pray	4,558536	Verb	4	5,053808
better	5,899236	Adverb	6	bitter	4,046946	Adjective	6	4,973091
loose	4,620621	Adjective	5	less	5,045187	Adverb	4	4,832904
taste	4,709804	Noun	5	test	4,92419	Noun	4	4,816997
trade	4,54614	Noun	5	train	4,977486	Noun	5	4,761813
pride	4,441669	Noun	5	proud	4,921855	Adjective	5	4,681762
save	5,209812	Verb	4	shave	4,138789	Verb	5	4,674301
rest	5,327744	Noun	4	wrist	4,014468	Noun	5	4,671106
warn	4,403768	Verb	4	warm	4,716716	Adjective	4	4,560242
file	4,643437	Noun	4	fail	4,390478	Verb	4	4,516958
medal	4,063422	Noun	5	middle	4,949846	Noun	6	4,506634
rice	4,178325	Noun	4	race	4,791246	Noun	4	4,484786
note	4,728315	Noun	4	net	4,191655	Noun	3	4,459985
path	4,389785	Noun	4	bath	4,492685	Noun	4	4,441235
fry	3,924291	Verb	3	fly	4,928923	Verb	3	4,426607
run	5,544177	Verb	3	pun	3,269558	Noun	3	4,406868
cloud	4,069985	Noun	5	crowd	4,572185	Noun	5	4,321085
chase	4,51559	Noun	5	chess	3,872759	Noun	5	4,194174
bike	4,412737	Noun	4	bake	3,802379	Verb	4	4,107558
fork	3,946011	Noun	4	fort	4,18836	Name	4	4,067186
desk	4,642082	Noun	4	disc	3,412408	Noun	4	4,027245
pile	4,119849	Noun	4	pale	3,904618	Adjective	4	4,012234
glue	3,770401	Noun	4	clue	4,245594	Noun	4	4,007997
MEAN	4,783157		4,366667		4,631125		4,466667	**4,707141**
SD	0,679312		0,76489		0,707393		0,973204	**0,543751**
MAX	6,601354		6		6,288836		7	**6,601354**
MIN	3,770401		3		3,269558		3	**3,269558**

4. Filler homophones (30)

Word	ZipfSUBTLEX-US	DomPOS	Length	Word	ZipfSUBTLEX-US	DomPOS	Length	MeanZipf
there	6,637708	Ex	5	their	5,815763	Pronoun	5	6,226735
hour	5,20976	Noun	4	our	6,14569	Pronoun	3	5,677725
wood	4,431083	Noun	4	would	6,246862	Verb	5	5,338973
wait	5,918626	Verb	4	weight	4,559241	Noun	6	5,238933
son	5,613018	Noun	3	sun	4,842552	Noun	3	5,227785
knight	4,427285	Name	6	night	5,936952	Noun	5	5,182118
knot	3,568296	Noun	4	not	6,733802	Not	3	5,151049
some	6,236761	Determiner	4	sum	4,011165	Noun	3	5,123963
knows	5,388501	Verb	5	nose	4,84304	Noun	4	5,11577
made	5,74861	Verb	4	maid	4,35816	Noun	4	5,053385
way	6,153141	Noun	3	weigh	3,852936	Verb	5	5,003039
peace	4,842185	Noun	5	piece	5,094608	Noun	5	4,968396
him	6,541548	Pronoun	3	hymn	3,216113	Noun	4	4,878831
blew	4,49893	Verb	4	blue	5,009754	Adjective	4	4,754342
none	5,043267	Pronoun	4	nun	3,843284	Noun	3	4,443276
air	5,142542	Noun	3	heir	3,718345	Noun	4	4,430444
find	5,918995	Verb	4	fined	2,935287	Verb	5	4,427141
higher	4,444428	Adjective	6	hire	4,406779	Verb	4	4,425603
hair	5,185707	Noun	4	hare	3,58409	Noun	4	4,384899
bare	3,921244	Adjective	4	bear	4,758554	Noun	4	4,339899
sole	3,721586	Adjective	4	soul	4,885784	Noun	4	4,303685
mind	5,684812	Noun	4	mined	2,882899	Verb	5	4,283856
sail	4,138171	Verb	4	sale	4,404439	Noun	4	4,271305
tale	4,079295	Noun	4	tail	4,378194	Noun	4	4,228744
waist	3,71179	Noun	5	waste	4,725924	Verb	5	4,218857
rays	3,638187	Noun	4	raise	4,741467	Verb	5	4,189827
flew	4,187257	Verb	4	flu	3,940194	Noun	3	4,063725
seas	3,642082	Noun	4	sees	4,570588	Verb	4	4,106335
allowed	4,6511	Verb	7	aloud	3,269558	Adverb	5	3,960329
flour	3,501349	Noun	5	flower	4,35704	Noun	6	3,929195
MEAN	4,860909		4,266667		4,535635		4,3	**4,698272**
SD	2,304051		0,907187		0,962051		0,83666	**0,557703**
MAX	6,637708		7		6,733802		6	**6,733802**
MIN	3,501349		3		2,882899		3	**2,882899**

Appendix C – Questionnaire

Demographics Questionnaire

Age: _____ Gender: _______________

Name of your current study program: ____

Type of your current study program: Bachelor Master Other: ___________

Year of your current study program: 1 2 3 4

Name(s) of previous study program(s): ____

Have you ever had (check all that apply):

- ☐ vision problems?
- ☐ a language disability?
- ☐ hearing impairment?
- ☐ a learning disability?

If yes to any of the above, please explain (including any corrections, e.g., hearing aids):

____

____

Language Background Questionnaire

What is your native language (the first language you ever learned)? ___________________

If you grew up with more than one native language, please list them all:

____

What is your **mother's** native language? __________________________

What is your **father's** native language? __________________________

Please list all the languages you can speak, in order of proficiency from high to low:

1. ________________________________
2. ________________________________
3. ________________________________
4. ________________________________
5. ________________________________
6. ________________________________

For most Dutch students, English is the first foreign language they learned. If you started learning any other foreign language(s) *before* English, please indicate those here:

____

How old were you when you first began **listening to English?** ________ years

How old were you when you first began **speaking in English?** ________ years

How old were you when you first began **learning English in school?** _______ years

How many total years of **English language-focused instruction** have you received, from all your levels of education combined? ________ years

Are you currently taking any courses taught in English? If so, please list them here:

____

____

Have you ever taken any linguistics courses? If so, please list them here:

____

____

____

Have you ever lived in or visited (for 1 month or more) a country where **English** is the primary language? If so, please give the following information about your experiences:

Country	**Year of Visit**	**Length of visit or residence (in months and/or years)**

Please rate your level of **English proficiency** in the following four skills:

	No ability	**Low**	**Adequate**	**Good**	**Excellent**	**Perfect**
Speaking	0	1	2	3	4	5
Listening	0	1	2	3	4	5
Writing	0	1	2	3	4	5
Reading	0	1	2	3	4	5

When you speak English, how strong of a **Dutch accent** do you think you have?

No accent at all	**Very light accent**	**Light accent**	**Moderate accent**	**Heavy accent**	**Very heavy accent**
0	1	2	3	4	5

How frequently do you think others identify you as a non-native speaker of English **based on your accent?**

Never	**Rarely**	**Sometimes**	**Often**	**Usually**	**Always**
0	1	2	3	4	5

How many hours do you spend actively **listening to** English in a typical week? (Think of time you spend watching movies and TV shows in English, attending English-taught lectures, participating in English conversations, etc.):

_____________ hours of English listening per week

How many hours do you spend **speaking in** English in a typical week? (Think of time you spend speaking English in class, at work, in sports/clubs, with friends or family, etc.)

