And & En, do proficient L2 speakers adapt the vowel to the target language?

Slump, Debora

And & *En*, do proficient L2 speakers adapt the vowel to the target language?

D. Slump

Master thesis

Linguistics

Universiteit Leiden

Thesis supervisor: Dr. W.F.L. Heeren

Second reader: Dr. N.H. de Jong

15-07-2022

Abstract

In forensic speaker comparison, there may be cases in which the speech samples to be compared are not in the same language. Research in forensic phonetics has however almost exclusively been done in monolingual contexts. This study aims to give more clarity about what features can be used in cross-linguistic speaker comparison. This is done by investigating whether proficient first language (L1) Dutch speakers of second language (L2) English adapt the vowel in *and* to the target language. L2 sound learning models predict that Dutch learners will have difficulty learning the difference between /ɛ/ and /æ/ and will therefore only be able to approximate nativelike pronunciation. The vowels in *en* and *and* of 35 proficient, female Dutch speakers of English were analyzed, both as a discourse marker and a coordinating conjunction. Mixed-effect models showed that for both word classes, the F1 and F2 were language-dependent. This means that the use of /æ/ as a characteristic in cross-linguistic speaker comparison is not useful.

Introduction

Forensic Speech Science is one of the most fast-growing fields in applied linguistics (French & Stevens, 2013). The most common practice in Forensic Speech Science is speaker comparison. Forensic speaker comparison consists of comparing a voice in a speech sample, for example a criminal recording, with the voice of a suspect (French & Stevens, 2013). This is done to help the court determine whether the suspect is the person talking in the speech sample. Speaker comparison is no easy task. Unlike a fingerprint or DNA, someone's voice does not have one or more biologically fixed features that are always the same. One person's voice may differ because of a variety of factors. Although some features are partly determined by biology, such as the fundamental frequency (F0), many features are influenced by the speaker's social and regional background, as well as the situational context (Foulkes & French, 2012). There are also no features of speech that can be found in every single utterance. Forensic speaker comparison on its own cannot determine the identity of a speaker, but it can provide crucial evidence that can be used alongside other information. For these speaker comparisons, it is therefore important to know what features of speech are speaker-specific, meaning they remain largely consistent in one's speech but have large variation between different speakers.

A lot of the research done in forensic phonetics focusses on finding features of speech and voices that have small within-speaker variation, but large between-speaker variation. The research has almost exclusively been conducted in a monolingual context. The Code of Practice of the International Association for Forensic Phonetics & Acoustics states "[m]embers should exercise particular caution with cross-language comparisons" (IAFPA, 2020). In forensic speaker comparison, there are however situations in which the speech materials to be compared are not in the same language (van der Vloed & Bouten, 2014). It is therefore important to know what features of speech are largely language-independent within speakers and can therefore be used in cross-linguistics comparisons.

People who speak a foreign language often have a foreign accent. A foreign accent is commonly described as a difference in pronunciation in non-native speech that deviates from the norms of native speech (Gut, 2007). According to speech learning models, this is because speakers cannot distinguish sounds that are very similar but not quite identical, in their first language (L1) and their second language (L2). Speakers may categorize the two sounds into the same phonetic category, and pronounce the L2 sound in an L1 way. One sound where this may occur is the /æ/ vowel in English, for which Dutch listeners found English word pronounced with /ɛ/ easier to recognize than when pronounced with /æ/ (Broersma et al., 2010). This indicates that Dutch learners of English might not have different categories for the two sounds.

This study aims to give more clarity into which aspects of speech can be useful in cross-linguistics comparisons. If there is L1 transfer of the vowel in the /æ/, this segment could be included in cross-linguistics comparisons. The remainder of the introduction first gives an overview of several models of learning L2 sounds and the prediction they make for how Dutch learners of English produce the /æ/ vowel. Following this, different uses of *and* are explained and how these different uses might affect L2 pronunciation. The introduction ends with the research questions and hypotheses for the current study.

Learning L2 sounds

Late learners of a second language usually have a foreign accent when speaking their L2 (Tahta et al., 1981). Factors that can influence this accent are the age of acquisition of L2 (Moyer, 1999), the amount of time spent in a country where the L2 is spoken (Flege & Fletcher, 1992), and how often

the L2 is used (Flege et al., 1999; for an overview see Piske et al., 2001). This section first explains the link between production and perception. Following this, three models of learning L2 sounds are introduced, ending the section with their predictions of the realization of the /æ/ vowel by Dutch learners of English.

Production-perception link

Several models aim to explain how people learn the sounds of a new language. Most of these models are primarily concerned with perception. It is however also useful to know how speakers perceive L2 segments in order to know in what way people produce L2 sounds. This is because of the perception-production link. This link plays an important role in the development of L2 pronunciation and refers to the notion that a learner first has to be able to perceive a sound in order to be able to produce it (Isbell, 2016). Evidence for the perception-production link has been found for both the L1 as well as the L2 (for a review, see Isbell, 2016). Bradlow et al. (1997) investigated the effect of training in the perceptual identification of the English /r/-/l/ contrast on /r/-/l/ production by 11 L1 Japanese speakers. They gave the Japanese speakers intensive perception training on the /r/-/l/ contrast, consisting of an extensive number of /r/-/l/ minimal word pairs. They then compared results from a pretest with the results of a post-test in both of which they gathered perception and production data of the /r/-/l/ contrast. A control group performed the same pretest and post-test but did not receive the training. The accuracy of /r/-/l/ identification went up from 65% to 81% before and after the training for the trained group. The control group showed no improvement between the pretest and the post-test. For the production of the /r/-/l/ contrast, the trained group also showed a significant improvement from the pretest to the post-test, whereas the control group showed no difference in production accuracy. Because the subjects had not received any production training, the authors conclude that any improvement in production is the result of transfer of gained knowledge in the perception domain to the production domain.

Similar results have been found by Motohashi-Saigo and Hardison (2009). They trained 30 beginning L1 English learners of L2 Japanese in Japanese geminate perception. Geminates are consonants that have a larger duration than singleton consonants (e.g. *kite* 'coming' versus *kitte* 'postage stamp'). In Japanese, duration of consonants (and of vowels) is a contrastive feature, as opposed to in English. Participants received either auditory-only training or auditory-visual training. The auditory-only group received information about geminates and was presented a stimulus and had to indicate whether they heard a singleton, a geminate or a long vowel. After their response, they received feedback about their choice. The auditory-visual group received a demonstration of waveforms that showed the difference between singletons and geminates in addition to the auditory training. A control group received no training. All three groups were administered a pretest and a post-test to determine if perception and production of geminates had improved in the post-test compared to the pretest. Both groups that received training had improved in geminate identification and production. The control group showed no improvement. These findings too give support for the perception-production link, showing that in order to produce a sound well, a learner first needs to be able to perceive it well.

L2 sound learning models

Because of this perception-production link, it is useful to know how speakers perceive L2 segments in order to understand in what way people produce L2 segments. Several models aim to explain how people learn new L2 sounds. The most influential ones that postulate how the L1 influences L2 learning are the Perceptual Assimilation Model of L2 speech learning (PAM-L2) (Best & Tyler, 2007), the Second Language Linguistic Perception model (L2LP) (Escudero, 2005; Van Leussen & Escudero, 2015) and the Speech Learning Model (SLM) (Flege, 1995; Flege & Bohn, 2021). I will first briefly

explain the three models, and then the prediction they make for how L1 Dutch speakers of L2 English will produce the /æ/ vowel.

