The relationship between perception and production of English /d/-/t/ in initial position by Dutch native speakers

Student: Lisa Darko

Student ID: 10751335

Supervisor: Silke Hamann

BA English language and culture

Date: 30 June 2017

Acknowledgements

I hereby declare that I have read the UvA guidelines on Plagiarism and confirm that this thesis is my own work.

I would like to thank my thesis supervisor Silke Hamann for her help and support setting up the experiment and for her useful feedback on the paper. In addition I would also like to thank everyone who was willing to participate in the experiment.

Table of contents

1 Introduction

For second language (L2) speakers, sounding ‘native-like’ can be beneficial in multiple ways. Not accurately pronouncing L2 sounds, i.e. having a ‘nonnative’ accent, can lead to confusion or even incomprehensibility on the listener's part. It could also influence how the speaker is viewed socially (e.g. Schairer, 1992; van der Doel, 2006, Nejjari, Gerritsen, Haagen, & Korzilius, 2012). For these reasons, sounding like a native speaker is desirable for many L2 speakers.

There has been much debate on how native-like speech in the L2 can be obtained. The critical period hypothesis (CPH) (first proposed in 1959 by Penfield and Roberts in *Speech and Brain Mechanisms*) hypothesizes that an L2 can be produced with native-like pronunciation only when learned in ‘the critical period’ (roughly up to the age of twelve). Any language that is learned after this period has passed will not sound native-like. Evidence for the CPH has been found by many studies that have shown that L2 speakers who acquired the L2 later in life have stronger nonnative accents than those who learned the L2 in childhood (e.g., Flege, Munro, & MacKay, 1995; Johnson & Newport, 1991). The critical period the CPH has also received counterevidence, as several studies showed that some late L2 learners are able to speak an L2 without any detectable nonnative accent (e.g., Bongaerts, Van Summeren, Planken, & Schils, 1997; Moyer, 1999).

Other theories hypothesize that native-like pronunciation can be obtained only when the L2 speaker can distinguish the L2 sounds accurately, on both a phonemic and phonetic level. Most well-known theories are Flege’s Speech Learning Model (SLM), (Flege, 1995) and Best’s Perceptual Assimilation Model (PAM), (Best, 1994). According to these theories, when perceiving an L2, L2 learners categorize the sounds of that L2 according to the existing categories in their L1. Pronunciation problems in the L2 then arise when L2 learners are not able to distinguish between L2 sound contrasts because of conflicting contrasts and influences in the L1. Therefore L2 speakers have to perceive an L2 sound as a category of the L2 instead of the L1, in order to be able to pronounce the L2 sound accurately. An accurate perception thus precedes an accurate production, according to these theories.

No conclusive evidence has been found on whether perception indeed proceeds production, or whether it follows production when it comes to L2 acquisition. With regard to pronunciation teaching, it is important to know whether an accurate perception is necessary to obtain a native-like pronunciation. The present study aims to give more insight into this controversial issue.

This paper examines the production and perception accuracy of Dutch L2 speakers of English. Its interest is to investigate the perception-precedes-production theory with regard to consonants, specifically, the /d/- /t/ contrast in word initial position. It focuses on consonants, because most research that has been done on the subject has focused on vowel perception and production, and it focuses on /d/ and /t/ in word initial position, rather than word final position, because in final position, Dutch does not have the /d/- /t/ contrast (because of final devoicing), and English is losing it as well, therefore it would be difficult to judge whether a pronunciation is English-like or not. Both English and Dutch allow the phonemes /t/ and /d/ in initial position, but the English realization differs from the Dutch realization in Voice Onset Time (VOT): the interval between the start of the release of a plosive, and the onset of voicing. Due to time restriction the choice was made to focus only on the alveolar plosives.

Consequently, the research question of the current study is as follows: is the perception accuracy of the English /d/ and /t/ in word initial position by Dutch L2 learners positively correlated with the production accuracy of the English /d/ and /t/? It is hypothesized that if the perception of /t/ and /d/ closely resembles native speakers' perception of /d/ and /t/, then the production reflects this, in that the production has a native-like realization of /d/ and /t/, meaning that the perception accuracy is indeed positively correlated with the production accuracy.

Subjects participated in an experiment consisting of a production and a categorization task. In the production task they had to read words with /d/ and /t/ in initial position. In the categorization task they were presented with stimuli with a /d/ or /t/ in initial position, and had to decide whether they heard a /d/ or /t/. If the results show that perception accuracy correlates with production accuracy, then it is evidence for the perception-precedes-production theory.

2 Theoretical framework

2.1 Relationship between production and perception

Whether perception precedes or follows production in the case of L2 acquisition is subject to much debate. Substantial research has been done on the relationship between production and perception of English speech sounds by L2 speakers (e.g. Bolanos, 2013; Flege, 1993; Flege, Bohn, & Jang, 1997; Flege, Mackay & Meador, 1999; Lopez, Yu, & Bell-Berti, 2006; Rauber, Escudero, Bion & Baptista, 2005; Sheldon & Strange, 1982). Research in this area has focused primarily on the perception and production of vowels (e.g. Flege, 1993 Flege et al., 1997; Flege et al. 1999; Lopez et al., 2006; Rauber et al.) while the relationship between perception and production of consonants has received less attention, although studies do exist (e.g. Flege, 1993; Sheldon & Strange, 1982; Bolanos, 2013).

2.1.1 Perception precedes production

Some theories on L2 acquisition argue that perception precedes production. One of those theories has been proposed by Flege (1995), as part of his Speech Learning Model (SLM). The SLM was a reaction to the critical period hypothesis. According to one hypothesis of the SLM, production errors are often the result of perception errors – perception accuracy thus limits production accuracy. Flege states that "the phonetic systems used in the production and perception of vowels and consonants remain adaptive over the life span, and that phonetic systems reorganize in response to sounds encountered in an L2 through the addition of new phonetic categories, or through the modification of old ones" (1995, p. 233).

This means that if an L2 learner is able to perceive a difference between the L2 sound and the closest sound to it in the L1, a new phonetic category can be created, but if the L2 speaker cannot perceive a difference, no new phonetic category will be created. Flege calls this ‘equivalence classification’: one phonetic category is used to process both the L2 sound and the L1 sound closest to it. When producing the L2 sound, the L2 speaker will make use of his or her L1 phonetic category and pronounce the sound with L1 qualities. In this way perception influences production.

Many studies that tested the SLM have found evidence for this theory, but many of these studies were carried out by Flege himself, as can be observed in his overview of studies on the relationship between production and perception (Flege, 1999). In one of his studies he examined the production and perception of English vowels by highly experienced native Italian speakers and found a significant correlation (Flege et al., 1999). In another study he assessed the effect of English language experience on nonnative speakers' production and perception of English vowels. The experienced nonnative subjects produced and perceived English vowels more accurately than the inexperienced nonnative subjects and Flege also states that the nonnative subjects' degrees of accuracy in producing and perceiving English vowels were related because acoustic-based predictions concerning how the various groups of nonnative subjects would perform in the vowel perception experiment were largely supported (Flege et al., 1997).

Other studies on the relationship between vowel production and perception include Lopez et al. (2006) and Rauber et al. (2005). Lopez et al. investigated the relationship between production and perception of English vowels by Mandarin L1 speakers and found that the more accurately the L2 speaker distinguished a vowel contrast, the more distinctly he or she produced that contrast. Rauber et al. examined highly proficient Brazilian L1 speakers. The two English vowel sounds that were poorly discriminated were produced with values similar to those of a single vowel in the learners' L1. These findings support the SLM theory of equivalence classification. Rauber et al. state that the results of their study “[give] more evidence to the fact that perception precedes production of sounds and that it is a prerequisite for accurate L2 production.” (p. 2916).

One study that focuses on consonants, is Flege’s (1993) *Production and perception of a novel, second-language phonetic contrast*. In this study Chinese subjects’ production and perception of the /t/- /d/ contrast in final position of English words were examined. Vowel duration was measured to determine the subjects’ differences in producing /t/ and /d/. Early and late learners were compared and the late learners’ vowel duration differences were much smaller than those of native speakers, and were correlated significantly with the degree of foreign accent in English. Moreover, the differences in L2 accuracy between the early and late learners were consistent with the hypothesis that L2 production accuracy is limited by the adequacy of perceptual representations for sounds in the L2. The data analysis of the individual differences did not support this hypothesis however.

2.1.2 Production precedes perception

Several studies (e.g., Borrell, 1990; Neufeld, 1988; Brière, 1966) have pointed out that when learning an L2, not all sounds that are correctly perceived are also correctly produced. Other studies have actually found evidence that production precedes perception.

One of those was the study by Goto in 1971. He reported findings of L2 speakers with a more accurate production than perception: Japanese speakers of English attempting to distinguish between the /l/- /r/ contrast, which does not exist in the Japanese language. In 1982, Sheldon and Strange replicated and extended Goto’s study and found that Japanese subjects with a high level of English had a relatively good pronunciation of /l/ and /r/ but poor auditory discrimination. They state that "at least for the contrast studied here, perceptual mastery of a foreign contrast does not necessarily precede adult learners' ability to produce acceptable tokens of the contrasting phonemes, and may, in fact, lag behind production mastery" (Sheldon & Strange, 1982, p. 245).