_____________ hours per of English speaking per week

Debriefing Qustionnaire

You may answer in English or in Dutch.

Did you use a particular strategy for the listening task **before** watching the video? If so, please explain.

Did you use a particular strategy for the listening task **after** watching the video? If so, please explain.

How well do you think you understood the explanation in the video?

Not at all	Very little	Somewhat	Well	Very Well	Perfectly
0	1	2	3	4	5

Do you have any other comments about the experiment?

Appendix D − Vocabulary task results

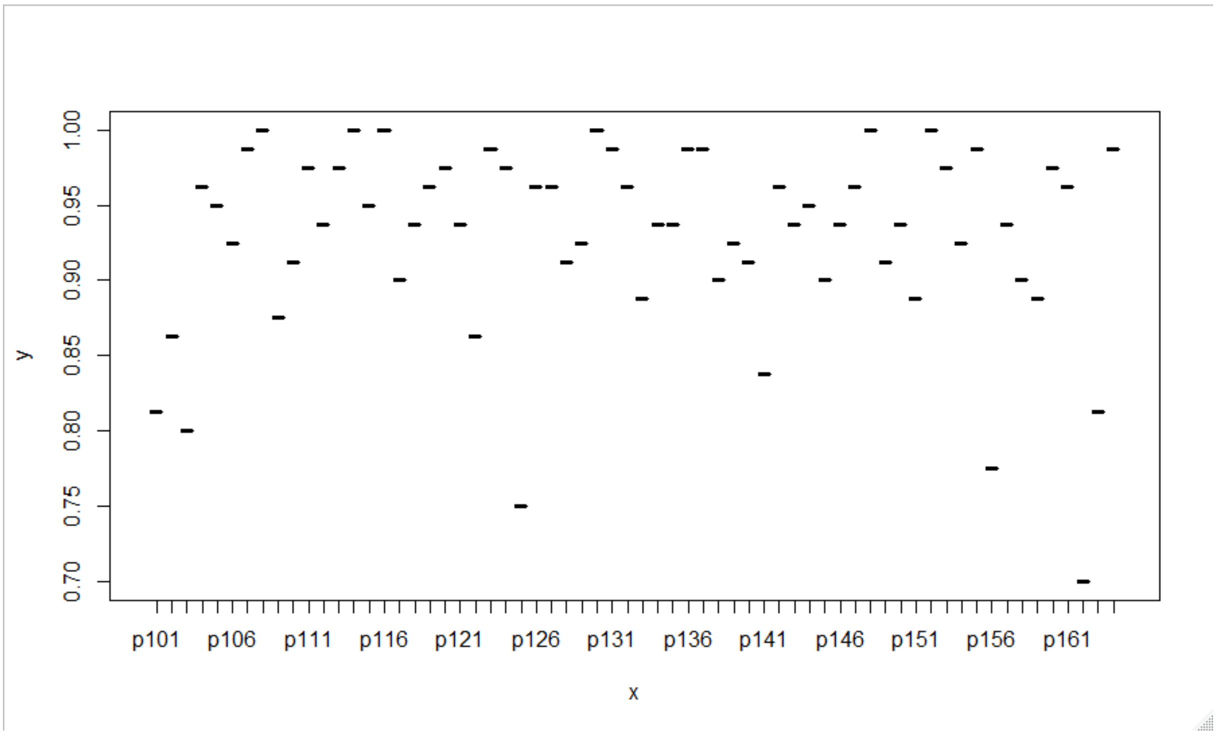

Figure 17. **Vocabulary scores for all participants, range 70-100%.**

When the data was categorised for word difficulty, most words were known by all participants (scores of 100%). There were 32 words that not all participants knew, and did not get a 100% score. The mean scores and standard deviations of these words are displayed in Table 17 below.

Table 17. **The scores for the 32 words that were known by not all participants, i.e., 'trek' was least well-known, only 31,2 percent of participants said 'yes' to knowing this word.**

	word	**mean**	**SD**
1	trek	0.312500	0.4671766
2	reddish	0.500000	0.5039526
3	bead	0.531250	0.5029674
4	lag	0.546875	0.5017331
5	radish	0.562500	0.5000000
6	cattle	0.609375	0.4917474
7	kettle	0.640625	0.4836103
8	expanse	0.687500	0.4671766

9	mantle	0.718750	0.4531635
10	flowed	0.734375	0.4451569
11	greed	0.750000	0.4364358
12	expense	0.859375	0.3503824
13	thread	0.859375	0.3503824
14	slight	0.875000	0.3333333
15	cart	0.890625	0.3145764
16	flesh	0.890625	0.3145764
17	mansion	0.890625	0.3145764
18	lend	0.906250	0.2937848
19	fate	0.921875	0.2704897
20	bat	0.937500	0.2439750
21	seed	0.937500	0.2439750
22	track	0.953125	0.2130420
23	float	0.968750	0.1753681
24	grade	0.968750	0.1753681
25	slide	0.968750	0.1753681
26	beat	0.984375	0.1250000
27	feed	0.984375	0.1250000
28	flash	0.984375	0.1250000
29	gas	0.984375	0.1250000
30	mention	0.984375	0.1250000
31	seat	0.984375	0.1250000
32	sight	0.984375	0.1250000