PAM-L2 & L2LP

Both PAM-L2 (Best & Tyler, 2007) and L2LP (Escudero, 2005) focus on sound perception and make predictions about the perceptual development of phonological contrasts. PAM-L2 is based on the Perceptual Assimilation Model (PAM) (Best, 1995). PAM and L2LP assume that there are no separate L1 and L2 perceptual systems. How well a listener will be able to discriminate L2 phonological contrasts therefore depends on how each of the L2 sounds are assimilated on the L1 phonological system (Tyler, 2019). The learner will first categorize the L2 sound according to the already existing L1 categories. Learning will only be necessary when a pair of L2 sounds do not match the L1 categories. In order for the learner to accurately learn the L2 categorization, they need to either create new L2 categories or adjust their existing categories. The models posit that there are several scenarios that can occur when an L2 learner encounters L2 sounds. Table 1 shows an overview of the different types of scenarios (in L2LP) or assimilations (in PAM) and how well the models predict the L2 sound contrast will be discriminated and figure 1 shows the three scenarios with examples.

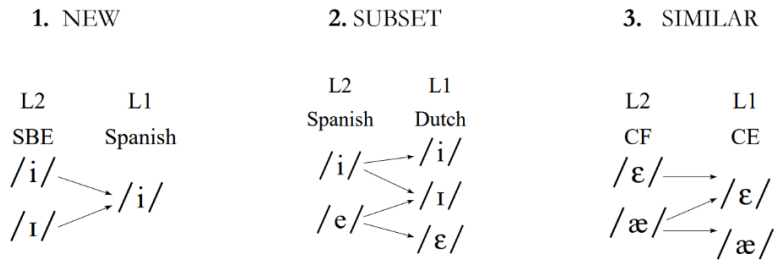
Table 1

Overview of the different assimilation types in L2LP and PAM

L2LP	PAM	Perception of two L2 sounds	Predicted discrimination
New scenario	Single-category assimilation	Versions of the same L1 category, both being an equally good or poor version. Fewer L1 categories than are needed for the L2.	Poor
Subset scenario	Uncategorized-categorized assimilation	One of the L2 sounds is not perceived as fitting in any L1 category. More L1 categories than are needed for the L2.	Very good
Similar scenario	Two-category assimilation	Good or acceptable versions of two different L1 categories. Equal amount of L1 categories as are needed for the L2.	Good

Figure 1

Three scenarios of sound perception in L2LP



Note. SBE = Southern British English, CF = Canadian French, CE = Canadian English. From "Linguistic Perception and Second Language Acquisition: Explaining the attainment of optimal phonological categorization" (p. 124) by P. Escudero, 2005, LOT Dissertation Series 113, Utrecht University.

The models state that it is more difficult for a learner to create a new category than it is to reduce the amount of categories or adjust the boundaries of the already existing categories. The models therefore predict that learning to accurately perceive the L2 sounds will be most difficult in the new scenario, because the learner needs to first perceive the difference between the two L2 sounds and then create a new category. In the subset scenario, the learner only has to reduce the number of categories and in the similar scenario the learner needs to adjust the boundaries of their already existing L1 categories, but will not need to create any new categories. The similar scenario is therefore predicted to be the least difficult for learners. PAM-L2 also uses the different assimilation types to predict how likely a learner is to form a new L2 category. The model predicts that when the type of assimilation changes because of the learner acquiring a new category, discriminating the two L2 sounds should be easier. Best and Tyler (2007) explain that a learner is not very likely to acquire a new L2 category if two L2 sounds are mapped on two different L1 categories (two-category assimilation). Because the learner perceives the L2 sounds as a version of an L1 category, the learner can easily discriminate the L2 sounds and perceptual learning is not likely to occur. For single-category assimilation, the learner will initially have difficulty discriminating the two sounds. Best and Tyler (2007) state that the likeliness of a listener learning the difference between such two L2 sounds might depend on whether the sounds are perceived as good or as poorer versions of the L1 category. The prediction is however, that most learners will not be very successful in learning to perceive a difference between the two L2 sounds. Factors as high frequency or dense phonological neighborhoods may increase the chance that a learner will be able to learn the difference. Lastly, in the case of uncategorized-categorized assimilation, the model predicts that learners are likely to form a new L2 category for the uncategorized sound. The assimilation then turns into a two-category assimilation, and the learner will more easily discriminate the two L2 sounds.

SLM

PAM-L2 and L2LP focus only on perception, whereas the Speech Learning Model (SLM) (Flege, 1995) also concerns production of L2 speech. SLM focuses on how speakers learn L2 vowels and consonants. SLM is based on the idea that many errors in L2 production are based in perception. L2 learners need accurate perceptual targets in order to guide the learning of L2 sounds correctly (Flege, 1995). This means that in order to learn how to pronounce an L2 sound, learners need to hear the difference between the new L2 sound and the neighboring L1 sounds. In SLM, Flege (1995; 1997) classifies L2 sounds as either identical, new or similar. Identical sounds are L2 sounds that do not significantly differ from the corresponding L1 sound as pronounced by native speakers of the two languages when acoustically analyzed. Identical sounds are easily and authentically produced by L2 speakers, because the speakers does not need to learn this sound, they already know and use it in their L1. New sounds are sounds that are both acoustically and perceptually distinct from any sound in the L1. According to Bohn and Flege (1992) for vowels, an L2 vowel should be classified as new only if it occupies a part of the phonetic vowel space that is not occupied by any vowel in the L1. According to SLM, new sounds will ultimately be pronounced authentically because a new phonetic category will be created for this sound. And because a new category is added, there should be no negative consequences for the sounds and categories that already existed in the speaker's phonetic system. Similar sounds are sounds that may be represented by the same IPA symbol, but when measured acoustically, differ in the two languages. SLM predicts that similar sounds will be the most difficult to learn, because L2 learners will not be able to distinguish the differences between the L2 sound and the L1 sound and will map the L2 sound on the already existing L1 category.

The model postulates that a learner needs phonetic categories in order to authentically produce L2 sounds (Flege, 1997). SLM predicts that L2 learners will, with sufficient L2 input, notice the difference between an L2 new sound and the sounds they know from their L1. This will lead to the creation of a

new phonetic category for the new L2 sound. Once the new category is established, and the learner has developed a rule for the realization of this sound, the model predicts that the pronunciation of the L2 sound will be authentic. For similar sounds however, the model states that no new category will be created, and therefore the pronunciation will only approximate the native pronunciation.

In their revised Speech Learning Model (SLM-r) Flege and Bohn (2021) maintain the hypothesis that learners relate L2 sounds to their already existing L1 categories, and that it is more likely that a new phonetic category is formed for L2 sounds that are perceived as less similar to L1 sounds. The categorical distinction between new, similar and identical sounds is however dropped. The prediction of the revised model stays the same as that of SLM. If speakers do not perceive a difference between an L2 sound and the closest L1 sound, they will not form a new phonetic category for the L2 sound, and will only approximate the pronunciation of the L2 sound. The more dissimilar they perceive the L2 sound to the closest L1 sound, the more likely they are to form a new phonetic category for the L2 sound and pronounce it authentically.