Flege & Eefting (1987b) used two language sets to test Dutch and English speakers' perception and production for both Dutch and English. They found that Dutch speakers produced a substantial VOT difference between Dutch and English, which indicated a good discrimination between the two languages. However, in the perception test, even the most proficient Dutch speakers (English majors) only showed a small shift in the location of perception boundaries for Dutch and English, indicating a poor discrimination between the two languages, which means that in this case perception did not precede production. Flege and Eefting speculate that the most proficient Dutch subjects had established a new phonetic category for the English /t/ and made use of their English phonetic category when identifying the stimuli in both the Dutch and English language set, and that is why the phoneme boundary shift is so small.

In Llisterri's (1995) overview of studies that examined the relationship between production and perception, Llisterri summarizes that it is not easy to establish a direct correlation between production and perception in an L2, and that it does not seem possible to infer production abilities from perceptual abilities or the other way around. He states that the link between perception and production is complex and that more factors need to be taken into account when it comes to examining the relationship between the two. For example, experience and language use seem to have a more marked influence on production than on perception, and social factors such as pressure to improve production may provide an explanation for cases in which production precedes perception. In addition, Llisterri argues, the relationship between production and perception might not only be influenced by contextual dependency, but also by the nature of acoustic cues (such as VOT).

2.2 Word initial /d/- /t/ contrast: Dutch vs. English

In both English and Dutch, phonemes /t/ and /d/ contrast phonologically in voicing. However, they are realized differently at the phonetic level: English and Dutch /t/ and /d/ differ in VOT. The term VOT was introduced by Lisker and Abramson (1964). They analyzed initial stop consonants in 11 languages and created three VOT categories: voicing lead or negative VOT (prevoicing), short lag VOT and long lag VOT (aspiration). The 11 languages were divided into three groups, depending on whether they have a maximum of two, three or four categories of plosives at each place of articulation. Dutch and English both are two-category languages. Lisker and Abramson observed that in the case of isolated words, the 6 two-category languages broadly fell into three VOT ranges: -125 ms to -75 ms (voicing lead), 0 ms to +25 ms (short lag), and +60 ms to +100 ms (long lag). The median values for these ranges were -100 ms, +10 ms and +75 ms respectively. Dutch, they say, occupies the ranges of about -100 and +10 ms, and English those of +10 and +75 ms.

2.2.1 VOT of English

Previous research that examined the production and perception of VOT by English L1 speakers has found various results. An overview of the results of four influential studies are shown in figure 1. Based on these previous results two ranges of English VOT (short lag and long lag) were determined to use for comparison in the present study.

Figure 1: Previous results of /d/ and /t/ VOT (in ms) by native English speakers

	Lisker & Abramson, 1964		**Flege & Eefting, 1987a**	**Klatt, 1975**	**Docherty, 1992**
	Mean	**Range**	**Mean**	**Mean**	**Mean**
/t/	70	30-105	89	65	64
/d/	5	0-25	18	23	21

From figure 1 it can be observed that in English, /t/ is realized with long lag VOT values and /d/ with short lag values (applying Lisker & Abramson's categories). It is clear that VOT values vary considerably between speakers: /t/ is produced with values ranging from 30 ms to 105 ms, and /d/ ranges from 0 ms to 26 ms.

2.2.2 VOT of Dutch

In Dutch, /d/ in initial position is generally realized with voicing lead (vibration of the vocal folds starts before the stop release) (Lisker & Abramson, 1964; Van Dommelen, 1983; Van Alphen & Smits, 2004). However, Van Alphen and Smits reported that, across the 10 speakers they examined, prevoicing was absent in 25% of the voiced plosives, showing that Dutch speakers not always produce voiced plosives with prevoicing. In the current study the subjects are expected to produce the /d/ with prevoicing more often than without prevoicing.

The Dutch /t/ is generally an unaspirated stop with short lag VOT values (Lisker & Abramson, 1964; Van Alphen & Smits, 2004). Dutch /t/ is thus realized with a similar short lag VOT as English /d/. Research did find that when speaking English, Dutch speakers tend to produce /t/ with longer VOT values than when speaking Dutch. Flege and Eefting (1987b) found that almost every Dutch subject had a longer mean VOT for English than for Dutch, although still not as long as that of native speakers. Interestingly, while English majors reached a mean long lag VOT value while speaking English, people that did not study English still had a mean short lag VOT.

Flege (1995) argues that "category formation for English stops may be prevented by the continued perceptual linkage of L1 and L2 sounds (i.e., by equivalence classification)" (p. 258). If the nonnative difference cannot be perceived and consequently no L2 categories can be formed for /t/ and /d/, this means, according to Flege, that /t/ and /d/ in the L2 will be produced with L1 VOT values, instead of L2 values. In the present study, the majority of native Dutch subjects are therefore expected to produce the /t/ in English words with short lag VOT values, which reflect the VOT values of their native language, instead of long lag values, as is common for native English speakers. For the perception task it is predicted that the subjects tend to identify the English /d/ as /t/, because the English /d/ is produced with the same VOT values as the Dutch /t/. The perception boundary between English /t/ and /d/ for the Dutch listeners is thus expected to be different from that of native English listeners. Because Dutch listeners presumably hear a /t/ at short lag and long lag values, while native listeners presumably only hear a /t/ at long lag values, Dutch listeners are expected to perceive a /t/ more often than native English listeners would.

3 Method

Two tasks were carried out to test subjects' perception and production of /d/ and /t/: a production task and a categorization task. In the production task subjects had to read words with /da/ or /ta/ in initial position and in the categorization task they were presented with /da/-/ta/ stimuli.

3.1 Participants

A total of 20 native Dutch speakers participated in this study. As the effects of experience, age, motivation, language experience and language use lays outside the scope of the study, these variables were not taken into account. For the purpose of the current study, variation in accuracy ratings is desirable because in that case the relationship between perception and production is easier to examine. If subjects are more experienced and therefore perceive the phonemes more accurately, then it is in the interest of the study to see if their production then positively correlates. Therefore variation in language proficiency among the subjects was desirable, however, due to the difficulty of finding participants, no time was spend on making sure subjects varied in language proficiency. Nonetheless variation of language experience was present, as some subjects study English at university (10%) and all other subjects' years of formal education of English varied from 4 to 9 years. All subjects were monolingual, with Dutch as L1, learned English in a formal setting (i.e. at school) (TV and other media was not taken into account although it presumably had an effect) and had not started learning English in formal education before the age of 10.

As testing in the sound lab at the University of Amsterdam meant that subjects would have to come to the university, which was an extra obstacle for them and difficult to do in the given timespan, the experiment was carried out at the subjects' homes. Background noise was eliminated as much as possible, sufficient for the recordings to be satisfactory analyzed. The experiment, including instructing time, took about 10 minutes to complete. Subjects did not receive any reward for participating in the experiment.

3.2 Stimuli

3.2.1 Production

As shown in Appendix 1, there were 10 test items in total: 5 words with /da/ in initial position and 5 words with /ta/ in initial position. In addition to the test items, there were 20 filler items, which did not contain a /t/ or /d/ in initial position. Some of the filler items were chosen because they were minimal pairs or rhyming words with the test items, with the goal of distracting the subjects. Test items and filler items were placed in a random order in a PowerPoint presentation. The order was the same for each subject. All items were placed in a sentence, see Appendix 1, to make speech more natural.

A Marantz professional recorder with a sampling frequency of 48 kHz was used to record the output of the participants.

3.2.2 Perception

A /da/ – /ta/ continuum was created prior to testing. The creation of the continuum points was manually done in Praat. A sample of a male Dutch speaker was used for endpoint /ta/, the most aspirated sound of the continuum. The burst was manipulated by shortening the aspiration noise, to make the sound more subtle and hopefully more ambiguous. The midpoint of the continuum, /da-ta/ (the stimulus that was intended to be the most ambiguous, because it reflects by convention the Dutch realization of /t/ but the English realization of /d/, and that the subjects expectedly would perceive as /t/) was created from the /ta/ sound by shortening the aspiration noise. For the continuum points in between the /ta/ endpoint and /da-ta/ midpoint, three stimuli were created by cutting +-20 ms out of the aspiration noise. For the beginning point /da/ (the most voiced sound of the continuum), another recording of the same speaker was used, in which the prevoicing was cut to 89 ms. In order to maintain the most natural prevoicing for the continuum points in between the /da/ beginning point and /da-ta/ midpoint, three stimuli were created by cutting out periods of +-20 ms of the beginning point's prevoicing, in which the natural lines of the prevoicing were retained. The end result was a continuum were VOT started at -89 ms and was increased in +- 20 ms steps to 80 ms.

3.3 Experiment

3.3.1 Production task

The experiment started with the production task to assure subjects were not aware what specific sounds were being tested, something that could possibly influence their production.