Appendix E – Boxplots identification task

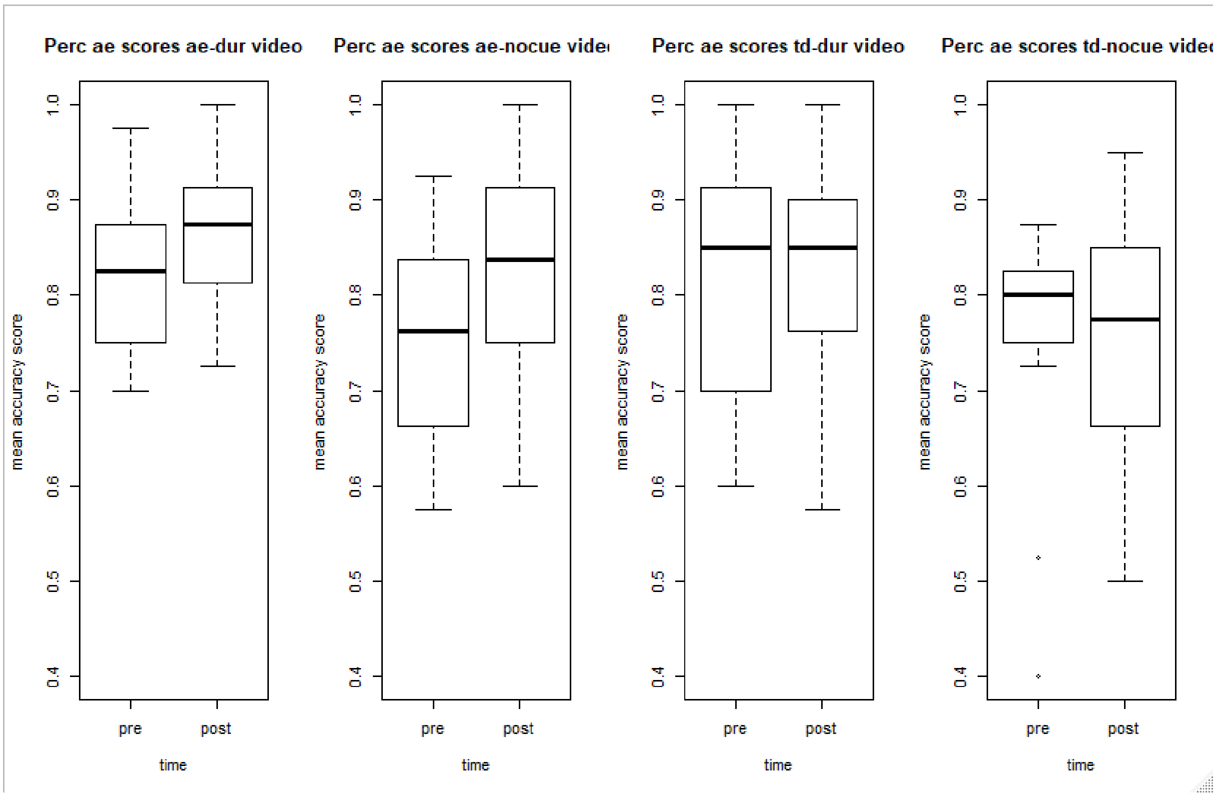

Figure 18. **Identification task scores for *ae*-items.**

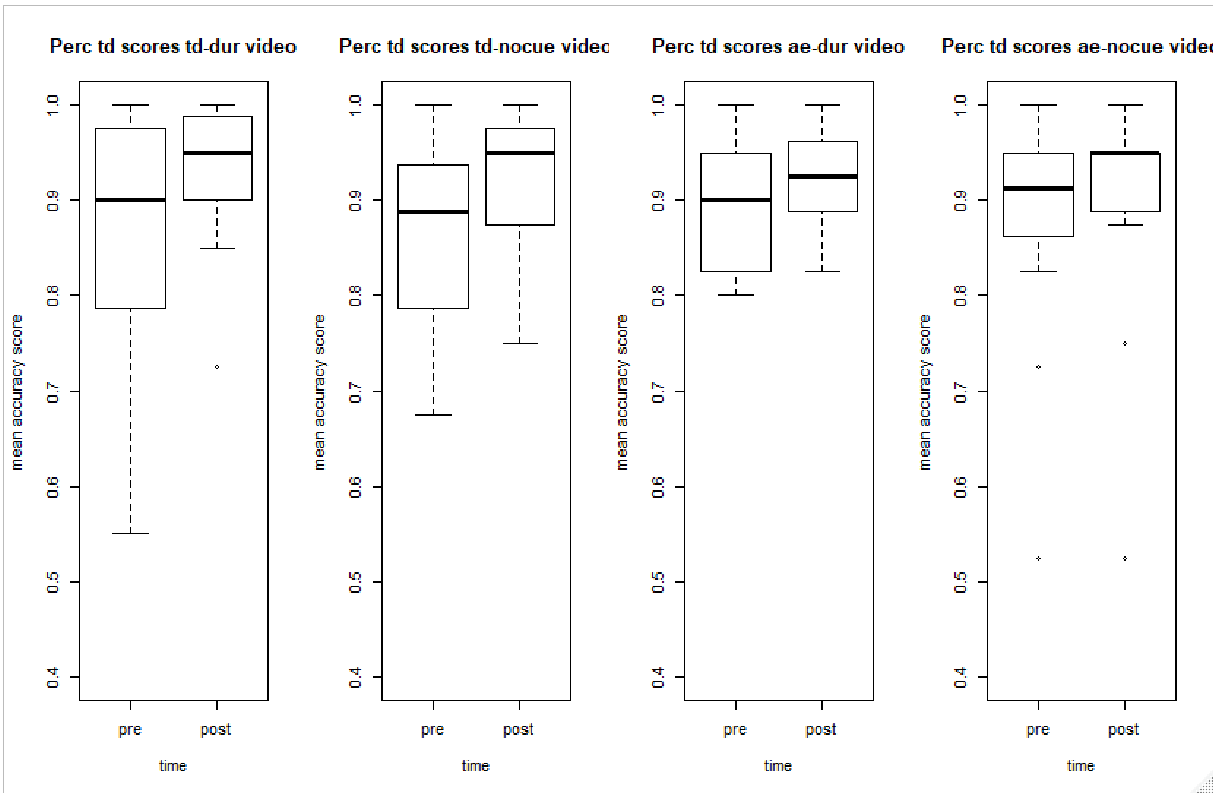

Figure 19. **Identification task scores for td-items.**

Appendix F − Boxplots metalinguistic awareness task

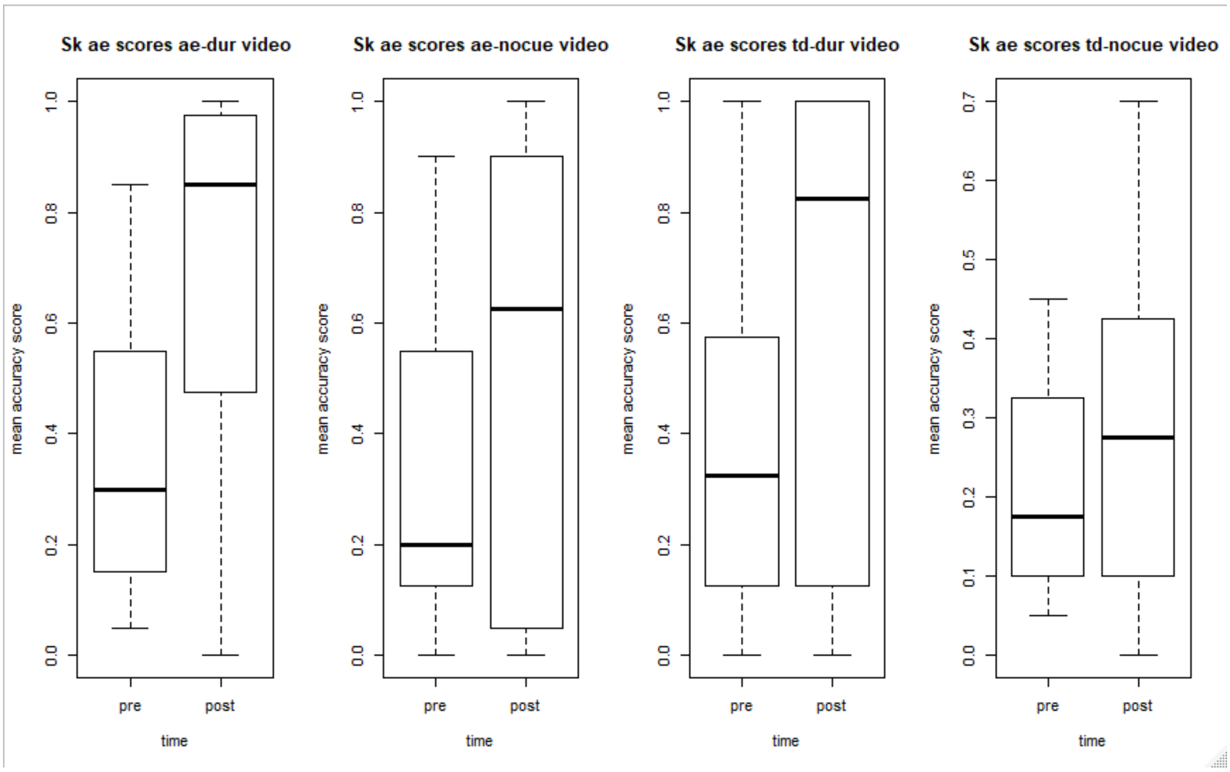

Figure 20. **Metalinguistic awareness scores for /ae/-/e/ word pairs.**

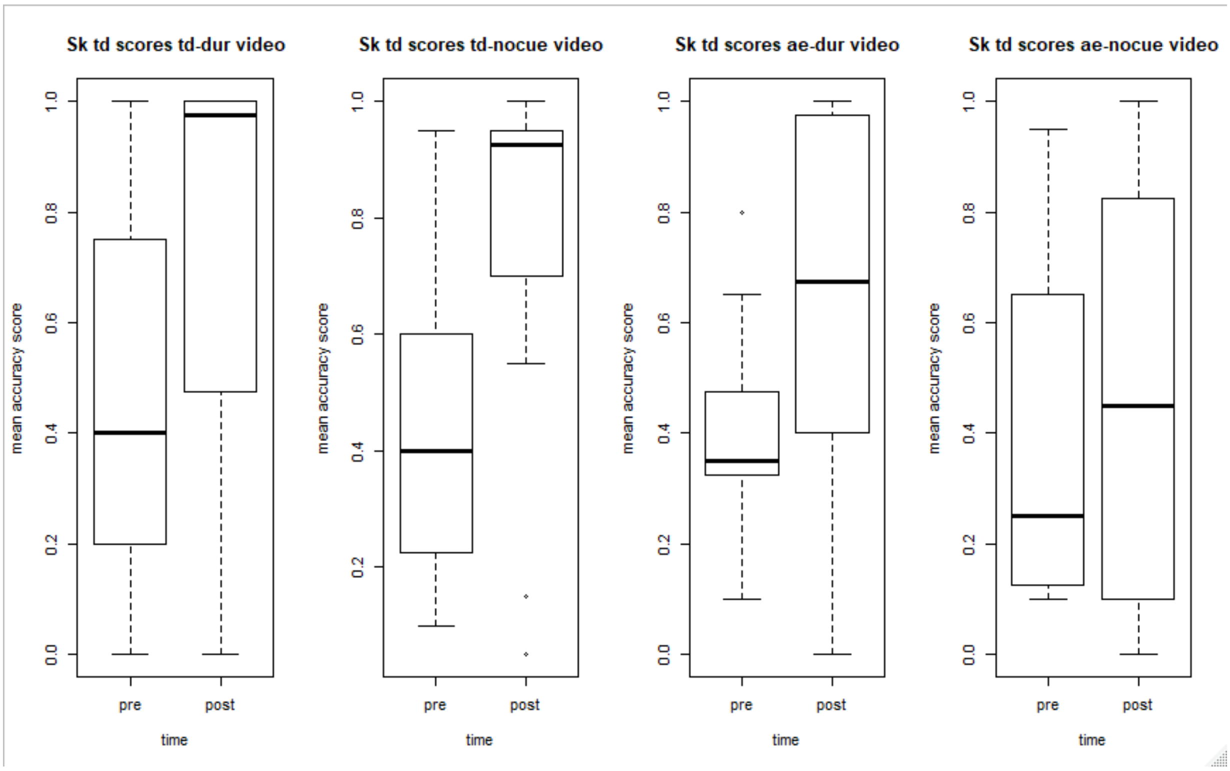

Figure 21. **Metalinguistic awareness scores for /t/-/d/ word pairs.**

Appendix G – R code

1. Analysis Section 2.1. Participants

```
rm(list = ls())      #empty environment


#------------install/load packages-----------#
library(lme4)
library(effects)
library(dplyr)
setwd(".../Data_analyses_thesis")    #set local directory


questionnaire_data <- read.csv("questionnaires_complete.csv", sep=";")
    #load questionnaire data
str(questionnaire_data)


#descriptives questionnaire
summary(questionnaire_data$age)
sd(questionnaire_data$age)


summary(questionnaire_data$isFemale)


summary(questionnaire_data$startAge_ENlearning)
sd(questionnaire_data$startAge_ENlearning)


summary(questionnaire_data$listening_proficiency)
sd(questionnaire_data$listening_proficiency)


summary(questionnaire_data$mean_proficiency)
sd(questionnaire_data$mean_proficiency)


summary(questionnaire_data$listening_hpw)
sd(questionnaire_data$listening_hpw)


summary(questionnaire_data$speaking_hpw)
sd(questionnaire_data$speaking_hpw)
```