Prediction of the models

For all three models, the prediction of how difficult and/or accurate the perception and thus production of an L2 sound is, is dependent on how the learner perceives and categorizes the L2 sound in relation to L1 sounds and their existing L1 categories. The prediction the models make for the production of the /æ/ vowel in Dutch learners of English will therefore depend on how these learners categorize this vowel.

One study that investigates L2 production and the perception of L2 pronunciation is that of Broersma et al. (2010). Broersma et al. (2010) studied cross-linguistic production and perception of Japanese and Dutch accented English. In their study, both Japanese and Dutch speakers listened to English words being pronounced with a Japanese or Dutch accent. They found that for words with the /æ/ vowel, Dutch listeners did not find it easier to recognize the word when pronounced with a more English vowel, in this case meaning with a vowel height more similar to the English /æ/. Instead, Dutch listeners recognized words faster when the vowel height was more similar to the Dutch /ɛ/ vowel. This was in contrast to other aspects of speech. For several other segments, they found that both Japanese and Dutch listeners found it easier to recognize words when the pronunciation was closer to native English pronunciation than to an accented pronunciation, even though listeners may be more accustomed to the accented pronunciation. For example, in Dutch, voiced obstruents like /b,d,v,z/ surface as voiceless at the end of a syllable, although the underlying phonological form maintains the voiced obstruent. This is known as final devoicing. And although Dutch listeners are capable of hearing and categorizing voicing contrasts in English in a final position (Broersma, 2005), they do not use this in word recognition (Broersma & Cutler, 2008). Dutch listeners hear the word ‘move’ when a native English speaker pronounced [mu:f]. Dutch speakers themselves also often realize voiced final obstruents as voiceless when they are speaking English (Nejjari et al., 2012). In spite of being familiar with accented English in which final obstruents are devoiced, Dutch listeners recognized words more easily if they were pronounced more English, in this case with a final obstruent that was more voiced. Thus, for most speech aspects, Dutch listeners recognized words quicker when the pronunciation was closer to native English pronunciation than to an accented pronunciation. Only for the /æ/ - /ɛ/ contrast, Dutch speakers more easily recognized a word when the vowel height was closer to the Dutch /ɛ/ rather than the English /æ/. This was not the case for the Japanese listeners. The authors suggest that this shows that Dutch listeners might not be aware that replacing the /æ/ with /ɛ/ in Dutch accented English deviates from the target pronunciation of English.

The study of Broersma et al. (2010) suggests that Dutch listeners do not perceive the English /æ/ vowel as a different category vowel than the /ɛ/ vowel. In PAM-L2 and L2LP, this puts the /æ/-/ɛ/ distinction for Dutch learners of English in the single-category assimilation, or the new scenario. The models therefore predict that Dutch listeners will have difficulty perceiving the difference between the two sounds because they will have to form a new category for the /æ/ sound, and that most learners will not be very successful in learning to perceive a difference between the two L2 sounds. SLM classifies /æ/ as a new L2 sound for Dutch speakers and therefore predicts that, with sufficient L2 input, Dutch speakers will be able to create a new phonetic category for the vowel and learn how to pronounce it authentically. In SLM-r however, the categorical distinction between new, similar and identical sounds is dropped and it posits that the greater the perceived difference between an L2 sound and an existing L1-cateogry, the more likely it is for the learner to form a new category for the L2 sound. Based on Broersma et al. (2010), Dutch listeners do not easily perceive the difference between /æ/ and /ɛ/. Therefore, the model predicts that it will be difficult for Dutch learners of English to form a new category for /æ/. In SLM, if Dutch speakers are not aware of the difference between /æ/ and /ɛ/, they will not be able to learn to pronounce the /æ/ vowel authentically, and pronunciation will only approximate native pronunciation.

In conclusion, all three models predict that it will be difficult for Dutch learners of English to learn to perceive the difference between /æ/ and /ɛ/. Both PAM-L2 and L2LP only make predictions about perception, but based on the before mentioned perception-production link (Isbell, 2016), we might extend the predicted difficulties to the production of /æ/. This is also in line with the prediction of SLM that Dutch learners of English will only be able to approximate native pronunciation of /æ/.

Uses of *and*

A very frequent word that lends itself for cross-linguistics comparison is *and,* and its Dutch equivalent *en*. High frequent words are useful for forensic analysis, because they are more likely to be present in the speech samples that are being analyzed. Nolan (1983) gives high frequency as one of the characteristics that an ideal speaker-specific token should have. *And* is the sixth most frequent word in the SUBTLEX-UK database (Van Heuven et al., 2014) and the tenth most frequent word in the SUBTLEX-US database (Brysbaert et al., 2012). *En* is the ninth most frequent word in the SUBTLEX-NL database (Keuleers & Brysbaert, 2010).

And is commonly classified as a coordinating conjunction (Gleitman, 1965). Coordinating conjunctions connect clauses or same-category words or phrases (Curzan & Adams, 2006). *And* is however also used as a discourse marker (Schiffrin, 1987). Discourse markers are words or short phrases that a speaker uses to indicate how their turn fits into the discourse and to indicate their attitude about the discourse, without adding new information. *And* is often used to signal that what follows *and* should be heard as separate but parallel to (a part of) the preceding discourse (Fraser, 1988). In (1) some examples of *and* used as a coordinating conjunction are shown and (2) shows some examples of *and* used as a discourse marker.

(1) a. Uh my hobbies are soccer **and** dancing.
b. The waves were nice **and** the instructors were nice.
c. The best thing of living with only girls is that it's really clean **and** neat in the unit.
d. So we went to all graveyards **and** all the churches.

(2) a. So I had uhm about three months holiday.
And uhm I wanted to earn some money.
b. But that went well.
And uhm I had to uh learn how to drive the car.

c. They were planning to go in a couple of weeks.
And uh so I had to find people somewhere else to travel with.

Schourup (1999) describes several characteristics of discourse markers. He names connectivity as their main characteristic. Discourse markers relate utterances to the discourse they are part of. Discourse markers are syntactically optional; removing the discourse marker in a sentence does not change the grammaticality of that sentence. Schourop (1999) furthermore states that discourse markers are also semantically optional in that when omitted, the relationship between the sentence and the discourse expressed by the discourse marker is still available, only not explicitly. This optionality does not mean that discourse markers are redundant or useless. Discourse markers help guide a listener to the interpretation intended by the speaker and they rule out interpretations that the speaker did not intend. The third main characteristic that Schourop (1999) describes is non-truth-conditionality. Discourse markers do not contribute anything to the truth-conditions of an utterance. These three characteristics, connectivity, optionality, and non-truth-conditionality, are thought to be necessary aspects of discourse markers. Discourse markers are also often thought to have a weak clause association, meaning that they are outside the syntactic structure or only loosely tied to it (Brinton, 1996). Discourse markers are mostly found in an utterance-initial position (Hansen, 1997; Schiffrin 1987; 2006). Although some discourse markers can also be found in non-initial positions (Schourop, 1999), this is not expected for the word *and*. Since *and* is a coordinating conjunction, it connects clauses, words or phrases. In that use, it can therefore not appear in an utterance-initial position, as it needs to be preceded by the first of the two items that it is connecting. Therefore, when *and* has a sentence-initial position, it is most likely being used as a discourse marker.

