UNIVERSITEIT VAN AMSTERDAM

Perception of the non-native phone **[g]** in Dutch:

where lies the VOT boundary?

Author:	Lisan Schuttenhelm
Student no.:	5740495
Supervisors:	Dr. Silke Hamann
	Prof. Dr. Paul Boersma
Course:	MA Thesis General Linguistics
Domain:	Phonetics
College:	University of Amsterdam
Year:	2012/2013
Date:	08-08-2013

ABSTRACT:

This study investigates the perception of non-native phone [g] by Dutch listeners. Existing literature on L2 acquisition suggests that non-native sounds initially assimilate to native phoneme categories. The present experiment shows that this is also the case for Dutch perception of [g] by demonstrating that 29 out of 43 participants do not discriminate between [g] and [k]. The articulatory consequence is productions such as [kukəl] and [kol] for the loan words 'Google' and 'goal' respectively. The 29 participants that were unable to discriminate between [g] and [k] had no problems discriminating between [d] and [t] (which are native Dutch phonemes), indicating that this inability to discriminate is specific to the /g/-/k/ contrast. The 14 participants that did discriminate between [g] and [k] show that the non-native phone [g] can eventually be acquired. These 14 participants are actively occupied with a L2 that employs /g/ as a phoneme and have had substantial exposure to this L2. Foreign language experience was expected to directly influence the ability to discriminate non-native sounds. However, 10 highly experienced listeners did not discriminate between [g] and [k], contrary to the expectation. The data presented in this paper can therefore neither confirm nor reject that foreign language experience plays a crucial role in non-native perception. The AX discrimination task shows a peak for /g/-/k/ around -20 ms, while it shows a peak for /d/-/t/ around 0 ms. This peak for /g/-/k/ would generally be expected to be at a positive VOT point and also more positive than that of /d/-/t/. These findings can be interpreted as odd and may be due to methodological effects, or, if not, the boundary for /g/-/k/ might reflect the acquisition of /g/ by L2 learners who establish a separate category and are perhaps still in the phase of adjusting the boundaries of that category to optimally match the L2 input. Whatever the case, the data presented here can again not make solid claims about this matter, for it cannot be excluded that the finding is due to methodological effects. Future research thus will require an improved methodology as well as an objective way of measuring foreign language experience, in order to be able to make more solid claims about non-native perception.

PREFACE:

The idea for this subject was formed when I heard my mother pronounce the internet search engine 'Google' ([gugəl]) as ' kookle' ([kukəl]). She replaced /g/ by /k/. I then started wondering how this could be possible. Can she not produce /g/? Can she not perceive /g/? Is it because /g/ is not originally in the Dutch language? Why can other Dutch people pronounce /g/ (as it appears) effortlessly? I aim to discuss these questions in the following thesis, which I write in order to complete the Master General Linguistics at the University of Amsterdam. A perception experiment will be conducted in order to reveal what happens when Dutch participants perceive [g]. The expectation is that my mother is not the only one, and that the perception of [g], which is not a contrastive phoneme in Dutch, is in fact influenced by (the amount of) second language (L2) experience. Essential in explaining such phenomena are L2 perception theories and accordingly views on how (either L1 or L2) phonological systems are formed. Such theories will be discussed and used to shed light on the phonological issue at hand.

ACKNOWLEDGEMENTS:

I would hereby like to thank the people that made it possible for me to complete this thesis. It is no secret that it has not been easy, as constant procrastination had its consequences. I am therefore very happy to finally present this final endeavor. Thank you mom, grandma (also financially) and Arno, for pushing me in a healthy way. For having the patience with me to let me do this in my own way. It must have been frustrating to see this trajectory go on for many, many more months than it should have. I appreciate your support. Of course I would also like to thank my supervisors Silke Hamann and Paul Boersma for their advice and great knowledge, and for providing me with clear directions whenever my mind was clouded. Thanks to David Weenink for lending his voice for the stimuli and Dirk Vet for providing the acoustic gear needed. Also Karin Wanrooij did me a great favour by 'sharing' her Spanish participants with me for the pilot. And it should be no surprise that I want to thank everyone participating in this experiment, you all made this happen!