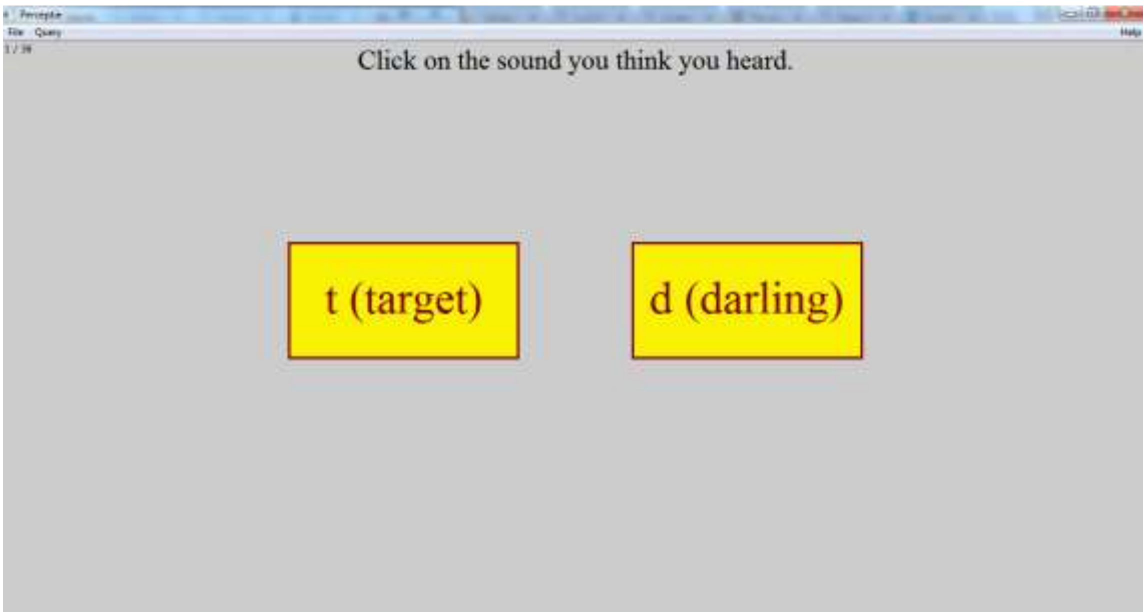
The test and filler items were presented in PowerPoint on the experimenter's laptop. Subjects were told to read the instructions, and to inform the researcher if anything was unclear or they had remaining questions. An introduction text was then presented on the screen, which explained the procedure: an English sentence would appear on the screen, and they had to read the whole sentence out loud, after which they had to wait until the next slide with the next sentence would present itself. The slides moved at a rate of three seconds and all test and filler items were presented twice. First subjects got two example sentences, which they had to read so they could get familiar with the rate at which the slides moved. After that the recorder was switched on and the actual task began. The production task took about three minutes to complete.

3.3.2 Categorization task

After the production task, the perception task immediately followed. In a forced-choice categorization task designed and run in Praat, the /da/- /ta/ continuum was presented in a random order to each subject. In the Praat script the number of replications of stimulus was set on 4, so all stimuli were presented to the subjects four times. The randomize option was set to <PermuteAll> so each subject got the same number of replications of each stimulus, but all in different random order.

Again an introduction text explained the procedure. Subjects were instructed to listen to a syllable and decide whether they heard a /d/ (as in darling) or /t/ (as in target) in initial position. They had to guess in case of uncertainty. Then they had to click on the corresponding button, /d/ (darling) or /t/ (target) on the screen, after which the next sound would be presented. Subjects listened to the sounds through Samsung headphones. The task took about one minute to complete.

Figure 2: Screenshot of categorization task in Praat



3.4 Analysis

3.4.1 Production

After all data was gathered, the data had to be analyzed in order to compare production and perception accuracy. Subjects' production of /d/ and /t/ was analyzed with the Praat software by examining the sound waves and spectrograms in their recordings. Although background noise was eliminated as much as possible, for several subjects some background noise remained, which was visible in the sound waves of the recordings. Closure time and VOT tiers were created using the text grid annotation tool and the sound recording and text grid were viewed and edited together. The range of the spectrogram was set to 0 - 10.000 Hz under spectrum settings. Pulses and intensity options were switched on, to facilitate the analyzing process. The beginning points and endpoints of closure time and VOT were marked by setting boundaries at respective points in the tiers. Closure time and VOT were then measured as the time in ms between the boundaries.

The closure time was determined by measuring the time between the end of the friction noise of the preceding /s/ and the stop release. The boundary at the end of the friction noise was set at the place where the spectrogram became considerately less dark. The boundary of time of stop release was set at the place where the spectrogram became considerately darker, which usually was also clearly visible in the waveform as sudden irregular waves. VOT was determined by measuring the time between the stop release and the start of voicing. The time of stop release was marked in the VOT tier at the same place as in the closure time tier. The start of voicing was determined by looking at the first periodic wave in the waveform and also at where the blue vertical stripe in the spectrogram began, which both indicated the start of vibration of the vocal folds. The blue stripe did not always begin exactly where the first regular wave began, but it was used as a helping tool. The point of release was assigned zero-time, so when the voice onset preceded the stop release (prevoicing), it meant it had a negative VOT and when the start of voicing followed the stop release, it had a positive VOT. For the /t/-words the voice onset was the start of the voicing of the /a/. This point was set at the first clear periodic wave after the stop release in the waveform, and the beginning of the blue vertical stripe in the spectrogram. For the /d/-words the start of voicing sometimes was the first clear periodic wave period before the stop release. Often there was no prevoicing present in the /d/-sounds, in which case the VOT was measured as in the /t/-words, so the length of the burst. In some cases, there was initial prevoicing, which meant that after the end of the friction noise of the preceding /s/ there was prevoicing of the /d/, but this did not continue until the stop release. In these cases, the initial voicing was measured, divided by the closure time, and multiplied by 100, giving a percentage of voicing instead of full voicing (see Appendix 2).

The measurements of closure time, VOT, duration of the burst and possible initial voicing were put in a table (see Appendix 2) for every /t/ and /d/ realization, which amounted to a total of 200 /t/-realizations and 200 /d/-realizations (10 realizations of /t/ and 10 realizations of /d/ per subject). After this, each subject's mean VOT for /t/ and /d/ was calculated (see table 1a). For /d/, negative values (so with prevoicing), were added up to positive values (with no prevoicing).

The subjects' VOT was compared to that of native speakers, and when VOT matched the VOT of native speakers, their production was considered 'native-like'. As a reference, the results of previous studies mentioned earlier in §2.2.1 were used to determine the ranges of native speakers' VOT for the realization of /d/ and /t/. The VOT ranges were determined to be 0 - 30 ms for /d/ (short lag) and 30 - 105 ms for /t/ (long lag).

3.4.2 Perception

The results of the categorization task were collected to a table and viewed in Praat. For each continuum point (e.g. (1) -87 ms, (2) -63 ms etc.) the amount of times the subjects chose /d/ was counted and converted in percentages. As the stimuli were repeated three times (so presented four times in total), the possible percentages were 0%, 25%, 75%, and 100%. Then the data was put in a table (see Appendix 3) and converted into graphs.

On the basis of the table and the graphs the perception boundary for each subject was determined, after which was determined whether each boundary was native-like or nonnative.

4 Results

4.1 Production

Table 1a shows the mean VOT of each subject for /t/ and /d/. The mean realization of /t/ varies from 17.7 ms to 84.4 ms. 11 subjects (55%) have a mean /t/ with a short lag value and 9 subjects (45%) with a long lag value. The mean realization of /d/ varies from -67.1 ms to 13.7 ms. 15 subjects (75%) have a mean /d/ with a voicing lead value, and only 5 (25%) with a short lag value. It should be noted that the separate realizations of both /d/ and /t/ within subjects varied substantially, especially in /d/-words, where subjects sometimes produced /d/ with prevoicing but other times without prevoicing. Also some values are very close to the boundaries set for this study. For example subject 1's mean of /t/ is almost 30 ms and subject 10's mean of /d/ is almost 0 ms.

Table 1a: Mean VOT (in ms) of each subject

Subject	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
/t/	29.5	42.2	52.4	25.4	72	27.7	20	17.7	41.2	26.9
/d/	8.1	-2.1	-52	6.8	-20	-8.9	13.7	-4.1	-46.9	-0.6
Subject	***11***	***12***	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***
/t/	28.9	25.8	38.2	24	35.1	21.5	84.4	34.8	41.4	26
/d/	12.8	-44.4	-24	-32	-4.6	6.2	-29.9	-67.1	-10.4	-27.2

For each subject it was determined whether their mean VOT was considered native-like (N) or nonnative (NN). For /d/ a mean between 0 and 30 ms (short lag) was considered native-like, and if it was under 0 ms (so with prevoicing) it was considered to be nonnative. For /t/ it meant that if the mean was between 0 and 30 ms (short lag values), it was considered to be nonnative, and if it was 30 ms or longer (long lag values), it was considered native-like. In table 1b this classification can be observed.

Table 1b: mean VOT (in ms) of each subject, with classification

Subject	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
/t/	29.5	42.2	52.4	25.4	72	27.7	20	17.7	41.2	26.9
	NN	N	N	NN	N	NN	NN	NN	N	NN
/d/	8.1	-2.1	-52	6.8	-20	-8.9	13.7	-4.1	-46.9	-0.6
	N	NN	NN	N	NN	NN	N	NN	NN	NN
Subject	***11***	***12***	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***
/t/	28.9	25.8	38.2	24	35.1	21.5	84.4	34.8	41.4	26
	NN	NN	N	NN	N	NN	N	N	N	NN
/d/	12.8	-44.4	-24	-32	-4.6	6.2	-29.9	-67.1	-10.4	-27.2
	N	NN	NN	NN	NN	N	NN	NN	NN	NN

4.2 Perception

Figure 3a, 3b and 3c show the number of times each subject chose /d/ in the categorization task, converted into percentages. Most subjects display a constant downward trend, which can be observed in figure 3a, a number of subjects also portray a downward trend but with an occasional rise (see figure 3b), and a few subjects exhibit an overall downward trend but do not identify the beginning point or endpoint as a /d/ or /t/ a 100% of the time (see figure 3c).

The /d/-/t/ boundary was determined for each individual subject based on figure 3a, 3b and 3c and was put in a table (see table 2a).