To my knowledge, there have not been any studies on the use of *and* as a discourse marker in L2, nor on the phonetic L2 realization of vowels in discourse markers. Both of these topics are useful to investigate in order to know if they are language-independent features of L2 speech that might be useful for cross-linguistic speech comparison. There have however been studies about the use of other discourse markers in the L2. The majority of these studies find that learners generally use discourse markers less frequently than native speakers do. This has been found for learners of several languages, including Chinese (Zhao, 2013), Spanish (Romero Trillo, 2002) and German (Müller, 2005) learners of English. Learners do not only use discourse markers less frequently than native speakers do, they also use a smaller variety of discourse markers than native speakers do. This sometimes leads to an overuse of certain discourse markers. This has for example been found for the use of *well* by German learners of English (Müller, 2005) and the use of *yes*, *so*, and *I think* for Japanese learners of English (Shimada, 2014). Gilquin (2016) investigated the use of certain discourse markers in L2 English, including *and so* and *and then*. Just as the before mentioned studies, she found that L2 speakers, with an exception for *well,* show an underuse of discourse markers compared to native speakers. L2 speakers used *well* more than native speakers did. She also found that learners who had spent more time in an English-speaking country used discourse markers more frequently than learners who had not spent any time in an English speaking country, and that the frequency of discourse marker use increased as time spent in an English-speaking country increased. This increase in frequency corresponds with a more native-like use of discourse markers. When looking at the individual discourse markers, only *well* showed a decrease of frequency with an increase of time spent in an English-speaking country. Because learners showed an overuse of *well* compared with native speakers, this decrease of usage is also closer to native-like use of the discourse marker.

De Marco (2016) investigated the use and acoustic characteristics of three discourse markers (*allora* 'then', *quindi* 'therefore', *però* 'but) in L2 Italian by learners of different fluency levels. She measured the duration of the discourse markers, the duration of the vowels in the discourse markers, and

speech rate. She found that learners used some discourse markers (e.g. *allora* 'then') to take time to construct their utterance, by lengthening the discourse marker. When used in this way, the discourse marker mainly had an initial position. Other discourse markers (e.g. *quindi* 'therefore') were used to take time in order to better articulate a long utterance. She also found that the measured acoustic values were only close to those of native speakers in very proficient learners. These functions of discourse markers are very similar to the use and function of filled pauses. Filled pauses are sounds of hesitation (like *uh* or *um*) that speakers use to signal uncertainty or to plan the following utterance (Clark & Fox Tree, 2002). Filled pauses are produced relatively unconsciously (Clark & Fox Tree, 2002; Hughes et al., 2016) and have therefore been theorized to be transferred from the L1 to the L2 (De Leeuw, 2007; Clark & Fox Tree, 2002). Since L2 discourse markers can be used in a similar way to filled pauses, and the word *and* is similar to its Dutch counterpart *en*, it is possible that, when used as a discourse marker, there is L1-transfer and the Dutch learners produce the vowel more like the Dutch /ɛ/ than like the English /æ/. The transfer is hypothesized to occur because the concerning words are uttered relatively unconsciously. It is therefore possible, following the L1-transfer hypothesis, that Dutch learners of English will pronounce *and* in an utterance-initial position less nativelike than *and* in a non-initial position.

To investigate the L1-transfer hypothesis, De Boer and Heeren (2020) compared the production of filled pauses by proficient L1 Dutch learners of L2 English in their L2 with their L1 realizations. They found a cross-linguistic difference in L1 and L2 in the F1 and F2 of the vowels and in the *um:uh* ratio. Other characteristics, like duration, F0, and F3 stayed consistent across the two languages. This showed that proficient L1 Dutch learners of English adapted the vowel in the filled pauses towards the target language. Similar results were also found for L1 Afrikaans learners of L2 Spanish (García-Amaya & Lang, 2020) and proficient Japanese learners of English (Rose, 2017). These results indicate that filled pauses, and specifically the vowels, are language specific, and are not fully transferred from the L1 to the L2. Based on the before-mentioned similarities in use and function of filled pauses and L2 discourse markers, it does not seem likely that there will be more L1-transfer in the vowel in *and* when used as a discourse marker compared to when used as a coordinating conjunction for proficient Dutch learners of English.

In summary, *and* can be used as a coordinating conjunction, but also as a discourse marker, in which case it will usually be in an utterance-initial position. Although there has not been much research on the realization of discourse markers in speakers' L2, it has been found that learners generally use less and a smaller variety of discourse markers. The more proficient speakers are and the more time they spend in a country where the target language is spoken, the more native-like their use of discourse markers becomes. De Marco's (2016) study found that acoustic values (duration, vowel duration and speech rate) were only close to those of native speakers in very proficient speakers. There have not been any studies on the vowel formants in L2 discourse marker, nor specifically for *and*. However, De Marco (2016) found that learners use L2 discourse markers to take time to construct and plan their following utterance. This is very similar to a use of filled pauses. It has been theorized that because they are produced relatively unconscious, L1-transfer occurs in filled pauses. Following this theory, there might be more L1-transfer in *and* used as a discourse marker than in *and* used as a coordinating conjunction for Dutch learners of English. De Boer and Heeren (2020) however found that proficient L1 Dutch speakers of English adapt the vowel of filled pauses, specifically the F1 and F2 to the target language, which contradicts the L1-transfer hypothesis. Based on these results, it does not seem likely for there to be more L1-transfer in *and* as a discourse marker compared to a coordinating conjunction for Dutch learners of English.

Research question and hypotheses

The current study investigates whether Dutch speakers of English adapt the vowel in *and* to the target language when speaking their L2 English. It also investigates whether there is a difference in realization of the vowel in *and* when used as a discourse marker or as a coordinating conjunction. For the first question, the expectation is that it will be difficult for Dutch learners of English to learn to perceive the difference between /æ/ and /ɛ/ and that they therefore will only be able to approximate native pronunciation. Because the speakers used in this study are all proficient speakers of English, it is also not expected that they use the Dutch /ɛ/ vowel. The hypothesis is therefore that the realization of the /æ/ is in between the realization of the Dutch /ɛ/ and the English /æ/.

Dutch children often start learning English in primary school and are obligated to take English courses up to the final exam of secondary school. In education, as well as in the government communication, Standard British English is used (Nejjari et al., 2012). At the same time, American English is widely present on TV, in movies and in music. Dutch people therefore often speak English with a combination of British and American characteristics (Nejjari et al., 2012). In both British English and American English, the F1 of /æ/ is higher than of /ɛ/, and the F2 of /æ/ is lower than of /ɛ/ (Ghorshi et al., 2008). The expectation therefore is that the participants produce the English /æ/ with a higher F1 and a lower F2 compared to the Dutch /ɛ/.

For the second question, the expectation is that there will not be a difference in how learners produce the vowel in *and* as a discourse marker or as a coordinating conjunction. This is based on the findings of De Marco (2016) that L2 discourse markers, when in an initial position, have a similar function as filled pauses. When *and* functions as a coordinating conjunction, it connects clauses, words or phrases. In that use, it can therefore not appear in an utterance-initial position, as it needs to be preceded by the first of the two items that it is connecting. Therefore, when *and* has a sentence-initial position, it is most likely being used as a discourse marker. Thus, when used as a discourse marker, *and* will mainly have a similar function as filled pauses. De Boer and Heeren (2020) found that the first two vowel formants in filled pauses are language specific and therefore not transferred from the L1. There is therefore, based on the literature, little reason to expect a difference in realization of *and* as a discourse marker and *and* as a coordinating conjunction.