2. Analysis Section 3.1. Explicit phonetic instruction

```
rm(list = ls())      #empty environment


#------------install/load packages-----------#
library(lme4)
library(effects)
library(dplyr)
setwd(".../Data_analyses_thesis")    #set local directory
```

```
#DATA SPLIT UP GROUPS
perception_data_split <- read.csv("perception_data_splitupgroups.csv", sep=",")
    #load perc data
perception_data_split$time <- factor(perception_data_split$time, levels = c("pre",
    "post"))     # order time pre->post


perc_data_summary <- perception_data_split %>%
  group_by(video_contrast, video_cue, time) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))




#make AE data frame (split up data)
perception_data_ae <- subset(perception_data_split, contrast_type == "ae")   #make
    subset ae contrast


data_summary_ae <- perception_data_ae %>%
    #calculate means and sds for plot
  group_by(video_contrast, time) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))




#make TD data frame (split up data)
perception_data_td <- subset(perception_data_split, contrast_type == "td")


data_summary_td <- perception_data_td %>%
    #calculate means and sds for plot
  group_by(video_contrast, time) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))




#add questionnaire data to perception data (splitfiles)
questionnaire_data2 <- read.csv("questionnaires_complete2.csv", sep=";")
    #load questionnaire data 4 columns


perc_quest_split <- merge(perception_data_split, questionnaire_data2)


perc_quest_split_ae <- merge(perception_data_ae, questionnaire_data2)
perc_quest_split_td <- merge(perception_data_td, questionnaire_data2)




#3.1.1


#BOXPLOT AE items: ae video vs td video
perc_ae_aevideos <- subset(perception_data_ae, video_contrast == "ae")
    #split up perc ae data by video contrast
perc_ae_tdvideos <- subset(perception_data_ae, video_contrast == "td")


data_summary_ae_aevideos <- perc_ae_aevideos %>%
  group_by(participant, video_cue, time) %>%
```

```
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


data_summary_ae_tdvideos <- perc_ae_tdvideos %>%
  group_by(participant, video_cue, time) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


par(mfrow = c(1,2))


plot(data_summary_ae_aevideos$time, data_summary_ae_aevideos$score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Identification task: ae
    items, ae-video", ylim = c(0.4,1.0))


plot(data_summary_ae_tdvideos$time, data_summary_ae_tdvideos$score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Identification task: ae
    items, td-video", ylim = c(0.4,1.0))




#MODEL AE
model_ae <- glmer(accuracy ~ time + video_contrast + time*video_contrast +
    (1|participant) + (1|correct_word),
                  family = binomial(link="logit"), data = perception_data_ae)
summary(model_ae)


plot(predictorEffects(model_ae, ~ time), main = "The effect of ae-video versus td-
    video on ae items", lines=list(multiline=TRUE),
     axes=list(grid=TRUE),
     confint=list(style="bars")) #try to change the label of the lines.




#MODEL AE with listening_proficiency
model_ae_lis_prof2 <- glmer(accuracy ~ time + video_contrast +
    listening_proficiency + time*video_contrast
                        + (1|participant) + (1|correct_word), family =
    binomial(link="logit"), data = perc_quest_split_ae)
summary(model_ae_lis_prof2)




#MODEL AE with mean proficiency
model_ae_mean_prof2 <- glmer(accuracy ~ time + video_contrast + mean_proficiency +
    time*video_contrast
                        + (1|participant) + (1|correct_word), family =
    binomial(link="logit"), data = perc_quest_split_ae)
summary(model_ae_mean_prof2)




#MODEL AE with listening_hpw
model_ae_lis_hpw2 <- glmer(accuracy ~ time + video_contrast + listening_hpw +
    time*video_contrast
                        + (1|participant) + (1|correct_word), family =
    binomial(link="logit"), data = perc_quest_split_ae)
summary(model_ae_lis_hpw2)
```

```
#Compare three models with each other
#compare model_ae_lis_prof2 / model_ae_mean_prof2 / model_ae_lis_hpw2


anova(model_ae_lis_prof2, model_ae_mean_prof2)
anova(model_ae_mean_prof2, model_ae_lis_hpw2)
anova(model_ae_lis_prof2, model_ae_lis_hpw2)




#3.1.2


#BOXPLOT TD items: td video vs ae video
perc_td_tdvideos <- subset(perception_data_td, video_contrast == "td")
perc_td_aevideos <- subset(perception_data_td, video_contrast == "ae")
    #split up perc td data by video contrast


data_summary_td_tdvideos <- perc_td_tdvideos %>%
  group_by(participant, video_cue, time) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


data_summary_td_aevideos <- perc_td_aevideos %>%
  group_by(participant, video_cue, time) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


par(mfrow = c(1,2))


plot(data_summary_td_tdvideos$time, data_summary_td_tdvideos$score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Identification task: td
    items, td-video", ylim = c(0.4,1.0))


plot(data_summary_td_aevideos$time, data_summary_td_aevideos$score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Identification task: td
    items, ae-video", ylim = c(0.4,1.0))




#MODEL TD
model_td <- glmer(accuracy ~ time + video_contrast + time*video_contrast +
    (1|participant) + (1|correct_word),
                  family = binomial(link="logit"), data = perception_data_td)
summary(model_td)


plot(predictorEffects(model_td, ~ time), main = "The effect of ae-video versus td-
    video on td items", lines=list(multiline=TRUE),
     axes=list(grid=TRUE),
     confint=list(style="bars")) #try to change the label of the lines.
```

```
#Model to check if there is effect of time for td items ae-videos
model_td_aevideo <- glmer(accuracy ~ time + (1|participant) + (1|correct_word),
                  family = binomial(link="logit"), data = perc_td_aevideos)
summary(model_td_aevideo)


plot(allEffects(model_td_aevideo))




#MODEL TD with listening proficiency
model_td_lis_prof2 <- glmer(accuracy ~ time + video_contrast +
    listening_proficiency + time*video_contrast + (1|participant) +
    (1|correct_word),
                         family = binomial(link="logit"), data =
    perc_quest_split_td)
summary(model_td_lis_prof_2)




#MODEL TD with mean proficiency
model_td_mean_prof2 <- glmer(accuracy ~ time + video_contrast + mean_proficiency +
    time*video_contrast
                        + (1|participant) + (1|correct_word), family =
    binomial(link="logit"), data = perc_quest_split_td)
summary(model_td_mean_prof2)


#MODEL TD with listening_hpw
model_td_lis_hpw2 <- glmer(accuracy ~ time + video_contrast + listening_hpw +
    time*video_contrast
                        + (1|participant) + (1|correct_word), family =
    binomial(link="logit"), data = perc_quest_split_td)
summary(model_td_lis_hpw2)




#the huge model TD with all proficiency measures
huge_model_td <- glmer(accuracy ~ time + video_contrast + listening_proficiency +
    mean_proficiency + listening_hpw + time*video_contrast
                      + (1|participant) + (1|correct_word), family =
    binomial(link="logit"), data = perc_quest_split_td)
summary(huge_model_td)




#Compare three models with each other


#compare model_td_lis_prof_2 / model_td_mean_prof2 / model_td_lis_hpw2


anova(model_td_lis_prof2, model_td_mean_prof2)


anova(model_td_mean_prof2, model_td_lis_hpw2)


anova(model_td_lis_prof2, model_td_lis_hpw2)
```