Figure 3a: percentage of /d/ chosen in categorization task. This figure shows the subjects who display a constant downward trend.

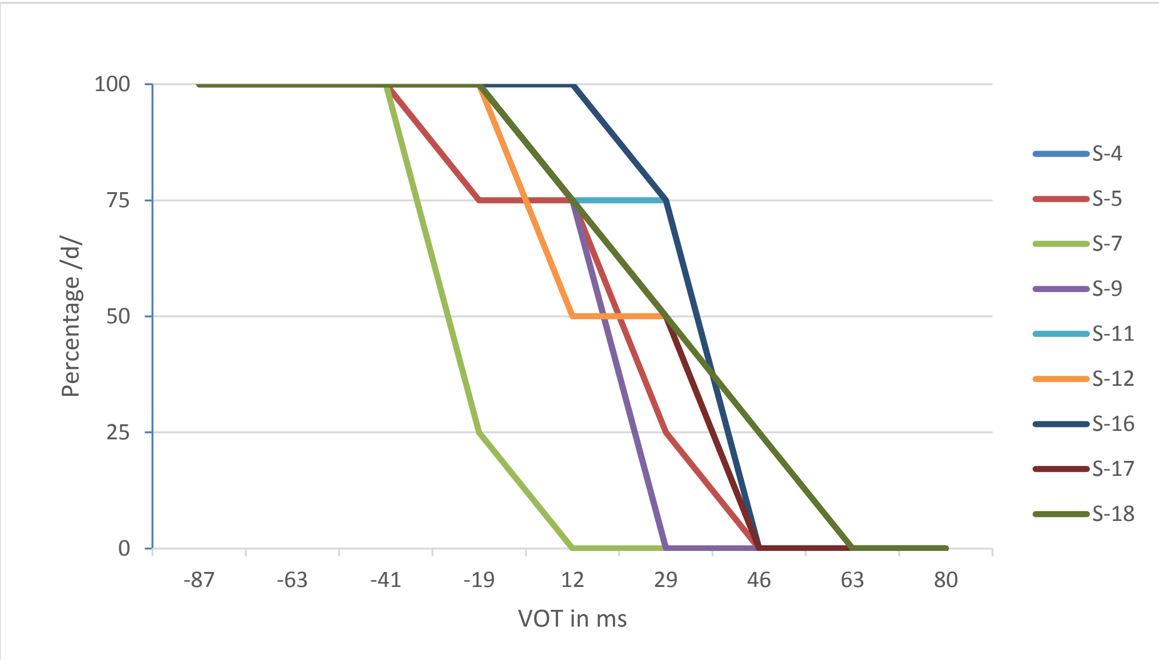


Figure 3b: percentage of /d/ chosen in categorization task. This figure shows the subjects who display a downward trend with an occasional rise.

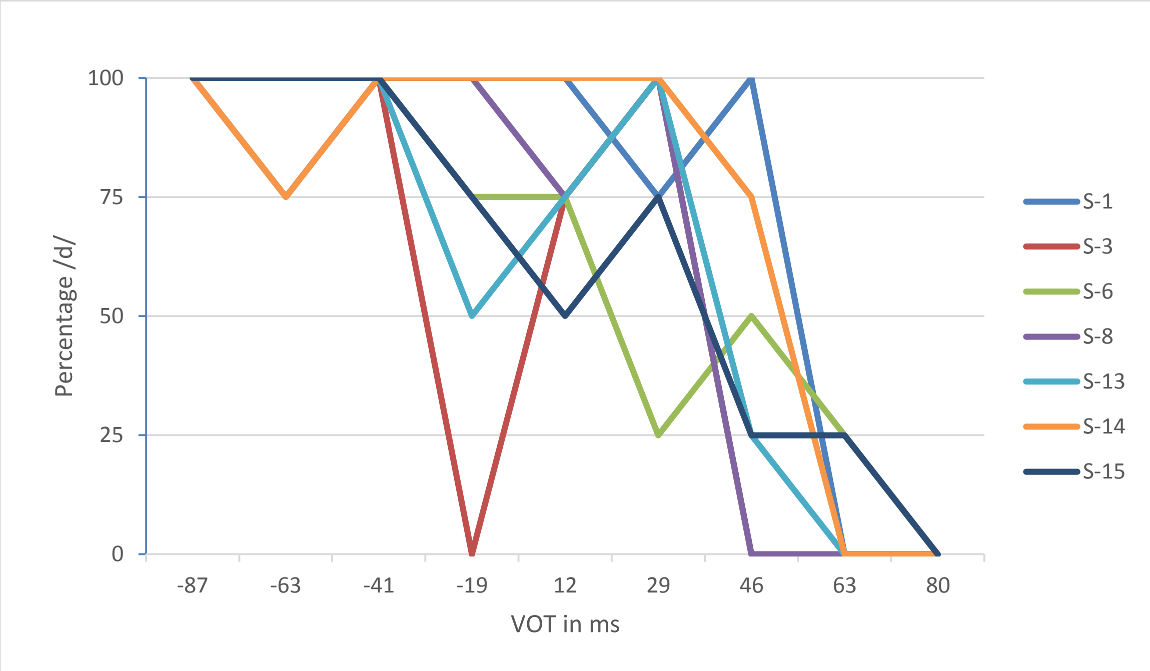
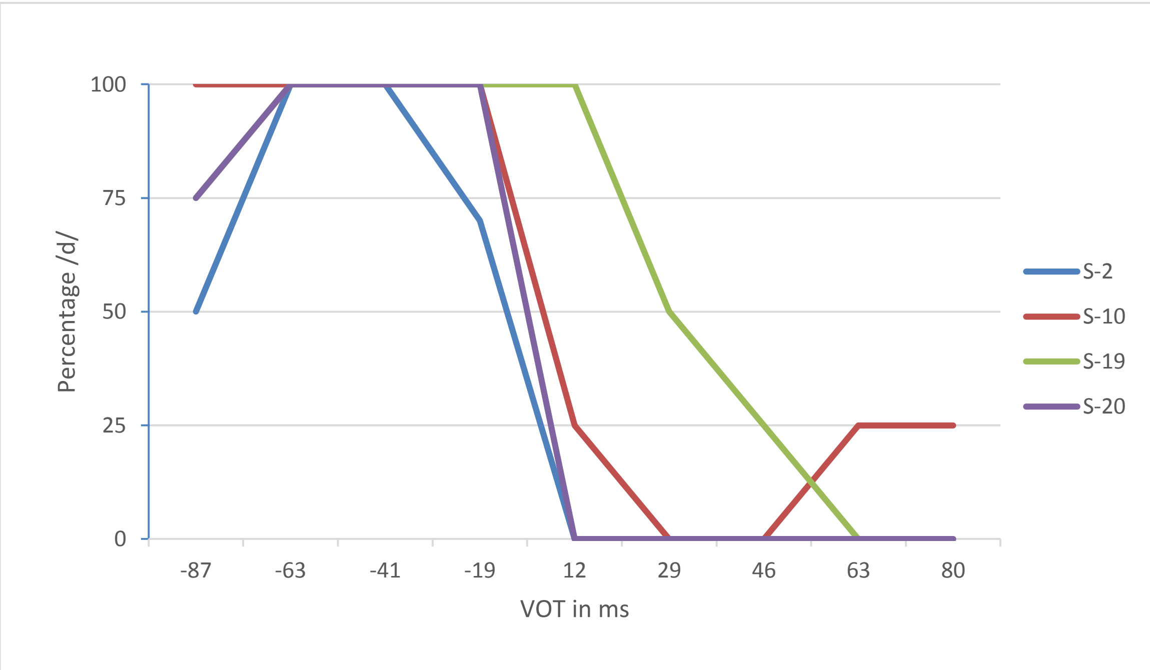


Figure 3c: percentage of /d/ chosen in categorization task. This figure shows the subjects who did not identify the starting point of the continuum a 100% of the time as a /d/ or the endpoint a 100% of the time as a /t/.



It is visible from figure 3a and 3b that the perception results of most subjects portray a constant downward trend, with sometimes an occasional rise of 25% or 50% in the middle of the continuum, which is what can be expected. There are some exceptions. As illustrated in figure 3c, subject 2, 19 and 20 identified the /d/ beginning point (-87 ms) not a 100% of the time as a /d/ but only in 50 or 75% of the cases. They did identify the next two points on the continuum a 100% as a /d/. A similar thing is true for subject 10 with the /t/ endpoint (80 ms) which was, together with the second to last stimulus (63 ms), identified as a /d/ 25% of the time, while 29 and 46 ms were identified as a /d/ 0% of the time. Even more surprising is subject 3, who shows a drop of 100% followed by a rise of 75% (see figure 3b). After identifying -87 ms and -63 ms a 100% of the times as /d/, subject 3 displays a steep drop to 0% at -41 ms, and then rises again to 75% at -19 ms.

Table 2a: boundary range (in ms) of each subject

Subject	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Boundary	46-63	-19-12	12-29	12-29	12-29	46-63	-41-19	29-46	12-29	-19-12
Subject	***11***	***12***	***13***	***4***	***15***	***16***	***17***	***18***	***19***	***20***
Boundary	-19-12	29-46	29-46	29-46	46-63	29-46	29-46	29-46	29-46	-19-12

For each subject it was determined whether their VOT perception boundary was considered native-like or nonnative. The boundary between /d/ and /t/ was set on 30 ms based on previous research, which meant that only when subjects had a boundary range of 29-46 ms their perception was considered to be native-like, because it meant that only after 29 ms they considered the stimuli to be a /t/. If subjects had any other boundary, their perception was considered to be nonnative. In table 2b, table 2a is extended with the native-like/nonnative classification.