Methods

Speakers

35 speakers were selected from the Database of the Longitudinal Utrecht Collection of English Accents (D-LUCEA; Orr & Quené, 2017). D-LUCEA consists of recordings of students of the Utrecht University College (UCU). All the selected speakers were native speakers of Dutch. Only female speakers were selected to create a homogenous speaker group. All speakers were native speakers of Dutch without an audible accent and all were first-year students of UCU. To get admitted to UCU, Dutch students need to have followed at least 6 years of secondary education in English, and score at least an 8 out of 10 for English in secondary education (Quené et al., 2017). They are therefore estimated to have at least B2 proficiency level on the Common European Framework of Reference for Languages (Council of Europe, 2001; Fasoglio & Tuin, 2018). At the time of the used recordings, the students were at the start of their first year at UCU. Their ages ranged from 17 to 20 years and the mean age was 18.4 years (SD = 0.73 yr).

Materials and procedure

The recordings were done in a quiet furnished office with the participant and one or more facilitators. In the full recordings, the speakers performed several tasks, including reading passages,

formal and informal spontaneous monologues and a dialogue with the facilitator. Some tasks were done in both English and Dutch, other tasks were only done in English. For the current study, the 2-minute spontaneous informal monologues were used. For this task, speakers talked for 2 minutes about an informal topic that they chose themselves. Speakers typically talked about topics like their vacation, life at UCU, hobby's and their favorite books. The spontaneous informal monologues were done in both Dutch and English. Two speakers first did the task in English and then in Dutch, the other speakers did the task first in Dutch, followed by English.

The data was analyzed using PRAAT (Boersma & Weenik, 2022). To investigate the acoustic qualities of the vowel in *en* in Dutch and *and* in English, the onset and offset of the vowel and the nasal were manually segmented. Segmentation was done by placing boundaries at the onset of the vowel, where the signal showed voicing and at the offset of the vocalic and nasal part of the token. Using the waveform and spectrogram, segmentation was supported by repeated listening. To investigate if participants produced the vowel in *and* differently in *and* as a discourse marker compared to *and* as a coordinating conjunction, it was also indicated if the token was a discourse marker or a coordinating conjunction. As explained above, because *and* as a coordinating conjunction connects words, phrases or clauses, it is very unlikely it will have the function of a coordinating conjunction when in an utterance-initial position. In this study, it is therefore assumed that *and* in an utterance-initial position is used as a discourse marker. It was therefore indicated whether the token was in an utterance initial of non-initial position. Any instances of creaky voice were not included, as it has been shown that creaky voice can change formant frequencies (Moosmüller, 2001). This lead to a total of 669 /ɛ/ tokens and 603 /æ/ tokens. Of the Dutch /ɛ/ tokens, 366 were in an initial position and 303 in a non-initial position. Of the English /æ/ tokens, 305 were in an initial position and 298 in a non-initial position (see table 2). The spectral measurements were measured using a PRAAT script (see appendix A). The first and second formants (F1 and F2) were measured over the mid 50% of the vowel.

Table 2

Distribution of /ɛ/ and /æ/ vowels across language and word class

Language	Vowel	Word Class		
		CC	DM	Total
Dutch (L1)	/ɛ/	303	366	669
English (L2)	/æ/	298	305	603
Total		601	671	1272

Note. CC = coordinating conjunction, DM = discourse marker.

Statistical analysis

The statistical analysis was done using R (R Core Team, 2022) and package *lme4* (Bateset al., 2015). To analyze whether Dutch learners of English adapt the vowel in *and* to the target language, and if there is a difference in their realization of this vowel in an initial and non-initial position, a mixed-effects analysis was done. For both F1 and F2, fixed effects were Language (Dutch, English), Word Class (coordinating conjunction, discourse marker) and the interaction between Language and Word Class. For the factor Language, Dutch (L1) was the reference level. For the factor Word Class, coordinating conjunction was the reference level, as that is the main function of *and*. For the random effects, there was an intercept for speaker and a random slope for Language by Speaker and for Word Class by Speaker. No obvious deviations from homoscedasticity or normality were observed during visual inspection of residual plots. P-values were obtained through likelihood ratio testing in which predictors were included stepwise. First, fixed effects and their interaction were tested, afterwards, random effects and random slopes were tested.

Results

For F1, an interaction between Language and Word Class was included in the optimal model [$\chi^2(1) = 23.01$, $p < 0.001$]. F1 was higher for English compared to Dutch and this effect was bigger for Discourse Markers than for Coordinating Conjunctions, by approximately 40 Hz. Both Language and Word Class were also part of the optimal model as fixed effects [Language: $\chi^2(1) = 58.57$, $p < 0.001$; Word Class: [$\chi^2(1) = 225.78$, $p < 0.001$].] When speaking English, the F1 was about 20 Hz higher compared to Dutch. In a discourse marker, F1 was about 50 Hz higher compared to a coordinating conjunction. The optimal model also included a random slope for Language by Speaker [$\chi^2(2) = 8.01$, $p = 0.018$], and for Word Class by Speaker [$\chi^2(3) = 22.33$, $p < 0.001$]. This shows that both language and word class dependent changes in F1 varied per speaker (see figure 2). Table 3 gives an overview of the optimal models for F1 and F2.

For F2, an interaction between Language and Word Class was included in the optimal model [$\chi^2(1) = 8.93$, $p = 0.03$]. F2 was lower for English compared to Dutch and this effect was bigger for Discourse Markers than for Coordinating Conjunctions, by approximately 40 Hz. Both Language and Word Class were part of the optimal model as well [Language: $\chi^2(1) = 52.96$, $p < 0.001$; Word Class: $\chi^2(1) = 64.14$, $p < 0.001$]. When speaking English, F2 was about 30 Hz lower compared to Dutch. In a discourse marker, F2 was about 80 Hz higher compared to a coordinating conjunction. The optimal model also included a random slope for Language by Speaker [$\chi^2(2) = 9.82$, $p = 0.007$] and for Word Class by Speaker [$\chi^2(3) = 20.51$, $p < 0.001$]. This shows that both language and word class dependent changes in F2 varied per speaker (see figure 2). Table 4 shows the mean F1 and F2 of the vowels per condition.

Table 3

Optimal linear mixed-effect models predicting F1 and F2 of the vowel of en *in L1 Dutch and* and *in L2 English*

	F1		F2	
	Coefficient (SE)	t	Coefficient (SE)	t
Intercept	614 (10.9)	56.3	1920 (20.1)	95.7
Language: English	21 (8.1)	2.6	− 31 (12.4)	− 2.5
Word Class: discourse marker	53 (8.7)	6.1	82 (12.4)	6.6
Interaction Language: English & Word Class: discourse marker	43 (9.1)	4.7	− 43 (14.6)	− 3.0