3. Analysis Section 3.2. Duration cue

```
rm(list = ls())      #empty environment


#------------install/load packages-----------#


library(lme4)
library(effects)
library(dplyr)
setwd(".../Data_analyses_thesis")   #set local directory       #set local directory


# DATA SPLIT UP GROUPS


perception_data_split <- read.csv("perception_data_splitupgroups.csv", sep=",")
    #load perc data
perception_data_split$time <- factor(perception_data_split$time, levels = c("pre",
    "post"))     # order time pre->post


perc_data_summary <- perception_data_split %>%
  group_by(video_contrast, video_cue, time) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


# make AE data frame (split up data)


perception_data_ae <- subset(perception_data_split, contrast_type == "ae")
    #make subset ae contrast


data_summary_ae <- perception_data_ae %>%
    #calculate means and sds for plot
  group_by(video_contrast, time) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


# make TD data frame (slit up data)
perception_data_td <- subset(perception_data_split, contrast_type == "td")


data_summary_td <- perception_data_td %>%
    #calculate means and sds for plot
  group_by(video_contrast, time) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))




# DATA INTACT GROUPS
perception_data_intact <- read.csv("perception_data_intactgroups.csv", sep=",")
    #load perc data
perception_data_intact$time <- factor(perception_data_intact$time, levels =
    c("pre", "post"))     # order time pre->post
```

```
#3.2.1

# within the videos that are about ae, which video (dur/nocue) led to better
    results?
# AE summary of perception data in boxplots
#Data: intact groups

perc_ae <- subset(perception_data_intact, contrast_type == "ae")        #make
    subset ae items

data_summary_ae_total <- perc_ae %>%                                    #calculate
    means and sds for plot
  group_by(group, time) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))

perc_ae_aedur <- subset(perc_ae, group == "ae-dur")        #split up data on ae
    items by group
perc_ae_tddur <- subset(perc_ae, group == "td-dur")
perc_ae_aenocue <- subset(perc_ae, group == "ae-nocue")
perc_ae_tdnocue <- subset(perc_ae, group == "td-nocue")

# make summaries for every group, calculating means and sd's
data_summary_perc_ae_aedur <- perc_ae_aedur %>%
  group_by(participant, time) %>%
  summarise(ae_aedur_score_mean = mean(accuracy), ae_aedur_score_sd = sd(accuracy))

data_summary_perc_ae_aenocue <- perc_ae_aenocue %>%
  group_by(participant, time) %>%
  summarise(ae_aenocue_score_mean = mean(accuracy), ae_aenocue_score_sd =
    sd(accuracy))

data_summary_perc_ae_tddur <- perc_ae_tddur %>%
  group_by(participant, time) %>%
  summarise(ae_tddur_score_mean = mean(accuracy), ae_tddur_score_sd = sd(accuracy))

data_summary_perc_ae_tdnocue <- perc_ae_tdnocue %>%
  group_by(participant, time) %>%
  summarise(ae_tdnocue_score_mean = mean(accuracy), ae_tdnocue_score_sd =
    sd(accuracy))



# BOXPLOT PERC mean accuracy scores per group: AE items
par(mfrow = c(1,4))

plot(data_summary_perc_ae_aedur$time,
    data_summary_perc_ae_aedur$ae_aedur_score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Perc scores ae-dur video,
    ae items", ylim = c(0.4,1.0))

plot(data_summary_perc_ae_aenocue$time,
    data_summary_perc_ae_aenocue$ae_aenocue_score_mean,
```

```
     xlab = "time", ylab = "mean accuracy score", main = "Perc scores ae-nocue
    video, ae items", ylim = c(0.4,1.0))


plot(data_summary_perc_ae_tddur$time,
    data_summary_perc_ae_tddur$ae_tddur_score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Perc scores td-dur video,
    ae items", ylim = c(0.4,1.0))


plot(data_summary_perc_ae_tdnocue$time,
    data_summary_perc_ae_tdnocue$ae_tdnocue_score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Perc scores td-nocue
    video, ae items", ylim = c(0.4,1.0))




# BOXPLOT AE dur vs nocue
par(mfrow = c(1,2))


plot(data_summary_perc_ae_aedur$time,
    data_summary_perc_ae_aedur$ae_aedur_score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Identification task: ae
    items, ae-dur video", ylim = c(0.4,1.0))


plot(data_summary_perc_ae_aenocue$time,
    data_summary_perc_ae_aenocue$ae_aenocue_score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Identification task: ae
    items, ae-nocue video", ylim = c(0.4,1.0))




# MODEL AE dur vs nocue

# Data: Split up groups


perception_data_ae_aevideos <- subset(perception_data_ae, video_contrast == "ae")
    #make subset ae videos


data_summary_aevideos <- perception_data_ae_aevideos %>%
    #calculate means and sds for plot
  group_by(video_cue, time) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


model_ae_aevideos <- glmer(accuracy ~ time + video_cue + time*video_cue +
    (1|participant) + (1|correct_word),
                           family = binomial(link="logit"), data =
    perception_data_ae_aevideos)
summary(model_ae_aevideos)


plot(predictorEffects(model_ae_aevideos, ~ time), main = "The effect of dur versus
    nocue video on ae items", lines=list(multiline=TRUE),
     axes=list(grid=TRUE),
     confint=list(style="bars"))
```

```
#3.2.2

# within the videos that are about td, which video (dur/nocue) led to better
    results?
# Data: intact groups

# TD summary of perception data in boxplots
perc_td <- subset(perception_data_intact, contrast_type == "td")        #make
    subset td contrast

data_summary_td_total <- perc_td %>%
    #calculate means and sds for plot
  group_by(group, time) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))

perc_td_aedur <- subset(perc_td, group == "ae-dur")        #split up perc td data
    by group
perc_td_tddur <- subset(perc_td, group == "td-dur")
perc_td_aenocue <- subset(perc_td, group == "ae-nocue")
perc_td_tdnocue <- subset(perc_td, group == "td-nocue")

# make summaries for every group, calculating means and sd's
data_summary_perc_td_aedur <- perc_td_aedur %>%
  group_by(participant, time) %>%
  summarise(td_aedur_score_mean = mean(accuracy), td_aedur_score_sd = sd(accuracy))

data_summary_perc_td_aenocue <- perc_td_aenocue %>%
  group_by(participant, time) %>%
  summarise(td_aenocue_score_mean = mean(accuracy), td_aenocue_score_sd =
    sd(accuracy))

data_summary_perc_td_tddur <- perc_td_tddur %>%
  group_by(participant, time) %>%
  summarise(td_tddur_score_mean = mean(accuracy), td_tddur_score_sd = sd(accuracy))

data_summary_perc_td_tdnocue <- perc_td_tdnocue %>%
  group_by(participant, time) %>%
  summarise(td_tdnocue_score_mean = mean(accuracy), td_tdnocue_score_sd =
    sd(accuracy))



# BOXPLOT PERC mean accuracy scores per group: TD items
par(mfrow = c(1,4))

plot(data_summary_perc_td_tddur$time,
    data_summary_perc_td_tddur$td_tddur_score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Perc td scores td-dur
    video", ylim = c(0.4,1.0))
```

```
plot(data_summary_perc_td_tdnocue$time,
    data_summary_perc_td_tdnocue$td_tdnocue_score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Perc td scores td-nocue
    video", ylim = c(0.4,1.0))


plot(data_summary_perc_td_aedur$time,
    data_summary_perc_td_aedur$td_aedur_score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Perc td scores ae-dur
    video", ylim = c(0.4,1.0))


plot(data_summary_perc_td_aenocue$time,
    data_summary_perc_td_aenocue$td_aenocue_score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Perc td scores ae-nocue
    video", ylim = c(0.4,1.0))




# BOXPLOT TD dur vs nocue
par(mfrow = c(1,2))
plot(data_summary_perc_td_tddur$time,
    data_summary_perc_td_tddur$td_tddur_score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Identification task: td
    items, td-dur video", ylim = c(0.4,1.0))


plot(data_summary_perc_td_tdnocue$time,
    data_summary_perc_td_tdnocue$td_tdnocue_score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Identification task: td
    items, td-nocue video", ylim = c(0.4,1.0))




# MODEL TD dur vs nocue


# Data: Split up groups
perception_data_td_tdvideos <- subset(perception_data_td, video_contrast == "td")
    #make subset ae videos


data_summary_tdvideos <- perception_data_td_tdvideos %>%
    #calculate means and sds for plot
  group_by(video_cue, time) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


model_td_tdvideos <- glmer(accuracy ~ time + video_cue + time*video_cue +
    (1|participant) + (1|correct_word),
                        family = binomial(link="logit"), data =
    perception_data_td_tdvideos)
summary(model_td_tdvideos)


plot(predictorEffects(model_td_tdvideos, ~ time), main = "The effect of dur versus
    nocue video on td items", lines=list(multiline=TRUE),
     axes=list(grid=TRUE),
     confint=list(style="bars")) #try to change the label of the lines.
```