Table 2b: boundary range (in ms) of each subject, with classification

Subject	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Boundary	46-63	-19-12	12-29	12-29	12-29	46-63	-41-19	29-46	12-29	-19-12
	NN	NN	NN	NN	NN	NN	NN	N	NN	NN
Subject	***11***	***12***	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***
Boundary	29-46	29-46	29-46	46-63	29-46	29-46	29-46	29-46	29-46	-19-12
	N	N	N	NN	N	N	N	N	N	NN

As shown in table 2b, 9 subjects (45%) have a native-like perception, and 11 subjects (55%) have not. Based on these results it would be expected that the 9 subjects that have a native-like perception, have a native-like production as well, and the 11 subjects that do not have a native-like perception, would certainly not have a native-like production.

4.3 Production and perception together

In table 3 the classification 'native-like' or 'nonnative' is presented for perception and production, together in one table (production split by /d/ and /t/).

Table 3: classification of production and perception for each subject

Subject	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Perception	NN	NN	NN	NN	NN	NN	NN	N	NN	NN
Production /t/	NN	N	N	NN	N	NN	NN	NN	N	NN
Production /d/	N	NN	NN	N	NN	NN	N	NN	NN	NN
Subject	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Perception	N	N	N	NN	N	N	N	N	N	NN
Production /t/	NN	NN	N	NN	N	NN	N	N	N	NN
Production /d/	N	NN	NN	NN	NN	N	NN	NN	NN	NN

It is illustrated in table 3 that production is not always native-like when perception is (subject 8, 12). But classification of /d/ and /t/ is different for many subjects. Sometimes the production of /t/ is native-like and /d/ is not (subject 13, 15, 17, 18 and 19) and sometimes the production of /d/ is native-like and /t/ is not (subject 11, 16).

More unexpected is the fact that in many cases production is native-like while perception is not. This is the case for subject 2, 3, 5 and 9 for /t/ and for subject 1, 4 and 7 for /d/. Subject 3, 4, 5 and 9 have a perception boundary of 12-29 ms. An issue arises because 29 ms is almost at the 30 ms boundary, so arguably these subjects do have a native-like boundary, however with the boundary set in this study they fall in the nonnative category. Production of /t/ and /d/ are never both native-like and 6 times (30%) both nonnative. Subject 6, 10, 14 and 20 have a nonnative perception boundary and a nonnative production mean for both /d/ and /t/.

5 Discussion

5.1 Implications of results

From table 3 it becomes clear that perception and production more often does not correlate than it does. First of all, the classification of /d/ and /t/ differs for many subjects. 14 subjects (70%) differ in whether their mean realization of /d/ or /t/ is native-like. /t/ is more often pronounced native-like than /d/ (45% against 25%). A possible explanation is that people are more aware of the fact that the /t/ in English is different from the Dutch /t/. Presence of aspiration is something that people are probably more aware of than absence of prevoicing. Flege and Eefting's (1987b) finding that Dutch subjects produce /t/ with longer VOT values in English then in Dutch, indicates that Dutch speakers are to some degree aware of acoustic differences between the English /t/ and the Dutch /t/. Another factor could be not wanting to be misunderstood. Although people are not likely to be consciously aware that the Dutch /t/ is pronounced similar to the English /d/, unconsciously they might know, or they might have experienced, that producing the English /t/ without aspiration can result in being misunderstood.

In the cases where subjects have a nonnative perception but a native-like production of either /d/ or /t/, it is not true that the phoneme which is native-like is almost classified as nonnative. As can be seen in table 1b and table 3, the subjects that have a native-like perception and only a native-like production of /t/, have mean realizations of /t/ that are at least 11.2 ms above the boundary (41.2 ms – 72 ms), and the subjects with a only a native-like production of /d/ have mean realizations of /d/ that are at least 6.8 ms above the boundary (6.8 ms – 13.7 ms). In the cases where subjects have a native-like perception but not a completely native-like production, the phoneme which is nonnative is also not almost native-like. Only one subject, with a nonnative production of /t/, has a VOT value that could be considered close to the boundary, namely 28.9 ms (subject 11, see table 1a). The others have a nonnative production of /d/ and in these cases the VOT values are not almost 0 ms, i.e. not almost native-like.

For the 6 subjects where the classification of /t/ and /d/ does correspond, both /t/ and /d/ are classified as nonnative. Only four times perception and production completely correspond, and in all cases, perception and production of both /d/ and /t/ are all classified as nonnative. There are no cases where a native-like perception corresponds with a native-like production.

From these results it evidently does not appear to be true that perception precedes production, because production is often native-like while perception is not, although never both /t/ and /d/, and production is never completely native-like when perception is.

This largely contradicts the findings of previous studies. It should be noted though that most studies investigated one of the aspects focused on in the present study, and have not focused on all aspects. The studies that examined the relationship between perception and production and found that perception preceded production, did not look at the /t/ and /d/ contrast, but most commonly at some sort of vowel contrast. It is possible that there are other things at play here.

Studies that have looked at the VOT of English and Dutch and Dutch L2 speakers of English, have found that Dutch speakers tend to produce the /t/ with short lag VOT values that are closer to their native language, instead of long lag values, as produced by native English speakers (see §2.2.1 and §2.2.2). In the current study only 11 subjects (55%) have a short lag VOT value for the mean /t/, which is only a slight majority. The remaining 45% have a native-like production of /t/, indicating that those subjects have established a phonetic category for the English /t/, but this cannot be the case because only 55.5% of those subjects have a native-like perception. A native-like production of /d/ even only co-occurs with a native-like perception for 2 subjects (10%).

Prevoicing in /d/-words was also less common than would be expected from previous research. First of all, almost all subjects, except subject 4, produced one or more /d/-words with initial prevoicing, instead of full prevoicing. In addition to this, no subject produced all /d/-words with (initial) prevoicing. With reference to Van Alphen and Smits (2004) (see §2.2.2), it could be expected that /d/ would not always be realized with voicing lead values. However, Van Alphen and Smits only reported an absence of prevoicing in 25% of the /d/ realizations. In the present study 25% of the mean /d/ VOT values are without prevoicing (see table 3: N-classifications of /d/). In total however, only 94 of 200 realizations, which is 47%, have prevoicing (24.5% is initial prevoicing). This means that in total 53% of the /d/-words were not produced with voicing lead values. This is a considerably higher percentage than was expected.

It was also expected that the subjects would identify /d/ as a /t/ in the categorization task, because the English /d/ is produced with the same VOT values as the Dutch /t/ (short lag). If this were true, the subjects should have identified the stimuli with a VOT of 12 ms and 29 ms as a /t/, instead of a /d/. This is not the case, most subjects identified both those VOT values as a /d/, some identified 29 ms as a /t/ but 12 ms as a /d/, and only a few identified both as a /t/. This might be explained in the context of the stimulus continuum. It seemed that the 63 and 80 ms stimuli were quite distinguishable as a /t/ and more obvious than the -87 and -63 as a /d/. Because the stimuli were presented in a random order, the contrast between the strong /t/'s and the rest of the stimuli seemed most noticeable, and it is possible that subjects tended to choose /d/ more often whenever it was not a strong /t/. If the /ta/

endpoints of the continuum were more subtle, i.e. with less aspiration, perhaps /d/ would have been identified as a /t/ more often.

In general, the perception precedes perception theory could not be corroborated. To answer the research question: the perception accuracy of the English /d/ and /t/ in word initial position by Dutch L2 learners does not positively correlate with the production accuracy of the English /d/ and /t/. No clear evidence that production precedes perception has been found either, because in the cases where perception is nonnative, production of /d/ and /t/ is never both native-like.

5.2 Problems during analysis

When analyzing the production data, the following problems arose (some once or twice, others frequently). Firstly, the burst of some subjects was hardly visible in the waveform, and it was hard to determine where the burst began and ended. The reason that the burst was so weak could be because the subject did not speak loud enough.

Another problem was that sometimes with /d/-words it was hard to determine where the voicing of the preceding /s/ ended and the (possible) prevoicing of the /d/ began.

A strange thing that could be observed was that 9 of the 200 realizations of /t/ displayed some sort of voicing, which could be seen in the waveform. This occurred with 5 different subjects, and 5 times with subject 17. The reason for this could have been background noise, though this does not explain why not all words displayed voicing. Background noise was worst for subject 17 however, so this could be an explanation. Although background noise was shut out as much as possible while conducting the experiments, outside noise or unexpected sounds were not completely absent and for some recordings the waveform did show some background noise. Although this did not prevent me from analyzing them, it did make it more difficult.

Lastly some subjects displayed great VOT differences between the words. For example subject 3 had a VOT of 13 ms for one realization of the word /taco/ and a VOT of 91 ms for a realization of the word /tall/. Such large ranges were not common however. VOT varied a lot as well between the same words but a different repetition.

5.3 Limitations of study and recommendations for future research

For the production task the choice was made to let the subjects read the test and filler items as part of a sentence. The word preceding the target words ended in /s/, but this was produced with voicing. Sometimes this voicing merged with the voicing of the /d/-words. VOT values might depend on whether the test items are produced as words in isolation or in natural speech. Replicating this study and changing the way the test items are elicited is necessary to verify the chosen method.