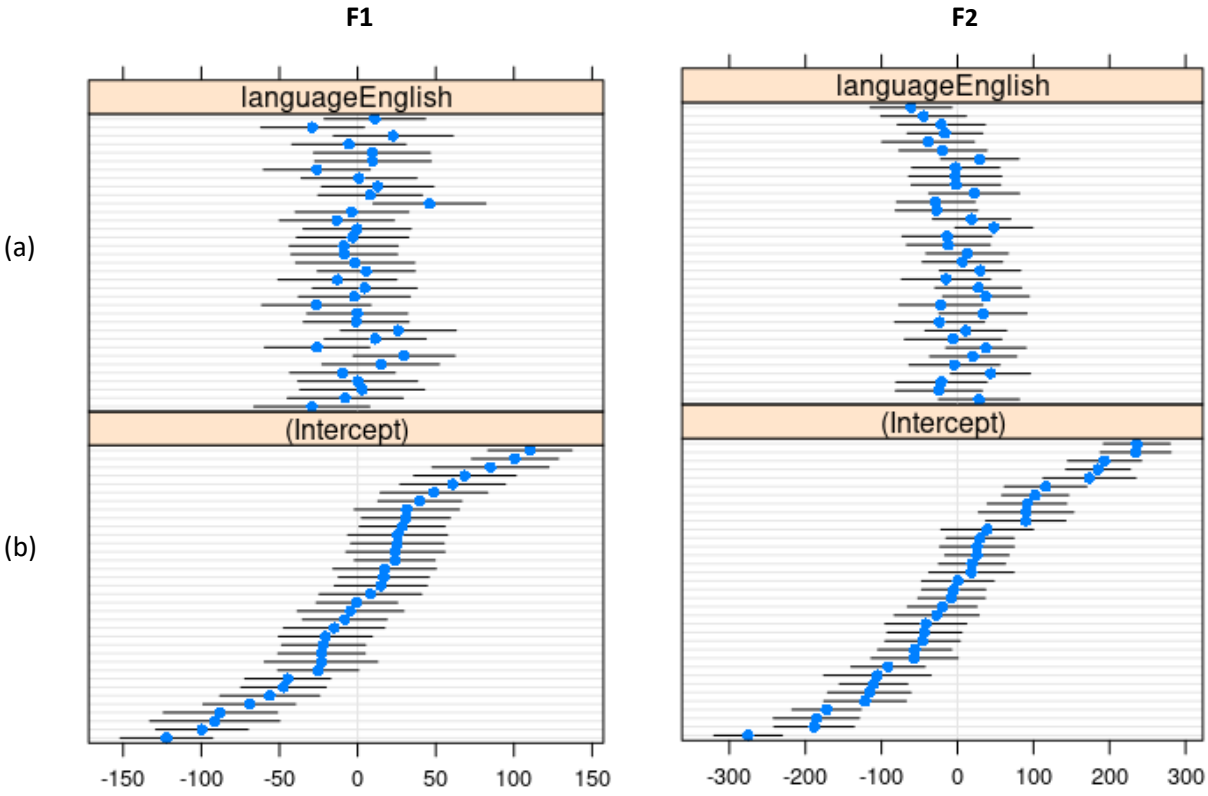
Table 4

Mean F1 and F2 by condition in Hertz

Means (Hz)	Dutch (L1)		English (L2)		Overall	
	F1	F2	F1	F2	F1	F2
CC	614	1920	635	1899	627	1909
DM	666	2002	680	1848	698	1974
Overall	644	1971	686	1911		

Figure 2

Caterpillar plots showing the by-speaker intercepts and L2 adaptation for the F1 and F2 models



Note. Horizontal lines represent the speakers, arranged from lowest to highest mean F1 and F2 in Dutch (L1). In (a), x = 0 is the intercept over all speakers, (b) shows by-speaker adaptations relative to their intercepts when speaking English (L2). (0 means no adaptation, -300 is the greatest adaptation)

Discussion

In this study, I investigated whether L1 Dutch speakers of L2 English adapt the vowel in *and* to the target language, or whether they transfer the vowel /ɛ/ from the L1. I also investigated whether there is a difference in realization of the vowel in *and* as a discourse marker or as a coordinating conjunction. Dutch does not have the /æ/ vowel, and Dutch learners have been found to perceive the English /æ/ vowel as a same category vowel as the /ɛ/ vowel (Broersma et al., 2010). The three different L2 sound learning models, PAM-L2, L2LP and SLM, all predict that Dutch learners of English would therefore only be able to approximate the native pronunciation. Because the speakers in this study are all proficient speakers, it was however expected that they would adapt the vowel towards the target language. In both American English and British English, the F1 of /æ/ is higher than of /ɛ/, and the F2 of /æ/ is lower than of /ɛ/ (Ghorshi et al., 2008). The expectation therefore was that participants would produce the English /æ/ with a higher F1 and a lower F2 compared to the Dutch /ɛ/.

In line with these predictions, speakers indeed produced the vowel in English differently than the vowel in Dutch. When speaking English, the F1 was higher and the F2 was lower. This shows that proficient learners of English indeed adapt the vowel of *and* to the target language. This is line with the findings of Flege (1997), who found that Dutch speakers who were judged to have a relatively mild foreign accent when speaking English produced the /æ/ with a higher F1 and a lower F2 compared to Dutch speakers who were judged to have a relatively strong foreign accent when

speaking English. The shift towards the target pronunciation also shows that these proficient speakers seem to at least to some extent have learned to make a distinction between /ɛ/ and /æ/ and not place /æ/ in the same category as /ɛ/.

The second question of this study was whether the speakers produced the vowel in *en* and *and* differently in a discourse marker compared to a coordinating conjunction. Because there have not been any studies on the use of *and* as a discourse marker in L2, nor on the phonetic L2 realization of vowels in discourse markers, there is little literature to base the expectations on. It has however been shown that discourse markers can be used as filled pauses (De Marco, 2016). Filled pauses are produced relatively unconsciously and they have therefore been theorized to be transferred from the L1 to the L2 (Clark & Fox Tree, 2002; De Leeuw, 2007). When investigating this claim however, studies did not find evidence for the L1-transfer but found instead that filled pauses, and specifically the vowels, are language specific and are not fully transferred from the L1 to the L2 (De Boer & Heeren, 2020; Rose, 2017). It was therefore not expected that word class would have an effect on the realization of the L2 /æ/ vowel.

Contrary to these predictions, the results did show a difference in pronunciation between the discourse marker and coordinating conjunction. In both L1 Dutch and L2 English, F1 was higher for discourse markers than for coordinating conjunctions. This effect was larger in English than it was in Dutch. F1 is related to the degree of openness of vowels, so in discourse markers, participants produced the vowel more open compared to coordinating conjunctions. F2 was lower in English for discourse markers compared to coordinating conjunctions, but higher in Dutch. F2 is related to the degree of backness of vowels, meaning that in English the vowel was more back in discourse markers compared to coordinating conjunction, whereas in Dutch it was more front. In Dutch, the /ɛ/ vowel in coordinating conjunction was more closed and more back compared to discourse markers, essentially more central. An explanation for this result could be that *en* as a coordinating conjunction is less likely to be stressed, as it is between two content words. Although *en* can be used to emphasis both the conjoined parts, in which case it would be stressed, this is not the most common way that *en* is used. It is therefore plausible that when used as a coordinating conjunction, *en* is less likely to be stressed. In Dutch, unstressed vowels are often reduced in their spectral quality (i.e. centralized) (Van Bergem, 1993). Because the discourse markers are used as filled pauses, they are more often prolonged and also more likely to be stressed and thus less likely to be reduced.