CONTENTS:

I. INTRODUCTION:

Languages differ in their phoneme inventories. One language can employ a sound contrast with which another language is unfamiliar, or which is not phonemically distinctive in another language. What happens with the perception and production of such a contrast by someone who is not familiar with it is an interesting topic of research. The acquisition of non-native phonemes by second language (L2) learners does generally not come easily, and sometimes not at all. Often a foreign accent is maintained. The extent to which a non-native phoneme can be successfully acquired is said to be influenced by variables such as age of onset (which is generally taken to be the most significant variable, especially in association with the concept of a critical or sensitive period, see for example Flege (1987:1) and Ruben (1997)), length of exposure to/ length of residence in a L2-speaking environment, native language (L1) background, quantity and quality of L2 input, amount of L1 and L2 use, phonetic training, language learning aptitude and motivation (Piske et al., 2001; Flege, 2009; Aliaga-García, 2007).

Generally theories accept the view that (at least initially) L2 acquisition is hindered by L1 interference (e.g. Kuhl, 2000, 2004; Flege, 1987:2; Brown, 2000; Best, 1994; Liberman et al., 1957). Often the consequences of this imperfect acquisition are rather harmless, except for perhaps social implications, as long as the difference in pronunciation does not comprise a phonemic distinction in the L2. However, it becomes problematic when it does concern a phonemic distinction. For example, a language with two stop categories (i.e. voiced and voiceless) like Dutch, will typically partially overlap with a language with three stop categories (voiced, voiceless and aspirated) like Thai. The following example describes a situation concerning the same number of phonemes, but different phonetic realizations. In word initial position, when voiceless stops become aspirated in English (Abramson & Lisker, 1964)- e.g. [p^hƐt] -, a Dutchman is likely to produce [pƐt] which may be interpreted by an Englishman as /bƐt/ and vice versa. A much more striking example though is that of Japanese, in which there is no phonemic distinction between the liquids [r] and [l] - which acoustically mainly differ in F3 (third formant) transition (Miyawaki et al., 1975). This distinction is present in many other languages such as English. In English, /red/ and /led/ form a minimal pair. In Japanese, this is not the case; they can use either [r] (or rather: [ɾ]) or [l] (or anything in between) for the same word because it will not influence its meaning. As a consequence of this lack of distinction, Japanese cannot discriminate between [r] and [l], which obviously leads to problems when acquiring a language such as 'Engrish'. An interesting point to be made here is that this inability to discriminate appears to be specific to the speech mode and is not a matter of general auditory perception. Next to conducting a discrimination task with synthesized [ra] and [la] stimuli, Miyawaki et al. (1975) also investigated discrimination of isolated F3 transitions, which are non-speech sounds. Where Japanese failed to discriminate between speech sounds [ra] and [la] (even after receiving at least ten years of formal English language training), they attained the same accuracy scores as did Americans in discriminating the non-speech stimuli. It is the linguistic context that causes the aforementioned perceptual effect; a discrepancy may be perceived on the acoustic level, but not on the phonological level (Brown, 2000).

Many of these factors also apply to the topic of this thesis, namely the perception of [g] by Dutch listeners. The Dutch phonological system lacks this phoneme (Abramson & Lisker, 1964) and subsequently it has no distinctive function. Technically, [g] does appear in cases of phonological assimilation (e.g. /zɑkduk/ is pronounced as [zɑgduk], /k/ becomes voiced under the influence of the following /d/), but this has no lexical consequences. The main difference between [g] and [k] is the voicing in [g]; an acoustic cue which differentiates all other stops in Dutch (/b/-/p/ and /d/-/t/). In theory [k] and [g] could be used interchangeably (like Japanese [r] and [l]), for these sounds would not form a minimal pair, e.g. 'koe' ([ku]) and 'goe' ([gu], which is a non-word). Although no Dutch speaker would actually use [g] instead of [k], some Dutch speakers appear to use [k] instead of [g] - something that becomes apparent with the coming of loan words. As mentioned in the preface, some Dutch speakers say [kukəl] instead of [gugəl] ('Google'). And once aware of this phenomenon, one can often catch a Dutchman pronouncing the English loan word 'goal' (/gol/) as [kol], even though the original English pronunciation more or less forms a minimal pair with Dutch 'kool' (/kol/). It would seem that these speakers map the sound [g] onto their existing L1 category /k/. The questions this paper aims to investigate are whether or not Dutch listeners acknowledge /g/ as a separate phonemic category contrasting with /k/, and if they do show a boundary (on the voice onset time continuum, discussed later) between these categories, whether this boundary might be located differently (again, on the voice onset time continuum) from other two stop categories languages which do employ both /g/ and /k/. A reasonable suspicion rises that the extent to which native Dutch speakers perceive and produce [g] depends on the level of experience with [g] through foreign languages. This could mean that at some point, after substantial L2 experience, a new category for /g/ is formed, which continues to adjust itself until it is set to optimally match the input occurrences (in line with gradual learning, see Boersma & Escudero, 2003; Aslin & Pisoni, 1980). A more detailed discussion of L2 phonological acquisition theories follows in chapter II. The hypotheses derived from these theories are stated in chapter III.