4. Analysis Section 3.3. Assymetry

```
rm(list = ls())     #empty environment


#------------install/load packages-----------#
library(lme4)
library(effects)
library(dplyr)
setwd(".../Data_analyses_thesis")   #set local directory


#INTACT DATA
perception_data_intact <- read.csv("perception_data_intactgroups.csv", sep=",")
    #load perc data
perception_data_intact$time <- factor(perception_data_intact$time, levels =
    c("pre", "post"))     # order time pre->post


#SPLIT DATA
perception_data_split <- read.csv("perception_data_splitupgroups.csv", sep=",")
    #load perc data split up groups
perception_data_split$time <- factor(perception_data_split$time, levels = c("pre",
    "post"))        # order time pre->post


perception_data_ae <- subset(perception_data_split, contrast_type == "ae")
    #make subset ae contrast
perception_data_td <- subset(perception_data_split, contrast_type == "td")
    #make subset td contrast


# Assymmetric difficulty RIGHT ANSWER TYPE analysis
# Data summaries subsets, split up data


#general
data_summary_right_answer_type <- perception_data_split %>%
  group_by(participant, video_contrast, video_cue, time, contrast_type,
    right_answer_type) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


#only ae items
data_summary_right_answer_ae <- perception_data_ae %>%
  group_by(participant, video_contrast, video_cue, time, contrast_type,
    right_answer_type) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


#only td items
data_summary_right_answer_td <- perception_data_td %>%
  group_by(participant, video_contrast, video_cue, time, contrast_type,
    right_answer_type) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))



#3.3.1


# BOXPLOT OF A AND E SEPERATELY
```

```
perception_data_e <- subset(perception_data_ae, right_answer_type == "e")
    #make subset right answer type is e
perception_data_a <- subset(perception_data_ae, right_answer_type == "a")
    #make subset right answer type is a


data_summary_right_answer_e_small <- perception_data_e %>%                   #data
    summary for this subset: means & sd's
  group_by(participant, time, contrast_type, right_answer_type) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


data_summary_right_answer_a_small <- perception_data_a %>%                   #data
    summary for this subset: means & sd's
  group_by(participant, time, contrast_type, right_answer_type) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))
par(mfrow = c(1,2))


plot(data_summary_right_answer_a_small$time,
    data_summary_right_answer_a_small$score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Scores for a", ylim =
    c(0.0,1.0))


plot(data_summary_right_answer_e_small$time,
    data_summary_right_answer_e_small$score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Scores for e", ylim =
    c(0.0,1.0))
par(mfrow = c(1,1))


# MODEL A vs E item general difficulty (over all groups): right answer type


#only right answer type
mod_right_answer_ae <- glmer(accuracy ~ right_answer_type + (1|participant) +
    (1|correct_word),
                          family = binomial(link="logit"), data =
    perception_data_ae)
summary(mod_right_answer_ae)
plot(allEffects(mod_right_answer_ae))




#time, video contrast & right answer type & three-way interaction
mod_right_answer_ae2 <- glmer(accuracy ~ time + right_answer_type + video_contrast
    + time*right_answer_type*video_contrast + (1|participant) + (1|correct_word),
                           family = binomial(link="logit"), data =
    perception_data_ae)
summary(mod_right_answer_ae2)
plot(allEffects(mod_right_answer_ae2))




#time and right answer type only, two-way interaction
mod_right_answer_ae3 <- glmer(accuracy ~ time + right_answer_type +
    time*right_answer_type + (1|participant) + (1|correct_word),
                           family = binomial(link="logit"), data =
    perception_data_ae)
```

```
summary(mod_right_answer_ae3)
plot(allEffects(mod_right_answer_ae3))



#3.3.2


# MODEL T vs D item general difficulty (over all groups): right answer type


# Bboxplots for t and d separately


perception_data_t <- subset(perception_data_td, right_answer_type == "t")
    #make subset right answer type is e
perception_data_d <- subset(perception_data_td, right_answer_type == "d")
    #make subset right answer type is a


data_summary_right_answer_t_small <- perception_data_t %>%                    #data
    summary for this subset: means & sd's
  group_by(participant, time, contrast_type, right_answer_type) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


data_summary_right_answer_d_small <- perception_data_d %>%                    #data
    summary for this subset: means & sd's
  group_by(participant, time, contrast_type, right_answer_type) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))
par(mfrow = c(1,2))


plot(data_summary_right_answer_t_small$time,
    data_summary_right_answer_t_small$score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Scores for t", ylim =
    c(0.0,1.0))


plot(data_summary_right_answer_d_small$time,
    data_summary_right_answer_d_small$score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Scores for d", ylim =
    c(0.0,1.0))
par(mfrow = c(1,1))




# MODEL T vs D item general difficulty (over all groups): right answer type
#only right answer type
mod_right_answer_td <- glmer(accuracy ~ right_answer_type + (1|participant) +
    (1|correct_word),
                             family = binomial(link="logit"), data =
    perception_data_td)
summary(mod_right_answer_td)
plot(allEffects(mod_right_answer_td))


#right answer type, time, video contrast
mod_right_answer_td2 <- glmer(accuracy ~ time + right_answer_type + video_contrast
    + time*right_answer_type*video_contrast + (1|participant) + (1|correct_word),
                              family = binomial(link="logit"), data =
    perception_data_td)
```

```
summary(mod_right_answer_td2)


plot(allEffects(mod_right_answer_td2))


#time and right answer type only, two-way interaction
mod_right_answer_td3 <- glmer(accuracy ~ time + right_answer_type +
    time*right_answer_type + (1|participant) + (1|correct_word),
                              family = binomial(link="logit"), data =
    perception_data_td)
summary(mod_right_answer_td3)


plot(allEffects(mod_right_answer_td3))
```