Multiple people reported after the categorization task that once or twice they clicked the wrong button or that they did not hear a difference between the sounds in the beginning of the task. For example, subject 10 clicked the /d/-button for stimuli 8 and 9 (63 and 80 ms), which were the second and third sound she heard, and after that she consistently chose /t/ for these sounds. In other words, only after subjects heard a few reference sounds, they could hear a contrast between different sounds. Presumably this interfered with the results and some results may not reflect the subjects' true perception boundary. The subjects did not practice the categorization task, perhaps a practice session is necessary to familiarize participants with the task, to assure they hear a difference between the sounds. What is more, the responses seemed to partly depend on the nature of the previously heard sound (e.g. a /d/ was more often identified as /d/ when it followed a strong /t/, than when it followed another /d/ (unless it was exactly the same sound). If presented in another sequence, the responses might have been different.

Moreover, the continuum had steps of approximately 20 ms, which might have been too big. The perception results show that many subjects had a sudden drop in identifying a /d/, e.g. going from 75% straight to 0%, instead of identifying the stimuli gradually from 100% to 0%. This indicates that the contrasts between the stimuli where perception drastically changed were too strong, i.e. the steps between the stimuli were too big, in which case the continuum needed more steps.

Another limitation of this study is the fact that it is disputable what exactly is 'native-like', because 'the' native speaker does not exist. Therefore the results of this study rely on the definition of what is native-like, in this case, the boundaries set for production and perception based on previous research. This is problematic, especially combined with a stimuli continuum with steps of approximately 20 ms in VOT. Point 6 of the continuum had a VOT of 29 ms and the boundary was set on 30 ms. If the VOT of point 6 was slightly longer, the perception boundary would have shifted to the left. To avoid this problem the stimuli continuum should be improved with smaller steps between stimuli, i.e. more stimuli, so the perception can be observed more precisely. Additionally, while voicing lead quite clearly only occupies negative values, there is no consensus on where the boundary of short lag and long lag lies exactly. If the boundary was set at a VOT of 46 ms or longer, the perception classification and therefore the comparison with production would have been considerably different (though perception and production would not have been more correlated). Lastly, due to time restrictions and limited resources, there was no control group of English native speakers. A control group would have made a direct comparison of the perception boundary of the Dutch L2 speakers to that of English native speakers possible.

Besides the fact that there was no control group, there were also only 20 Dutch subjects, again due to time restrictions. The present findings are limited by this small sample size of participants because it did not allow for a statistical analysis and the results in this study should therefore be interpreted with caution.

Further research is needed to address the limitations of this study and in addition future studies could also explore the effects of language experience and age on the relationship between production and perception of /d/ and /t/ or look at other consonant contrasts.

6 Conclusion

The aim of this study was to contribute evidence or contra-evidence to the theory that perception precedes production, and to provide more insight into the issue concerning the perception and production of the English consonant contrast /t/-/d/, which differs for native speakers of Dutch and native speakers of English in VOT. It was investigated whether the perception accuracy of the English /d/ and /t/ in word initial position by Dutch L2 learners positively correlates with the production accuracy of their English /d/ and /t/. Based on previous research it was expected that this would be the case, but the results found in this study suggest that in the case of /d/ and /t/ in initial position the perception-precedes-production theory does not hold true.

The experiment consisted of a production task where subjects were asked to read sentences with words with /d/ and /t/ in initial position, and a categorization task, where subjects had to identify syllables with /d/ and /t/ in initial position. The results from the experiment revealed that perception and production did not correlate. Interestingly, many cases of a native-like production with a nonnative perception were found, but production accuracy of /t/ and /d/ did also not correlate, so there was no valid evidence that production precedes perception.

However, there are some limitations to the study and a replication of this study should be performed to make sure these results are reliable. The methodology can be improved in several ways, such as: adding more steps to the continuum, making the last stimulus of the continuum less aspirated and therefore more ambiguous, and adding a control group. In addition, it would be interesting to assess the effects of the way the test items are elicited and the order in which the stimuli in the categorization task are presented.

While neither evidence has been found for a perception-precedes-production theory, nor a production-precedes-perception theory, the findings presented in this paper hopefully provide more insight into the issue of English L2 perception and production of /d/ and /t/ in initial position by native speakers of Dutch.

References

van Alphen, P. M., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of Phonetics, 32*(4), 455-491.

Bolanos, L. (2013). Perception and production in nonnative speech: Russian palatalization. *The Journal of the Acoustical Society of* America, *133*(5), 3610-3610.

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in second language acquisition*, *19*(4), 447-465.

Brière, E. (1966). An Investigation of Phonological Interference. *Language, 42*(4), 769-796.

Docherty, G. J. (1992). *The timing of voicing in English obstruents*. Berlin: Foris Publication.

Doel, W.Z. van der (2006). *How Friendly are the Natives? An Evaluation of Native-speaker Judgements of Foreign-accented British and American English*. (Doctoral dissertation, Utrecht University). Retrieved from: Utrecht University Repository.

van Dommelen, W. A. (1983). Some observations on assimilation of voicing in German and Dutch. In Marcel von den Broecke (Ed). et al. *Studies for Antonie Cohen: Sound structures*, (pp. 47-56). Dordrecht: Foris publications Holland.

Flege, J.E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America, 93*, 1589-1608.

Flege, J. E. (1995). Second language speech learning: theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 233-272). Timonium, MD: York Press.

Flege, J. E. (1999). The relation between L2 production and perception. In Ohala, J. J (Ed). et al., *Proceedings of the XIVth International Congress of Phonetics Sciences* (pp. 1273-1276). Berkeley, CA: Linguistics Department, University of California.

Flege, J.E., Eefting, W. (1987a). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics, 15*, 67-83.

Flege, J. E., & Eefting, W. (1987b). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication, 6*(3), 185-202.

Flege, J.E., Wang, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English / t/- /d/ contrast. *Journal of Phonetics, 17*, 299-315.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America, 97*(5), 3125-3134.

Flege, J.E., Bohn, O.S., & Jang, S. (1997). Effects of experience on nonnative speakers' production and perception of English vowels. *Journal of phonetics, 25*(4), 437–470.

Flege, J.E., Mackay, I.R.A., & Meador, D. (1999) Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*(5), 2973-2987.

Goto, H. 1971. Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia, 9*, 317-323.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60-99.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research, 18*(4), 686-706.

Lisker, L., & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384-422.

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In Kjell Elenius & Peter Branderud (eds.), *Proceedings of the 13th International Congress of Phonetic Sciences 4* (pp.92-99. Stockholm: KTH / Stockholm University.

Lopez, K. Yu, Y.H., & Bell-Berti, F. (2006). Perception and production of English vowels by native speakers of Mandarin. *The Journal of the Acoustical Society of America*, *119*(5), 3423-3423.

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in second language acquisition*, *21*(1), 81-108.

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

Neufeld, G. G. (1988). Phonological asymmetry in second language learning and performance, *Language Learning, 38*, 531-559.

Penfield, W., & Roberts, L. (1959). *Speech and Brain Mechanisms*. Princeton, NJ: Princeton University Press.

Rauber, A.S., Escudero, P., Bion R A. H., & Baptista, O. (2005). The Interrelation between the Perception and Production of English Vowels by Native Speakers of Brazilian Portuguese. In *Ninth European Conference on Speech Communication and Technology*. Lisbon: International Speech Communication Association.

Schairer, K.E. (1992). Native Speaker Reaction to Nonnative Speech. *The Modern Language Journal*, *76*(3), 309-319.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243-261.

Appendices

Appendix 1 List of words for production

Test items:

/T/-WORDS	/D/-WORDS
TARGET	DARLING
TACO	DART
TALL	DARTS
TALK	DARK
TAR	DAFT

Filler items:

MINIMAL PAIRS	RHYME WORDS	OTHER WORDS
BALL	STAR	YARD
CALL	START	HOME
FALL		RUFFLE
WALK		RUSH
CART		HONEY
PART		FEAST
BAR		FLAME
FAR		YELL
		HAPPY
		STARTLE

The words were presented in the following sentences: The words is … / What does … mean?