When speaking English, the speakers produced the vowel in coordinating conjunctions more closed and more front compared to discourse markers. This means that they adapt the /æ/ vowel more to the target language and pronounce it more native-like in discourse markers. One explanation for this could again be vowel reduction. In English, unstressed vowels are realized as schwa's ([ə]). This sound is more centralized than /æ/, meaning more back and more closed. Just as in Dutch, *and* as a conjunction coordinator is less likely to be stressed, and the vowel would therefore be more likely to be reduced. This does indeed happen for the F1, which is lower in coordinating conjunctions compared to discourse markers, meaning the vowel is more closed. The F2 however, is lower in discourse markers compared to coordinating conjunctions, meaning the vowel is more front. This is however closer to native production. The speakers have adapted the F2 more to the target language in discourse markers than in coordinating conjunctions. The effect of language, lowering the F2, therefore seemingly outweighs the reduction effect.

The results of this study show that proficient L1 Dutch speakers of L2 English do not transfer the vowel from *en* without adapting it. This was the case when *and* was used as a discourse marker as well as when it was used as a coordinating conjunction. When performing cross-linguistic comparisons, the most useful features are those that are largely language-independent within

speakers. Based on the results found in this study, the /æ/ vowel is not language-independent for proficient L1 Dutch speakers of L2 English, and is therefore not an appropriate feature to use in cross-linguistics speaker comparison.

Since there were no native speakers in this study, we can only make conclusions about the direction of the adaptation of the vowel of the L1 Dutch speakers, but not about how close they were to native pronunciation. Future studies could compare native speakers doing the same tasks as non-native speakers to get a clearer picture of how close to nativelike production L2 speakers are. The participants of this study all just started at UCU at the time of recording and had at least a B2 proficiency level. The L2 sound learning models discussed in the current study all predict that the /ɛ/-/æ/ distinction is difficult to learn for L1 Dutch learners of L2 English, and only proficient speakers will have different categories for the two sounds. Future research could explore how proficiency level affects the production of the English /æ/ vowel by Dutch speakers of English and if the realization of this vowel differs in different contexts or word classes. The current study only compared vowel formants *and* and *en.* There are many more aspects that could be language-independent and would therefore be useful to explore. In their study of cross-linguistic realization of filled pauses of Dutch speakers of English, de Boer and Heeren (2020) found that whereas F1 and F2 were language-dependent, fundamental frequency and duration did remain consistent across languages. Further research might therefore determine if these features also remain similar across languages in conjunctions or discourse markers.

Conclusion

This study aimed to determine which aspects of speech can be useful in cross-linguistics comparisons. If L1 transfer in the vowel /æ/ occurred, this segment could be included in cross-linguistics comparisons. Such transfer was not found, proficient L1 Dutch speakers of L2 English adapted the vowel in *and* to the target language. This effect was bigger in discourse markers compared to coordinating conjunctions, but present in both uses of *and*. /æ/ is therefore not a useful segment for cross-linguistic speaker comparison for Dutch speakers of English.

References

Bates, M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Best, C. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 171-204). York Press.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. Munro, & O. Bohn, *Language Experience in Second Language Speech Learning : In Honor of James Emil Flege* (pp. 13-34). John Benjamins Publishing Company.

Boersma, P., & Weenik, D. (2022). *Praat: doing phonetics by computer [computer program]*. Retrieved from www.praat.org

Bohn, O., & Flege, J. (1992). The production of new and similar vowels by German learners of English. *Studies in Second Language Acquisition, 14*(2), 131-158.

Bradlow, A., Pisoni, D., 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.

Brinton, L. (1996). *Pragmatic Markers in English: Grammaticalization and Discourse Functions.* De Gruyter Mouton.

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

Broersma, M., & Cutler, A. (2008). Phantom word activation. *System: An International Journal of Educational Technology and Applied Linguistics, 36*(1), 22-34.

Broersma, M., Aoyagi, M., & Weber, A. (2010). Cross-linguistic Production and Perception of Japanese- and Dutch-accented English. *Journal of the Phonetic Society of Japan, 14*(1), 60-75.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behaviour Research Methods, 44*(4), 997-997.

Clark, H., & Fox Tree, J. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*(1), 73-111.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Press Syndicate of the University of Cambridge.

Curzan, A., & Adams, M. (2006). *How English works: A linguistic introduction* (3rd ed.). Pearson Longman.

De Boer, M., & Heeren, W. F. (2020). Cross-linguistic filled pause realization: The acoustics of uh and um in native Dutch and non-native English. *The Journal of the Acoustic Society of America, 148*(6), 3612-3622.

De Leeuw, E. (2007). Hesitation Markers in English, German, and Dutch. *Journal of Germanic Linguistics, 19*(2), 85-114.

De Marco, A. (2016). The use of discourse markers in L2 Italian: A preliminary investigation of acoustic cue. *LIA: Language, Interaction and Acquisition, 7*(1), 67-88.

Escudero, P. (2005). *Linguistic perception and second language acquistion: explaining the attainment of optimal phonological categorization.* [Doctoral dissertation, Utrecht University].

Fasoglio, D., & Tuin, D. (2018). *Speaking skill levels of English attained in Dutch secondary education.* SLO.

Flege, J. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issue in cross-language research* (pp. 229-273). York Press.

Flege, J. (1997). English vowel productions by Dutch talkers: more evidence for the "similar" vs "new" distinction. In A. James, & J. Leather (Eds.), *Second-Language Speech: Structure and Process* (pp. 11-52). Mouton de Gruyter.

Flege, J., & Bohn, O. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland, *Second Language Speech Learning* (pp. 3-83). Cambridge University Press.

Flege, J., & Fletcher, K. (1992). Talker and listener effects on degree of perceived foreign accent. *The Journal of the Acoustical Society of America, 91*(1), 370-389.

Flege, J., Yeni-Komshian, G., & Liu, S. (1999). Age Constraints on Second-Language Acquisition. *Journal of Memory and Language, 41*(1), 78-104.

Foulkes, P., & French, P. (2012). Forensic Speaker Comparison: A Linguistic–Acoustic Perspective. In L. Solan, & P. Tiersma (Eds.), *The Oxford Handbook of Language and Law* (pp. 557 - 572). Oxford University Press.

Fraser, B. (1988). Types of English Discourse Markers. *Acta Linguistica Hungarica, 38*(1), 19-33.

French, P., & Stevens, L. (2013). Forensic Speech Science. In M. Jones, & R.-A. Knight (Eds.), *The Bloomsbury Companion to Phonetics* (pp. 183-197). Bloomsbury Academic.

García-Amaya, L., & Lang, S. (2020). Filled pauses are susceptible to crosslanguage phonetic influence: Evidence from Afrikaans-Spanish. *Studies in Second Language Acquisition, 42*(5), 1-29.

Ghorshi, S., Vaseghi, S., & Yan, Q. (2008). Cross-entropic comparison of formants of British, Australian and American English accents. *Speech Communication, 50*(7), 564-579.

Gilquin, G. (2016). Discourse markers in L2 English. In O. Timofeeva, A.-C. Gardner, A. Honkapohja, & S. Chevalier, *New Approaches to English Linguistics: Building Bridges* (pp. 213-249). John Benjamins Publishing Company.

Gleitman, L. R. (1965). Coordinating Conjunctions in English. *Language, 41*(2), 260-293.

Gut, U. (2007). Foreign Accent. In C. Müller (Ed.), *Speaker Classification I: Fundamentals, features, and methods* (pp. 75-87). Springer-Verlag Berlin Heidelberg.

Hansen, M.-B. M. (1997). Alors and donc in spoken French: A reanalysis. *Journal of Pragmatics, 28*(2), 153-187.