To gain insight in the perceptual processing of [g] by Dutch listeners, a discrimination test was conducted accompanied by a questionnaire on language background. A description of the used methodology can be found in chapter IV. Chapter V and VI consider the results of the experiment and the resulting discussion.

II. THEORETICAL BACKGROUND:

Insights from the broad research area of language acquisition offer guidance in hypothesizing what the participants of the current experiment (chapter IV) will experience perceptually. If it would indeed be the case that certain Dutch people 'cannot'[1] perceive or produce [g] (possibly due to a lack of experience with this sound), the cause should be found in the way the brain deals with foreign sounds. What happens when a foreign sound is perceived? Is it perceived as such but subsequently ignored by the L1 phonological system? Is the sound perceived as or assimilated to an existing L1 phoneme? Will one ever fully be able to acquire a non-native phoneme? In this section some theories that suggest answers to these questions will be discussed. This discussion serves as an overview and global theoretical

[1] It is not per se a matter of inability, but more a matter of experiencing difficulties because non-native sounds are perceived in terms of L1 phonology. Non-native sounds may (initially) assimilate to native categories. Acoustic differences can in fact be perceived (as Miyawaki et al. (1975) showed for non-speech sounds), but are filtered out as noise or treated as intra-category variation during the mapping of the signal onto L1 categories when this discrepancy yields no relevant contrast in meaning in L1 phonology (e.g. Flege, 1987:2).

background; all the specific implications, experimental data and potential problems that come with these theories are beyond the scope of this thesis and therefore cannot be reviewed in detail.

Before establishing a phoneme inventory according to their native language environment, infants are in fact not yet constrained by their native language in their discrimination of contrasts (Kuhl, 2004). At that stage they can also discriminate non-native contrasts. This ability seems to fade as soon as perception is adjusted to L1, which according to Kuhl (2004) and Werker & Tees (1984) is during the second half of the first year of life. Once phonological categories are established, discriminating non-native contrasts somehow potentially becomes problematic. This observation might seem drastic, but for example Best (1994) indicates that this effect of native language on perception of non-native contrasts - called interference - is not absolute or permanent, as supported by her data that show that certain contrasts can still be discriminated and acquired at a later age. Flege (1987:2) also argues that discrimination of non-native sounds at a later age is possible, depending on the (dis)similarity of the contrast with respect to L1. Kuhl (2004) seems less specific about predictions on foreign accent but quite firm on the concept of L1 interference, stating that the detection of phonological patterns early in life will lead to neural commitments and speech motor patterns which are difficult to modify later. All theories discussed here accept that the acquisition of a native language has an effect on acquiring a second language; they differ in the predictions their theories make.

Best (1994) describes how children (aged 6-8 months) initially employ a strategy of detecting general information in speech, in which phase they are able to discriminate many native and non-native contrasts as they are not yet constrained by their L1. While gaining language experience, this strategy becomes more and more language-specific. They start recognizing and producing language-particular elements in terms of gestural patterns[2], in which phase they attend only to native contrasts (aged 10-12 months). Interestingly, she found the age around 8-10 months to be an intermediate stage, which shows the beginnings of the emergence of a native phonological system when contrast discrimination is neither completely general nor language-specific. A quote summarizes her view on phonological development with respect to non-native phones: "The nature of the experiential effect on perception of non-native segments appears to be largely an adjustment of selective attention rather than a permanent revision of the initial state of sensory-neural mechanisms, it is neither absolute in extent nor irremediable in adulthood, and it varies in degree among specific types of non-native contrast and among individuals" (Best, 1994). She bases the criterion for the inability to perceive a non-native contrast on the similarity with gestural patterns found in L1: if the articulatory properties of a L2 phone are perceived to be similar to those of an existing category, the listener will fail to detect a difference; in which case the non-native phone will assimilate to that most similar L1 category. If the native and non-native phones are too different the listener will in fact perceive a contrast and no assimilation takes place. An important assumption is that assimilation is not absolute; some sensitivity to variation within a category will be retained. This makes possible that a listener can still hear discrepancies within a