Appendix 2 Table of production analysis

speaker	word	rep	context	segment	closure (ms)	burst (ms)	VOT (ms)	Initial voicing (ms)	percentage
1	talk	1	is	t	71	24	24		
1	dart	1	/	d	/	19	19		
1	tall	1	does	t	51	32	32		
1	daft	1	does	d	69	22	22		
1	target	1	is	t	40	33	33		
1	tar	1	is	t	49	33	33		
1	darling	1	is	d	68	9	-49	49	72%
1	taco	1	is	t	32	39	39		
1	darts	1	is	d	36	39	39		
1	dark	1	does	d	47	17	17		
1	talk	2	is	t	49	30	30		
1	dart	2	does	d	40	19	19		
1	tall	2	does	t	54	26	26		
1	daft	2	/	d	/	18	18		
1	target	2	is	t	54	27	27		
1	tar	2	is	t	71	25	25		
1	darling	2	is	d	13	35	35		
1	taco	2	does	t	31	26	26		
1	darts	2	does	d	19	24	24		
1	dark	2	is	d	43	15	15		
2	talk	1	is	t	41	29	29		
2	dart	1	does	d	77	15	-77		
2	tall	1	does	t	59	89	89		
2	daft	1	does	d	85	14	14		
2	target	1	is	t	46	38	38		
2	tar	1	is	t	189	28	28		
2	darling	1	is	d	49	22	22		
2	taco	1	is	t	42	29	29		
2	darts	1	is	d	49	17	17		
2	dark	1	does	d	62	19	-33	33	53%
2	talk	2	is	t	46	40	40		
2	dart	2	does	d	36	16	-36		
2	tall	2	does	t	48	63	63		
2	daft	2	does	d	263	16	16		
2	target	2	is	t	41	30	30		
2	tar	2	is	t	42	33	33		
2	darling	2	is	d	42	22	22		
2	taco	2	does	t	52	49	49		
2	darts	2	does	d	29	19	19		
2	dark	2	is	d	40	15	15		
3	talk	1	is	t	69	31	31		
3	dart	1	does	d	124	11	-83	83	67%

3	tall	1	does	t	80	91	91			
3	daft	1	does	d	143	16	-95	95		66%
3	target	1	is	t	110	56	56			
3	tar	1	is	t	112	65	65			
3	darling	1	is	d	87	14	-44	44		51%
3	taco	1	is	t	93	13	13			
3	darts	1	is	d	81	12	-63	63		78%
3	dark	1	does	d	79	14	-61	61		77%
3	talk	2	is	t	88	40	40			
3	dart	2	does	d	89	14	14			
3	tall	2	does	t	62	73	73			
3	daft	2	does	d	74	14	-57	57		77%
3	target	2	is	t	42	61	61			
3	tar	2	is	t	64	48	48			
3	darling	2	is	d	61	21	-45	45		74%
3	taco	2	does	t	69	46	46			
3	darts	2	does	d	90	13	-66	66		73%
3	dark	2	is	d	80	8	8			
4	talk	1	is	t	156	18	18			
4	dart	1	does	d	61	11	11			
4	tall	1	does	t	51	55	55			
4	daft	1	does	d	81	16	16			
4	target	1	is	t	82	21	21			
4	tar	1	is	t	125	14	14			
4	darling	1	is	d	115	12	12			
4	taco	1	is	t	77	21	21			
4	darts	1	is	d	81	14	14			
4	dark	1	does	d	130	15	15			
4	talk	2	is	t	143	26	26			
4	dart	2	does	d	119	13	13			
4	tall	2	does	t	101	40	40			
4	daft	2	does	d	115	17	-65			
4	target	2	is	t	92	15	15			
4	tar	2	is	t	101	21	21			
4	darling	2	is	d	87	18	18			
4	taco	2	does	t	83	23	23			
4	darts	2	does	d	84	14	14			
4	dark	2	is	d	57	20	20			
5	talk	1	is	t	74	61	61			
5	dart	1	does	d	97	22	22			
5	tall	1	does	t	62	79	79			
5	daft	1	does	d	63	26	26			
5	target	1	is	t	68	62	62			
5	tar	1	is	t	67	64	64			
5	darling	1	is	d	71	13	13			
5	taco	1	is	t	83	102	102			

5	darts	1	is	d	67	20	-60	60	90%
5	dark	1	does	d	81	22	22		
5	talk	2	is	t	40	80	80		
5	dart	2	does	d	94	28	-79	79	84%
5	tall	2	does	t	63	116	116		
5	daft	2	does	d	81	19	19		
5	target	2	is	t	27	81	81		
5	tar	2	is	t	107	66	66		
5	darling	2	is	d	69	22	-69		
5	taco	2	does	t	83	70	70		
5	darts	2	does	d	74	24	-31	31	42%
5	dark	2	is	d	63	15	-63		
6	talk	1	is	t	33	30	30		
6	dart	1	does	d	39	21	21		
6	tall	1	does	t	94	31	31		
6	daft	1	does	d	63	13	-63		
6	target	1	is	t	43	19	19		
6	tar	1	is	t	80	21	21		
6	darling	1	is	d	64	14	-26	26	41%
6	taco	1	is	t	48	26	26		
6	darts	1	is	d	22	14	14		
6	dark	1	does	d	57	22	-14	14	25%
6	talk	2	is	t	24	31	31		
6	dart	2	does	d	69	18	18		
6	tall	2	does	t	76	28	28		
6	daft	2	does	d	84	18	-84		
6	target	2	is	t	55	39	39		
6	tar	2	is	t	20	20	20		
6	darling	2	is	d	51	12	12		
6	taco	2	does	t	23	32	32		
6	darts	2	does	d	53	19	19		
6	dark	2	is	d	16	14	14		
7	talk	1	is	t	40	24	24		
7	dart	1	does	d	67	19	19		
7	tall	1	does	t	42	25	25		
7	daft	1	does	d	68	11	11		
7	target	1	is	t	60	20	20		
7	tar	1	is	t	64	14	14		
7	darling	1	is	d	62	17	17		
7	taco	1	is	t	55	12	12		
7	darts	1	is	d	72	11	11		
7	dark	1	does	d	35	16	16		
7	talk	2	is	t	60	14	14		
7	dart	2	does	d	57	14	14		
7	tall	2	does	t	75	27	27		
7	daft	2	does	d	83	12	12		

7	target	2	is	t	56	25	25		
7	tar	2	is	t	56	22	22		
7	darling	2	is	d	52	8	8		
7	taco	2	does	t	47	17	17		
7	darts	2	does	d	77	13	13		
7	dark	2	is	d	74	14	14		
8	talk	1	is	t	84	27	27		
8	dart	1	does	d	81	13	13		
8	tall	1	does	t	61	15	15		
8	daft	1	does	d	119	16	16		
8	target	1	is	t	54	16	16		
8	tar	1	is	t	102	17	17		
8	darling	1	is	d	88	17	17		
8	taco	1	is	t	66	11	11		
8	darts	1	is	d	134	17	-35	35	26%
8	dark	1	does	d	102	16	16		
8	talk	2	is	t	66	17	17		
8	dart	2	does	d	48	13	13		
8	tall	2	does	t	88	14	14		
8	daft	2	does	d	62	15	15		
8	target	2	is	t	67	18	18		
8	tar	2	is	t	91	12	12		
8	darling	2	is	d	101	13	13		
8	taco	2	does	t	59	30	30		
8	darts	2	does	d	93	14	-55	55	59%
8	dark	2	is	d	54	26	-54		
9	talk	1	is	t	75	33	33		
9	dart	1	does	d	56	41	-56		
9	tall	1	does	t	117	47	47		
9	daft	1	does	d	166	26	-60	60	36%
9	target	1	is	t	64	37	37		
9	tar	1	is	t	60	40	40		
9	darling	1	is	d	89	34	34		
9	taco	1	is	t	62	35	35		
9	darts	1	is	d	63	21	-63		
9	dark	1	does	d	88	16	-88		
9	talk	2	is	t	75	38	38		
9	dart	2	does	d	82	13	-25		
9	tall	2	does	t	116	54	54		
9	daft	2	does	d	68	18	-68		
9	target	2	is	t	63	50	50		
9	tar	2	is	t	97	44	44		
9	darling	2	is	d	64	25	-27		
9	taco	2	does	t	58	34	34		
9	darts	2	does	d	70	14	-70		
9	dark	2	is	d	46	10	-46		

10	talk	1	is	t	56	30	30			
10	dart	1	does	d	68	14	-68			
10	tall	1	does	t	71	30	30			
10	daft	1	does	d	252	17	17			
10	target	1	is	t	44	32	32			
10	tar	1	is	t	76	22	22			
10	darling	1	is	d	67	23	23			
10	taco	1	is	t	78	19	19			
10	darts	1	is	d	58	20	20			
10	dark	1	does	d	93	32	-80	80		86%
10	talk	2	is	t	54	23	23			
10	dart	2	does	d	77	19	19			
10	tall	2	does	t	49	35	35			
10	daft	2	does	d	31	21	21			
10	target	2	is	t	59	25	25			
10	tar	2	is	t	50	37	37			
10	darling	2	is	d	69	13	13			
10	taco	2	does	t	75	16	16			
10	darts	2	does	d	42	19	19			
10	dark	2	is	d	63	10	10			
11	talk	1	is	t	58	35	35			
11	dart	1	does	d	39	27	27			
11	tall	1	does	t	79	28	28			
11	daft	1	does	d	130	28	28			
11	target	1	is	t	73	23	23			
11	tar	1	is	t	63	32	32			
11	darling	1	is	d	71	26	26			
11	taco	1	is	t	145	30	30			
11	darts	1	is	d	64	23	23			
11	dark	1	does	d	-76	24	24			
11	talk	2	is	t	63	28	28			
11	dart	2	does	d	59	23	23			
11	tall	2	does	t	64	29	29			
11	daft	2	does	d	59	27	27			
11	target	2	is	t	54	31	31			
11	tar	2	is	t	64	28	28			
11	darling	2	is	d	55	22	-24	24		44%
11	taco	2	does	t	114	25	25			
11	darts	2	does	d	69	19	-47	47		68%
11	dark	2	is	d	45	21	21			
12	talk	1	is	t	36	30	30			
12	dart	1	does	d	29	21	21			
12	tall	1	does	t	23	32	32			
12	daft	1	does	d	67	26	-67			
12	target	1	is	t	28	27	27			
12	tar	1	is	t	82	25	25			