5. Analysis Section 3.4. Metalinguistic awareness

```
rm(list = ls())     #empty environment


#------------install/load packages-----------#
library(lme4)
library(effects)
library(dplyr)
setwd(".../Data_analyses_thesis")   #set local directory


# SK data: intact groups
sound_knowledge_data <- read.csv("sound_knowledge_data_intactgroups.csv", sep=",")
    #load sk data
sound_knowledge_data$time <- factor(sound_knowledge_data$time, levels = c("pre",
    "post"))     #order pre -> post


## SK: effects of time
sk_ae <- subset(sound_knowledge_data, contrast_type == "ae")   #subset sk data: ae
    pairs


data_summary_sk_ae <- sk_ae %>%                               #calculate means and
    sd's per group & time
  group_by(participant, group, time) %>%
  summarise(ae_score_mean = mean(accuracy), ae_score_sd = sd(accuracy))


sk_td <- subset(sound_knowledge_data, contrast_type == "td")    #subset sk data: td
    pairs


data_summary_sk_td <- sk_td %>%                               #calculate means and
    sd's per group & time
  group_by(participant, group, time) %>%
  summarise(td_score_mean = mean(accuracy), td_score_sd = sd(accuracy))


# BOXPLOT sk mean accuracy scores for AE pairs and TD pairs


par(mfrow = c(1,2))
```

```
plot(data_summary_sk_ae$time, data_summary_sk_ae$ae_score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Metalinguistic awareness:
    ae word pairs")


plot(data_summary_sk_td$time, data_summary_sk_td$td_score_mean,
     xlab = "time", ylab = "mean accuracy score", main = "Metalinguistic awareness:
    td word pairs")


# MODEL effect of time


#effect of time on ae pairs
mod_sk_time_ae <- glmer(accuracy ~ time + (1|participant) + (1|trial_id),
                   family = binomial(link="logit"), data = sk_ae)
summary(mod_sk_time_ae)


#effect of time on td pairs
mod_sk_time_td <- glmer(accuracy ~ time + (1|participant) + (1|trial_id),
                        family = binomial(link="logit"), data = sk_td)
summary(mod_sk_time_td)


#subset sk data: pretest
sk_pre <- subset(sound_knowledge_data, time == "pre")


# Homophone fillers
sk_pre_filler_homophones <- subset(sk_pre, pair_type == "homophone" & contrast_type
    == "filler")


# overview per participant: range 30 - 100
data_summary_fillerhomophones <- sk_pre_filler_homophones %>%
  group_by(participant, group) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


# mean and sd score overall
data_summary_fillerhomophones2 <- sk_pre_filler_homophones %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


# Minimal pair fillers
sk_pre_filler_minpairs <- subset(sk_pre, pair_type == "min_pair" & contrast_type ==
    "filler")


# overview per participant, range 60 - 100 %
data_summary_fillerminpairs <- sk_pre_filler_minpairs %>%
  group_by(participant, group) %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


# mean and sd score overall
data_summary_fillerminpairs2 <- sk_pre_filler_minpairs %>%
  summarise(score_mean = mean(accuracy), score_sd = sd(accuracy))


# MERGE sk pretest data summaries with perc pretest data
```

```
sk_ae_pre <- subset(sk_pre, contrast_type == "ae")              #subset sk pretest
    data: ae pairs
sk_td_pre <- subset(sk_pre, contrast_type == "td")              #subset sk pretest
    data: td pairs


data_summary_sk_ae_pre <- sk_ae_pre %>%                         #data summary ae
    pairs: calculate means and sd's
  group_by(participant) %>%
  summarise(sk_ae_pre_score_mean = mean(accuracy), sk_ae_pre_score_sd =
    sd(accuracy))


data_summary_sk_td_pre <- sk_td_pre %>%                         #data summary td
    pairs: calculate means and sd's
  group_by(participant) %>%
  summarise(sk_td_pre_score_mean = mean(accuracy), sk_td_pre_score_sd =
    sd(accuracy))


# Perception data intact groups
perception_data_intact <- read.csv("perception_data_intactgroups.csv", sep=",")
    #load perc data
perception_data_intact$time <- factor(perception_data_intact$time, levels =
    c("pre", "post"))     # order time pre->post


#take perception subset only pretest data
perception_data_pre <- subset(perception_data_intact, time == "pre")


# add means and sd scores of ae items as columns/vectors to the perception data
    pretest
merged_perc_sk_ONLYae <- merge(perception_data_pre, data_summary_sk_ae_pre)


# add meand and sd scores of td items as columns/vectors to this dataframe
merged_perc_sk_complete <- merge(merged_perc_sk_ONLYae, data_summary_sk_td_pre)


#--->>> merged_perc_sk_complete now contains means and sd's vectors of sk pretest
    data <<<---#


# making different data frames for ae or dt perc contrast type respectively


merged_ae_perc_data_withsk <- subset(merged_perc_sk_complete, contrast_type ==
    "ae")  #subset perc with sk data: contrast type = ae
merged_td_perc_data_withsk <- subset(merged_perc_sk_complete, contrast_type ==
    "td")  #subset perc with sk data: contrast type = td




#3.4.1


# CORRELATION ANALYSIS: sk pre on AE pairs versus perc pre on AE contrast type
# PLOT ae perc data against ae sk scores to see correlation
```

```
par(mfrow = c(1,1)) #plot one graph at a time

# Plotting raw accuracy scores does not work. So calculate means for perc scores.
# Calculate means and sd's of perc AE contrast per participant

data_summary_perc_pre_ae <- merged_ae_perc_data_withsk %>%
  group_by(participant) %>%
  summarise(perc_ae_pre_score_mean = mean(accuracy), perc_ae_pre_score_sd =
    sd(accuracy))

# Merge those means and sds with the AE file
merged_ae_withpercmeans <- merge(merged_ae_perc_data_withsk,
    data_summary_perc_pre_ae)
#PLOT the correlation for ae sk versus perc

plot(merged_ae_withpercmeans$perc_ae_pre_score_mean,
    merged_ae_withpercmeans$sk_ae_pre_score_mean,
     xlab = "perception pretest accuracy of ae items", ylab = "sk pretest mean
    accurary for ae word pairs",
     xlim = c(0.4,1.0), main = "The correlation of sk and perc mean scores for ae
    contrast" )

# TESTS FOR CORRELATION AE

# raw accuracy (perc) versus the means (sk) added in every row
cor.test(merged_ae_perc_data_withsk$accuracy,
    merged_ae_perc_data_withsk$sk_ae_pre_score_mean, method= "pearson") #or kendall
    or spearman

# but raw data versus means is probably not good in correlation, so:
# correlation between perc mean scores and sk mean scores

cor.test(merged_ae_withpercmeans$perc_ae_pre_score_mean,
    merged_ae_withpercmeans$sk_ae_pre_score_mean, method= "pearson") #or kendall or
    spearman

# but still many rows with the same number so:
# dataframe with only one row per participant: means of perc and sk tasks

merged_means_perc_sk_ae <- merge(data_summary_perc_pre_ae, data_summary_sk_ae_pre)
cor.test(merged_means_perc_sk_ae$perc_ae_pre_score_mean,
    merged_means_perc_sk_ae$sk_ae_pre_score_mean, method= "pearson") #or kendall or
    spearman

# carry out a linear regression between means to add the regression line to the
    plot
# using the dataframe with only one row per participant perc means and sk means
fit1 <- lm (merged_means_perc_sk_ae$sk_ae_pre_score_mean ~
    merged_means_perc_sk_ae$perc_ae_pre_score_mean)
summary(fit1)
```

```
plot(merged_means_perc_sk_ae$perc_ae_pre_score_mean,
    merged_means_perc_sk_ae$sk_ae_pre_score_mean,
     xlab = "Identification task: pretest mean accuracy on ae items", ylab =
    "Awareness: pretest mean accurary on ae word pairs",
     xlim = c(0.4,1.0), main = "The correlation of awareness and identification
    accuracy for the ae contrast", abline(fit1, lty = "dashed"))



#Change plot so that scales are similar
plot(merged_means_perc_sk_ae$perc_ae_pre_score_mean,
    merged_means_perc_sk_ae$sk_ae_pre_score_mean,
     xlab = "Identification task: pretest mean accuracy on ae items", ylab =
    "Awareness: pretest mean accurary on ae word pairs",
     xlim = c(0.0,1.0), main = "The correlation of awareness and identification
    accuracy for the ae contrast", abline(fit1, lty = "dashed"))



#3.4.2.


# PLOT td perc data against td sk scores to see correlation


# plotting raw accuracy scores does not work. So calculate means for perc scores:
# calculate means and sd's for perception data td contrast
data_summary_perc_pre_td <- merged_td_perc_data_withsk %>%
  group_by(participant) %>%
  summarise(perc_td_pre_score_mean = mean(accuracy), perc_td_pre_score_sd =
    sd(accuracy))


#merge it with the existing td perc&sk data frame
merged_td_withpercmeans <- merge(merged_td_perc_data_withsk,
    data_summary_perc_pre_td)


# PLOT the correlation for td sk versus perc


plot(merged_td_withpercmeans$perc_td_pre_score_mean,
    merged_td_withpercmeans$sk_td_pre_score_mean,
     xlab = "perception pretest accuracy of td items", ylab = "sk pretest mean
    accurary for td word pairs",
     xlim = c(0.4,1.0), main = "The correlation of sk and perc mean scores for td
    contrast")


# TESTS FOR CORRELATION


#raw accuracy versus sk means
cor.test(merged_td_perc_data_withsk$accuracy,
    merged_td_perc_data_withsk$sk_td_pre_score_mean, method= "pearson") #or kendall
    or spearman


#mean perc vs mean sk but many observations per participant
cor.test(merged_td_withpercmeans$perc_td_pre_score_mean,
    merged_td_withpercmeans$sk_td_pre_score_mean, method= "pearson") #or kendall or
    spearman
```

```
#making new dataframe for only the means of perc and the means of sk
merged_means_perc_sk_td <- merge(data_summary_perc_pre_td, data_summary_sk_td_pre)


#correlation with this simple df
cor.test(merged_means_perc_sk_td$perc_td_pre_score_mean,
    merged_means_perc_sk_td$sk_td_pre_score_mean, method= "pearson") #or kendall or
    spearman


# carry out a linear regression between means to add the regression line to the
    plot
# using the dataframe with only one row per participant perc means and sk means


fit2 <- lm (merged_means_perc_sk_td$sk_td_pre_score_mean ~
    merged_means_perc_sk_td$perc_td_pre_score_mean)
summary(fit2)


plot(merged_means_perc_sk_td$perc_td_pre_score_mean,
    merged_means_perc_sk_td$sk_td_pre_score_mean,
     xlab = "Identification task: pretest mean accuracy on td items", ylab =
    "Awareness: pretest mean accurary on td word pairs",
     xlim = c(0.4,1.0), main = "The correlation of awareness and identification
    accuracy for the td contrast", abline(fit1, lty = "dashed"))


#change plot so that scales are similar
plot(merged_means_perc_sk_td$perc_td_pre_score_mean,
    merged_means_perc_sk_td$sk_td_pre_score_mean,
     xlab = "Identification task: pretest mean accuracy on td items", ylab =
    "Awareness: pretest mean accurary on td word pairs",
     xlim = c(0.0,1.0), main = "The correlation of awareness and identification
    accuracy for the td contrast", abline(fit1, lty = "dashed"))
```