[2] With the perception primitive being articulatory information, there is a direct link between perception and production; which is an apparent advantage with respect to theories that have to postulate a specialized phonetic module to translate abstract phonetic representations into speech patterns. This runs counter to for example Kuhl (2000), who states that perceptual representations of speech are stored in memory early in life. These representations subsequently guide the development of motor speech.

category and recognize an assimilated phone as being less than native-like. Best (1994) posits the Phonological Assimilation Model to precisely predict which contrasts can be acquired and which cannot. According to PAM non-native contrasts can be assimilated to native phonemes in these ways:

- The phones of a non-native contrast are similar to two different categories in L1: each non-native phone will be assimilated to its corresponding most similar L1 category. A high level of discrimination is expected because a discrepancy is perceived.
- The members of a non-native contrast both assimilate equally well or poorly to a single L1 category: each non-native phone may be equally similar to or different from the native exemplar. Discrimination is expected to be poor because the sounds assimilated, but still possible because the listener can possibly perceive the assimilated non-native phone as being odd or differing from the native prototype.
- The phones of a non-native contrast both assimilate to a single L1 category, but one phone may be more similar to the native phoneme than the other. They differ in category goodness. If there is a large difference in similarity to the native prototype between the non-native phones (strong category goodness), high discriminability is expected - yet lower than for the inter-category discrimination. A weak category goodness in its extreme form is identical to the equal assimilation to one L1 category, for which poor discrimination was expected.
- The members of a non-native contrast do not assimilate with an L1 category because they are dissimilar beyond acceptance. Rather, the phones will be perceived as non-speech sounds. Moderate to good discrimination is expected, depending on how similar the non-speech sounds are perceived to be.

What would this model imply for, say, a Spanish /k/-/g/-contrast perceived by a Dutch listener? /g/ has no direct corresponding L1 category, but is articulatorily most similar to L1 /k/ (as can be assumed by the Dutch production examples of loan words mentioned earlier), and thus is most likely to assimilate with Dutch /k/. Spanish /k/, of course, will also assimilate with Dutch /k/. However, Spanish /k/ is nearly identical to Dutch /k/, while Spanish /g/ differs from Dutch /k/. The assimilation type is that of a strong category goodness, and thus discrimination is expected (albeit lower than in the case Dutch would have had a corresponding /g/ category). Considering that according to this model perceiving a discrepancy between [k] and [g] should be possible for Dutch listeners, a process of acquisition of /g/ should commence once the listener detects contrastive use of these two sounds in L2 input. An English /k/-/g/-contrast should roughly be processed in the same way. However, recall the phonological 'rule' mentioned earlier for stops in syllable initial position in English. Voiceless stops become aspirated, and voiced stops become partially voiced or even voiceless (Ladefoged & Maddieson, 1996). This means that initial English /g/ becomes more like Dutch /k/, while initial English /k/ becomes less like Dutch /k/. How much the non-native phones differ depends on how aspirated or how voiceless the phones become; one option is to say that they both differ equally from Dutch /k/ and therefore represent the second type of assimilation mentioned above, or a weak category goodness version of the third type of assimilation, in which case poor discrimination is expected. The other option is a moderately strong category goodness version of the third type of assimilation, if English /g/ is (almost) like Dutch /k/ while English /k/ greatly differs from its Dutch counterpart, in which case moderate to moderately high discriminability is expected. Also, the latter would imply that Dutch listeners will perceive English [g̊] as native /k/ and do just fine with producing [g] with the native speech pattern for /k/, but perceive [k^h] as a less than native-like /k/ and start acquiring /k^h/ as a separate phoneme instead of /g/. However, this is unlikely since this example only concerns the specific phonological context for syllable-initial stops. A learner of English will generally be confronted with a voiced /g/, contrasting with non-aspirated /k/. Let us for the remainder of this chapter assume that /g/ represents a typical voiced exemplar ([g]), regardless of what specific L2 might be the input (Spanish, English, etc.).