12	darling	1	is	d	101	18	-101			
12	taco	1	is	t	65	23	23			
12	darts	1	is	d	-69	25	25			
12	dark	1	does	d	113	30	-113			
12	talk	2	is	t	78	20	20			
12	dart	2	does	d	97	15	-56	56		58%
12	tall	2	does	t	78	19	19			
12	daft	2	does	d	79	25	25			
12	target	2	is	t	57	23	23			
12	tar	2	is	t	112	23	23			
12	darling	2	is	d	105	14	-51			
12	taco	2	does	t	36	37	37			
12	darts	2	does	d	99	13	-37	37		37%
12	dark	2	is	d	69	16	-69			
13	talk	1	is	t	79	34	34			
13	dart	1	does	d	86	30	30			
13	tall	1	does	t	98	45	45			
13	daft	1	does	d	85	25	25			
13	target	1	is	t	81	38	38			
13	tar	1	is	t	98	27	27			
13	darling	1	is	d	55	19	-29	29		53%
13	taco	1	is	t	67	34	34			
13	darts	1	is	d	66	18	18			
13	dark	1	does	d	65	19	-65			
13	talk	2	is	t	81	35	35			
13	dart	2	does	d	59	10	-59			
13	tall	2	does	t	62	39	39			
13	daft	2	does	d	65	21	-65			
13	target	2	is	t	66	40	40			
13	tar	2	is	t	65	45	45			
13	darling	2	is	d	54	17	-54			
13	taco	2	does	t	71	45	45			
13	darts	2	does	d	58	19	-58			
13	dark	2	is	d	25	17	17			
14	talk	1	is	t	78	35	35			
14	dart	1	does	d	58	23	-58			
14	tall	1	does	t	51	26	26			
14	daft	1	does	d	63	28	28			
14	target	1	is	t	64	22	22			
14	tar	1	is	t	52	20	20			
14	darling	1	is	d	76	21	-76			
14	taco	1	is	t	57	37	37			
14	darts	1	is	d	71	25	-71			
14	dark	1	does	d	60	16	-15			
14	talk	2	is	t	62	20	20			
14	dart	2	does	d	109	29	29			

14	tall	2	does	t	101	18	18		
14	daft	2	does	d	95	20	-81	81	85%
14	target	2	is	t	78	20	20		
14	tar	2	is	t	81	22	22		
14	darling	2	is	d	59	19	19		
14	taco	2	does	t	69	20	20		
14	darts	2	does	d	73	17	-73		
14	dark	2	is	d	81	30	-22		
15	talk	1	is	t	48	27	27		
15	dart	1	does	d	70	19	19		
15	tall	1	does	t	96	39	39		
15	daft	1	does	d	59	16	-22	22	37%
15	target	1	is	t	79	18	18		
15	tar	1	is	t	67	25	25		
15	darling	1	is	d	39	15	15		
15	taco	1	is	t	57	54	54		
15	darts	1	is	d	61	14	14	19	31%
15	dark	1	does	d	51	17	-32	32	63%
15	talk	2	is	t	54	57	57		
15	dart	2	does	d	62	18	-47	47	76%
15	tall	2	does	t	41	40	40		
15	daft	2	does	d	64	16	-40	40	63%
15	target	2	is	t	55	37	37		
15	tar	2	is	t	71	22	22		
15	darling	2	is	d	52	15	15		
15	taco	2	does	t	44	32	32		
15	darts	2	does	d	59	17	17		
15	dark	2	is	d	58	15	15		
16	talk	1	is	t	97	25	25		
16	dart	1	does	d	77	14	14		
16	tall	1	does	t	140	22	22		
16	daft	1	does	d	111	16	-44	44	40%
16	target	1	is	t	86	17	17		
16	tar	1	is	t	80	25	25		
16	darling	1	is	d	91	23	-35	35	38%
16	taco	1	is	t	75	22	22		
16	darts	1	is	d	74	17	17		
16	dark	1	does	d	65	16	16		
16	talk	2	is	t	101	22	22		
16	dart	2	does	d	79	18	18		
16	tall	2	does	t	114	27	27		
16	daft	2	does	d	85	16	16		
16	target	2	is	t	86	16	16		
16	tar	2	is	t	111	24	24		
16	darling	2	is	d	78	16	16		
16	taco	2	does	t	141	15	15		

16	darts	2	does	d	88	15	15		
16	dark	2	is	d	87	29	29		
17	talk	1	is	t	84	96	96		
17	dart	1	does	d	122	13	-122		
17	tall	1	does	t	90	123	123		
17	daft	1	does	d	125	15	15		
17	target	1	is	t	107	80	80		
17	tar	1	is	t	161	77	77		
17	darling	1	is	d	234	17	-59	59	25%
17	taco	1	is	t	89	63	63		
17	darts	1	is	d	91	16	-40	40	44%
17	dark	1	does	d	109	14	14		
17	talk	2	is	t	67	46	46		
17	dart	2	does	d	101	21	21		
17	tall	2	does	t	85	143	143		
17	daft	2	does	d	89	16	16		
17	target	2	is	t	63	76	76		
17	tar	2	is	t	75	76	76		
17	darling	2	is	d	141	24	-26	26	18%
17	taco	2	does	t	69	64	64		
17	darts	2	does	d	95	14	14		
17	dark	2	is	d	150	16	-132	132	88%
18	talk	1	is	t	56	36	36		
18	dart	1	does	d	81	19	-81		
18	tall	1	does	t	89	41	41		
18	daft	1	does	d	101	18	-93	93	92%
18	target	1	is	t	53	31	31		
18	tar	1	is	t	76	30	30		
18	darling	1	is	d	80	21	-80		
18	taco	1	is	t	59	27	27		
18	darts	1	is	d	84	13	-84		
18	dark	1	does	d	95	9	-80	80	84%
18	talk	2	is	t	74	28	28		
18	dart	2	does	d	80	16	-58	58	73%
18	tall	2	does	t	113	48	48		
18	daft	2	does	d	107	11	11	93	87%
18	target	2	is	t	63	38	38		
18	tar	2	is	t	85	30	30		
18	darling	2	is	d	53	22	-46	46	87%
18	taco	2	does	t	76	39	39		
18	darts	2	does	d	100	15	-78	78	78%
18	dark	2	is	d	82	17	-82		
19	talk	1	is	t	211	33	33		
19	dart	1		d	/	15	-112		
19	tall	1	does	t	145	52	52		
19	daft	1	does	d	124	15	-104	104	84%

19	target	1	is	t	58	30	30			
19	tar	1	is	t	152	23	23			
19	darling	1	is	d	78	2	-78			
19	taco	1	is	t	164	60	60			
19	darts	1	is	d	68	8	-59	59		87%
19	dark	1	does	d	192	5	-101			
19	talk	2	is	t	244	52	52			
19	dart	2	does	d	135	13	-92			
19	tall	2	does	t	110	63	63			
19	daft	2	does	d	167	19	-124	124		74%
19	target	2	is	t	141	39	39			
19	tar	2	is	t	128	23	23			
19	darling	2	is	d	95	4	-95			
19	taco	2	does	t	123	39	39			
19	darts	2	does	d	211	16	-130			
19	dark	2	is	d	109	8	-109			
20	talk	1	is	t	129	22	22			
20	dart	1	does	d	85	27	-60	60		71%
20	tall	1	does	t	121	43	43			
20	daft	1	does	d	116	12	-67			
20	target	1	is	t	136	22	22			
20	tar	1	is	t	133	17	17			
20	darling	1	is	d	107	17	17			
20	taco	1	is	t	122	31	31			
20	darts	1	is	d	107	17	17			
20	dark	1	does	d	66	25	-39	39		59%
20	talk	2	is	t	139	20	20			
20	dart	2	does	d	121	25	25			
20	tall	2	does	t	173	25	25			
20	daft	2	does	d	88	22	-39	39		44%
20	target	2	is	t	82	31	31			
20	tar	2	is	t	133	27	27			
20	darling	2	is	d	100	20	-100			
20	taco	2	does	t	89	22	22			
20	darts	2	does	d	142	23	-42	42		30%
20	dark	2	is	d	95	16	16			

Appendix 3 Table of categorization task answers in percentages

	-87 ms	-63 ms	-41 ms	-19 ms	12 ms	29 ms	46 ms	63 ms	80 ms
S-1	100%	100%	100%	100%	100%	75%	100%	0%	0%
S-2	50%	100%	100%	75%	0%	0%	0%	0%	0%
S-3	100%	100%	0%	75%	100%	0%	0%	0%	0%
S-4	100%	100%	100%	100%	75%	0%	0%	0%	0%
S-5	100%	100%	100%	75%	75%	25%	0%	0%	0%
S-6	100%	100%	100%	75%	75%	25%	50%	25%	0%
S-7	100%	100%	100%	25%	0%	0%	0%	0%	0%
S-8	100%	75%	100%	100%	75%	100%	0%	0%	0%
S-9	100%	100%	100%	100%	75%	0%	0%	0%	0%
S-10	100%	100%	100%	100%	25%	0%	0%	25%	25%
S-11	100%	100%	100%	100%	75%	75%	0%	0%	0%
S-12	100%	100%	100%	100%	50%	50%	0%	0%	0%
S-13	100%	100%	100%	50%	75%	100%	25%	0%	0%
S-14	100%	75%	100%	100%	100%	100%	75%	0%	0%
S-15	100%	100%	100%	75%	50%	75%	25%	25%	0%
S-16	100%	100%	100%	100%	100%	75%	0%	0%	0%
S-17	100%	100%	100%	100%	75%	50%	0%	0%	0%
S-18	100%	100%	100%	100%	75%	50%	25%	0%	0%
S-19	75%	100%	100%	100%	100%	50%	25%	0%	0%
S-20	75%	100%	100%	100%	0%	0%	0%	0%	0%