Hughes, V., Foulkes, P., & Wood, S. (2016). Strength of forensic voice comparison evidence from the acoustics of filled pauses. *International Journal of Speech, Language and the Law, 23*(1), 99-132.

IAFPA. (2020, September). *Code of Practice*. Retrieved from https://www.iafpa.net/about/code-of-practice/

Isbell, D. (2016). The Perception-Production Link in L2 Phonology. *MSU Working Papers in Second Language Studies, 7*(1), 57-67.

Keuleers, E., & Brysbaert, M. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behaviour Research Methods, 42*(2), 643-650.

Moosmüller, S. (2001). The influence of creaky voice on formant frequency changes. *International Journal of Speech, Language and the Law, 8*(1), 100-112.

Motohashi-Saigo, M., & Hardison, D. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Technology, 13*(2), 29-47.

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in Second Ļanguage Acquisition, 21*, 81-108.

Müller, S. (2005). *Discourse Markers in Native and Non-native English Discourse.* John Benjamins Publishing Co.

Nejjari, W., Gerritsen, M., Van der Haagen, M., & Korzilius, H. (2012). Responses to Dutch-accented English. *World Englishes, 31*(2), 248-267.

Nolan, F. (1983). *The phonetic bases of speaker recognition.* Cambridge University Press.

Orr, R., & Quené, H. (2017). D-LUCEA: Curation of the UCU Accent Project Data. In J. Odijk, & A. van Hessen (Eds.), *CLARIN in the Low Countries* (pp. 181-193). Ubiquity Press.

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

Quené, H., Orr, R., & Van Leeuwen, D. (2017). Phonetic similarity of /s/ in native and second language: Individual differences in learning curves. *The Journal of the Acoustical Society of America, 142*(6), EL519-EL524.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing.

Romero Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics, 34*(6), 769-784.

Rose, R. (2017). A comparison of form and temporal characteristics of filled pauses in L1 Japanese and L2 English. *Journal of the Phonetic Society of Japan, 21*(3), 33-40.

Schourop, L. (1999). Discourse markers. *Lingua*, 227-265.

Schriffin, D. (1987). *Discourse Markers.* Cambridge University Press.

Schriffin, D. (2006). Discourse marker research and theory:. In K. Fischer, & M.-B. M. Hansen (Eds.), *Approaches to Discourse Particles* (pp. 315-338). BRILL.

Shimada, K. (2014). Contrastive Interlanguage Analysis of Discourse Markers Used by Nonnative and Native English Speakers. *JALT Journal, 36*(1), 47-67.

Tahta, S., Wood, K., & Loewenthal, K. (1981). "Foreign accents: Factors relating to transfer of accent from the first language to a second language. *Language and Speech, 24*(3), 265-272.

Tyler, M. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. In A. Nyvad, M. Hejná, H. A., A. Jespersen, & M. Sørensen (Eds.), *A sound approach to language matters – In honor of Ocke-Schwen Bohn* (pp. 607-630). Department of English, School of Communication and Culture, Aarhus University.

Van Bergem, D. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication, 12*(1), 1-23.

van der Vloed, D., & Bouten, J. V. (2014). NFI-FRITS: A forensic speaker recognition database and some first experiments. *Proceedings of the Odyssey Speaker and Lang. Recogn. Workshop*, (pp. 6-13). Joensuu, Finland.

Van Heuven, W., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology, 67*(6), 1176-1190.

Van Leussen, J., & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology, 6*, 1-12.

Zhao, H. (2013). A Study on the Pragmatic Fossilization of Discourse Markers among Chinese English Learners. *Journal of Language Teaching and Research, 4*(4), 707-714.

Appendix A – PRAAT script used for data analysis

```
# for the acoustic analysis of vowel realizations
#    1. duration measurements
#    2. formant measurement, F1 & F2, in the vocalic part
#
#    30-05-2022, by W.F.L. Heeren for D. Slump
#

# here fill in your own input and output directories
# the input directory contains wave files and textgrids
inDir$ = "D:\Data"
outDir$ = "D:\Data"

# make a list of the wave file names in the directory
Create Strings as file list: "list", inDir$ + "\*.wav"
numberOfFiles = Get number of strings

# initialize the output file
appendFile: "vowel-measurements.txt", "filename", tab$, "segment", tab$, "StartTime", tab$, "language", tab$, "position", tab$,
"durationVowel", tab$, "durationN", tab$, "F1", tab$, "SD_F1", tab$, "F2", tab$, "SD_F2", newline$

# go through the list of names, one by one
for ifile from 1 to numberOfFiles

        select Strings list
        fileName$ = Get string... ifile
        # open the wave file
        Read from file: inDir$ + "\" + fileName$
        name$ = selected$("Sound")
                idSnd = selected("Sound")
                        To Formant (burg): 0, 3, 3500, 0.025, 50

                # open the corresponding textgrid
        Read from file: inDir$ + "\" + name$ + "_SIL_orth.TextGrid"
        numberOfIntervals = Get number of intervals: 5

        # go through the textgrid, interval by interval, check if there's an 'en/and'; if so, do measurements
        for intervalNumber from 1 to numberOfIntervals
                selectObject: "TextGrid 'name$'" + "_SIL_orth"
                label$ = Get label of interval: 5, intervalNumber
                if label$ != ""
                                        selectObject: "TextGrid 'name$'" + "_SIL_orth"
                                        appendFile: "vowel-measurements.txt", name$, tab$, label$, tab$

                                        # measure the filled pause's duration
                                        selectObject: "TextGrid 'name$'" + "_SIL_orth"
                                        start = Get start time of interval: 5, intervalNumber
                                        end = Get end time of interval: 5, intervalNumber


                                        # get language from tier 1
                                        duration = end - start
                                        midInterval = start + duration / 2
                                        phraseNumber = Get interval at time:  1,  midInterval
                                        language$ = Get label of interval: 1, phraseNumber
                                        positionNumber = Get interval at time:  6,  midInterval
                                        position$ = Get label of interval: 6, positionNumber

                                        appendFile: "vowel-measurements.txt", start, tab$, language$, tab$, position$, tab$

                                        # measure the segments' duration
                                        if label$ == "n"
                                                durationN = end - start
                                                appendFile: "vowel-measurements.txt", "", tab$, durationN
                                        else
                                                durationVowel = end - start
                                                appendFile: "vowel-measurements.txt", durationVowel, tab$, "", tab$

                                        endif
```

```
                                        # measure the vowel's F1 and F2
                                        if label$ == "n"
                                                appendFile: "vowel-measurements.txt", newline$

                                        elsif label$ != "n"
                                                # define measurement interval
                                                quart = duration / 4
                                                mid = duration / 2
                                                fstart = (mid - quart) + start
                                                fend = (mid + quart) + start
                                                selectObject: "Formant 'name$'"
                                                f1 = Get mean: 1, fstart, fend, "hertz"
                                                b1 = Get standard deviation: 1, fstart, fend, "hertz"
                                                f2 = Get mean: 2, fstart, fend, "hertz"
                                                b2 = Get standard deviation: 2, fstart, fend, "hertz"
                                                appendFile: "vowel-measurements.txt", f1, tab$, b1, tab$, f2, tab$, b2, newline$
                                        end

                                endif
                        endif
        selectObject: "TextGrid 'name$'"+ "_SIL_orth"
        endfor


        # clean up
        select all
        minusObject: "Strings list"
        Remove

endfor

# clean up
selectObject: "Strings list"
Remove
```