Flege (1987:2) states that L2 learners are unable to make effective use of sensory input in speech learning. Non-native L2 phonemes will not be produced authentically because the acoustic differences are not accurately perceived due to phonological filtering[3]. He poses the concept of Equivalence Classification, which is applicable to both L1 and L2 phonological development and, like Best's PAM, makes a distinction between 'new' (unlike any L1 phoneme) and 'similar' (acoustic properties similar to (but differing systematically from) those of a L1 phoneme) sounds. It makes the following predictions:

- Identical sounds are rightfully identified as a L1 phoneme.
- New sounds are initially identified via a L1 category (assimilation), but learners will eventually come to recognize (if enough L2 input) that it is a separate phoneme. Establishing a new category is possible, native-like pronunciation is possible.
- Similar sounds are identified as L1 phonemes. Establishing a new category is impossible. Pronunciation can improve with linguistic experience but will never reach native-like levels.

As a result of assimilation, learners will use L1 articulatory patterns for L2 phones. Flege (1987) assumes that L2 learners will never be able to authentically produce L2 phones that differ acoustically from L1 phonemes unless they establish a new category for the L2 phones. Equivalence Classification is advantageous when acquiring L1 phonology, because it helps the infant to perceive all variations of a phoneme (as the result of intra- and inter-speaker variations) as belonging to the same category. However, it is also this mechanism that will lead to a foreign accent in L2. It prevents the learner from making effective use of auditorily accessible acoustic differences between phones in L1 and L2. The fact that L2 learners tend to have problems with pronouncing new and similar sounds authentically suggests that older infants or adults have difficulties establishing the articulatory patterns for these sounds (possibly supporting the concept of a critical or sensitive period). Either this or they continue to perceive these sounds in terms of L1 categories. As regards the similar sounds, for which no separate category will be established, Flege (1987) hypothesizes that as they gain L2 experience, L2 learners will approximate but not achieve the phonetic norms of L2. He bases this on the assumptions that L2 learners are able to auditorily detect the acoustic differences distinguishing similar L1 and L2 phones and that the phonetic representations that guide segmental articulation continue to be modifiable throughout life as the result of phonetic input. This means that L2 speakers can eventually attain a pronunciation of a similar L2 phone that is different from its L1 counterpart but not L2 native-like either. This is in line with the Merger Hypothesis (Flege, 1981), which holds that L2 learners more or less merge

[3] The acquired L1 phonology interferes by filtering out perceptual acoustic differences that are not phonologically relevant in the L1.

the phonetic properties of similar L1 and L2 phones within a single category, resulting in an intermediate value somewhere in between that of L1 and that of L2 (as L2 experience is gained). For example, if non-native /k^h/ with an average voice onset time (VOT) of say +80 ms is assimilated to a single category with L1/k/ with an average VOT of say +50 ms, it is possible that the speaker (after significant linguistic experience) comes to pronounce this category with a VOT that is longer than the L1 value, but shorter than the L2 value, say +65 ms. Voice onset time is defined as the length of time that passes between the release of a stop consonant and the onset of voicing, i.e. vibrating of the vocal folds (Abramson & Lisker, 1964). The VOT for voiced stops is typically a negative value (voicing lead), while voiceless stops typically have positive VOT values (voicing lag/tenuis). The question now is whether [g] is a new sound for Dutch, or similar to an existing Dutch sound (namely /k/). Unfortunately Flege (1987) does not seem to provide clear instructions on how to make these decisions. If it were to be classified as a new sound, Dutch L2 learners of for example Spanish will initially assimilate [g] with /k/, but as they become experienced learners they will establish a separate category for /g/, and eventually pronounce /g/ according to L2 native standards while retaining a native-like pronunciation of L1 /k/. If, on the other hand, [g] were to be classified as similar to /k/, [g] will be assimilated with /k/ and no separate category for /g/ will be established. Highly experienced L2 learners would be likely to eventually pronounce /k/ with an intermediate VOT-value somewhere between that of L1 /k/ and L2 /g/ (i.e. merging). The former seems to be more consistent with informal observations.

Brown (2000) adopts an apparently more nativist approach to language acquisition. She states that it is not phonological representations that constrain L2 perception, but features. A phoneme has a unique structural representation in terms of phonetic features (feature geometry). Features concern properties such as type of sound and place and manner of articulation. These features are universal and are acquired in a gradual and systematic order. In a L1 phonological system, a phoneme is only specified by the features that are needed to distinguish it from other phonemes. If a particular feature of a phoneme is redundant or predictable (i.e. not necessary to distinguish it from other phonemes) the feature will be absent from the underlying representation (in line with the notion of Minimally Contrastive Underspecification (Brown & Matthews, 1997)). A learner has to detect the use of contrasting features in the input before (s)he can start to acquire the specific phoneme(s). Like the theories already discussed, Brown (2000) describes how infants start off with the general capacity to discriminate all contrasts, but show a decline in this capacity (with some exceptions) with the construction of L1 phonology. The newly acquired layer of phonological structure mediates between the acoustic signal and the linguistic processing system and is responsible for filtering out irrelevant noise. Again, this can have a negative effect on L2 perception: variations in acoustic signal are funneled into L1 categories by the L1 phonological system and are thus treated as intra-category variation. In fact, this same variation may actually be inter-category variation in a L2, yielding differences in meaning. Initially, in the beginning stage of L2 acquisition, non-native contrast will be mapped onto L1 categories. According to Brown (2000), L2 learners will acquire only those non-native contrasts that they perceive as distinct sounds. Acquisition is triggered when contrastive use of these distinct features in L2 is detected (for which accurate perception of that phonemic contrast is required). Conversely, if the learner does not perceive a discrepancy in a non-native contrast, no acquisition of that contrast will take place. The learner will fail to distinguish these sounds as they are perceived as the same category (assimilation). In order to predict what contrasts a L2 learner will or will not perceive Brown posits the Phonological

Interference Model. This model is based on the similarities or differences in features that are present in the L1 and L2 grammar. It holds the following predictions:

- If the native grammar lacks the feature used in a L2 contrast, the learner will be unable to accurately perceive the L2 phoneme and thus cannot acquire the contrast. The L2 phones will be mapped onto a single L1 category (assimilation).
- If the contrasting feature is present in the native grammar, this will facilitate perception of the non-native contrast, regardless of whether this specific segment is part of the L1 inventory. This will be a progress of gradual enhancement until native-like levels are attained.

A quote clarifies this positive L1 interference: "despite a lack of acoustic, phonetic or phonemic experience with a particular non-native contrast, a speaker's experience perceiving native phonemic contrasts along an acoustic dimension defined by a given underlying feature (for example, voicing) permits him or her to accurately discriminate any non-native contrast that differs along that same dimension" (Brown, 2000). This seems to be advantageous for Dutch learners of the /k/-/g/-contrast, because the feature distinguishing this contrast is [voice]; a feature that is present in Dutch phonology and distinguishes all other stops in the Dutch inventory (/b/-/p/ and /d/-/t/). The Phonological Interference Model thus predicts that Dutch listeners will be able to perceive the /k/-/g/-contrast, because Dutch employs the feature that distinguishes the L2 contrast, i.e. [voice], and acquisition of a separate category for /g/ will be triggered as soon as contrastive use in L2 is detected. Although they both use the concept of new and similar sounds, Flege (1987) and Brown (2000) seem to make opposite predictions. Flege (1987) claims that the acquisition of similar L2 phones is not possible, while Brown (2000) suggests that it is precisely these sounds that will be acquired (with 'similar' meaning that the L2 phone and its L1 counterpart (or any other L1 contrast) share the same phonetic distinctive feature). Conversely, according to Flege (1987), new L2 sounds can eventually be acquired, while Brown (2000) restricts this acquisition of new sounds to only those that employ a distinctive feature that is present in L1. Of course, one needs to keep in mind that Flege (1987) on the one hand is talking about phonemes, while Brown (2000) on the other hand is talking about features.

Some of the discussed theories seem to predict irremediable consequences for foreign accent. On a more hopeful note there is research such as that of Aliaga-García (2007), who investigates the role of phonetic training on L2 perception and production. Explicit phonetic training over a period of only six weeks already seemed to have a positive influence on L2 perception and production. Although the long-term effect is not investigated in her study, it seems safe to say that L2 input quality and quantity also certainly play roles in L2 acquisition. Kuhl (2000) also mentions that exaggerated acoustic cues (similar to child-directed speech in L1 acquisition, i.e. Motherese), exposure to multiple instances of a phoneme by many talkers and mass listening experience are effective L2-learning methods. Kuhl (2000, 2004, 2005), like the aforementioned theories, attributes difficulties in non-native perception to the emergence of the native language phonology but does so in a more specific manner. Her Native Language Neural Commitment (NLNC) hypothesis states that neural commitment to the acoustic and statistical properties of native language phonological units enhances future native language learning while not supporting alternative phonetic patterns, thus interfering with the processing of foreign language patterns. Brain measures show that when adults process foreign language speech sounds, a larger area of the brain is activated and for a longer period than when processing native sounds, which indicates neural inefficiency[4]. This inefficiency is a result of using native language listening strategies when processing foreign speech (i.e. L1 interference), which do not allow accurate categorization of L2. Kuhl (2005) states that appropriate phonetic training improves performance and increases neural efficiency, but that it will never lead to native-like levels. The NLNC hypothesis assumes that language acquisition depends on native language phonetic learning and that the degree to which infants remain able to detect non-native phonetic contrasts reflects the degree to which the brain remains open or uncommitted to native language speech patterns. This uncommittedness is expected to have a negative correlation with later language learning, because early native-language speech perception is required for later language learning. The hypothesis directly predicts that infants that discriminate only native contrasts at a certain age (namely 7 months) do better at later language learning, while infants that are still able to discriminate some non-native contrasts at that same age (while never exposed to foreign input) are expected to show a slower progress of language learning. The NLNC hypothesis has interesting implications for the concept of a critical or sensitive period[5]. Kuhl (2004) argues that a sensitive period 'opens' with the exposure to language, when a mapping process commences for which infants are neurally prepared. This is when networks in the brain start to commit to L1 speech patterns. As mentioned earlier, infants with successful pattern detection during early infancy are more likely to advance faster towards complex language structures. But when language input is substantially delayed it will be impossible for the infant to achieve native-like skills, indicating a critical period. Kuhl is an advocate of the theory of Statistical Learning (which will not be elaborated here). It implies that an infant's category reflects the distribution of a particular phoneme in the language input, and that they are sensitive to the degree of variability in that distribution. A distribution will continue to reflect variability with the gaining language experience of the child (experience with various speakers and productions). At some point, after substantial listening experience, the production of a phoneme by a new speaker will no longer cause a change in the underlying distribution. The category stabilizes. This could mark the 'closing' of the sensitive period for phonetic learning. After this point, learning would decline. No testable claims about L2 development are posed in the articles by Kuhl.

Finally, the Gradual Learning Algorithm as described by Boersma (1997) and Boersma & Hayes (2001) will be discussed here, in particular the application of the algorithm to specific L2 learning as described by Escudero & Boersma (2004). The GLA can simulate both L1 and L2 language learning. It is based on a probabilistic version of Optimality Theory, which determines the optimal output based on a given input by means of a constraint ranking. Constraints guide the interpretation of acoustic cues by, to give an example, telling the listener to not perceive a VOT of -120 ms as /b/. This constraint will typically occur with other constraints such as "[-120] is not /d/" and "[-120] is not /g/", where the latter two constraints will be ranked higher than the first mentioned constraint. An incoming acoustic cue of [-120] then will be perceived as /b/, because other candidates (e.g. /d/ and /g/) violate the higher ranked constraints. Such constraints can also be used to define a phonological category, which is formed on the

[4] Unfortunately, Kuhl (2000, 2004, 2005) does not elaborate on this point; she does not specifically discuss how adults learn foreign language phonology in the long term (can they acquire non-native sounds?) and what exactly this would look like neurologically (and precisely how it differs from neural commitment in children).

[5] The use of these terms can be rather confusing. In Kuhl (2004), Kuhl speaks of a sensitive period, while she uses the term critical period for the same observations and assumptions in Kuhl (2005).

basis of probabilistic information. The Gradual Learning Algorithm starts with equally ranked constraints and eventually leads to a native language-appropriate constraint ranking by a process of constraint reranking. If a learner detects that the output (i.e. the optimal candidate) is incorrectly selected, for instance because of semantic context, constraints are reranked so that the right candidate becomes more likely to be selected in the future. The knowledge behind the perception process is this formal perception grammar. Learners want to minimize the chance of miscomprehension, and do so by making decisions that lead to maximum-likelihood behavior (an optimal listener will always interpret an incoming sound the way the speaker most likely intended to), as follows from the learning algorithm. To achieve this, the appropriate constraint ranking must be selected. The same principles apply to L2 learning, but supplemented by the Full Access hypothesis and the Full Transfer hypothesis which were originally posed by Schwartz & Sprouse (1996). The Full Transfer hypothesis holds that the beginning L2 learner transfers his/her entire L1 system to the interlanguage system when starting L2 development. The Full Access hypothesis holds that the L2 learner has access to all the language development mechanisms that were also available for L1. They can use L1-like learning mechanisms to increase L2 native-like performance. These mechanisms typically include boundary shift, category creation and increased use of marked structures (Schwarz & Sprouse, 1996). When acquiring a L2, the learners will start out with the constraint ranking they have developed for their L1, and therefore the L1 categories they established. The L1 category boundaries decide how L2 phones will be perceived. This L2 initial state will most likely not be good enough for the L2 language environment, and this, according to the maximum-likelihood behavior, will encourage learners to start adjusting to their L2 by accessing their native language acquisition devices (or however one prefers to call it) and modify their structures (creating categories or reduce/split/merge existing categories) or processes (shifting boundaries). Distributional learning introduces new constraints (Escudero & Boersma, 2004). The continuous ability to adjust the underlying constraint set and the availability of perception processes predicts that L2 phonemes can be acquired and also, native-like L2 proficiency is possible. Assimilation of L2 phones to L1 categories is possible since L2 learners initially perceive in terms of their L1 category boundaries, but this theory predicts that this problem can be overcome by modifying the particular categories. An interesting point is that the manner of modification might be different from the development native speakers of the L2 go through when acquiring that particular contrast. So it might be the case that, although it leads to native-like perception, the underlying perceptual structure for a L2 learner is different from a native L2 speaker. Being able to create new distinctions and shift boundaries (as a result of gradual learning), the prediction is that L2 learners will eventually be able to reach native-like levels. In short:

- L2 input is initially perceived in terms of the L1 phonological system. The Gradual Learning Algorithm ensures that constraints are reranked until an optimal input-output relation is achieved, categories (and their boundaries) can be created or modified. Native-like proficiency is expected. Although acoustic discrepancies are perceived, initially L2 learners may discriminate a certain L2 contrast poorly if this contrast is assimilated to a single L1 category.

This model claims to combine the predictive power of Best's Phonological Assimilation Model, Flege's Speech Learning Model (not directly discussed here) and Schwartz & Sprouse's Full Access and Full

Transfer theories. However, the predictions made by Boersma's model go beyond those of abovementioned models, for those predicted that certain contrasts could never be fully acquired: contrasts that can in fact be acquired according to the GLA. For the perception of [g] by Dutch listeners, this theory is most likely to suggest that Dutch L2 learners initially perceive [g] as /k/ because there is no phoneme boundary present in L1. However, as a result of distributional learning, eventually (on the basis of L2 language input of course) a boundary between /g/ and /k/ will be established and perception will gradually be enhanced as the GLA reranks the constraints concerned, until an optimal input-output relation is established. This optimal ranking ensures optimal perception of both L1 and L2 /k/ and L2 /g/. It will be possible for Dutch L2 learners to achieve native-like proficiency[6].

III. HYPOTHESES:

On the basis of the theoretical background as described in the previous section, some hypotheses concerning the perception of [g] by Dutch listeners can be posed. Here follows a summary.

General:

- Perception shows L1 interference by means of assimilation of L2 sounds with L1 categories (as a result of phonological filtering).
- The articulatory patterns belonging to that L1 category will be used for production of the particular L2 sound; inexperienced Dutch L2 learners will initially pronounce /g/ as [k].
- Although L2 contrasts will initially be perceived in terms of L1 phonology on the level of language processing, the acoustic discrepancies will in fact have their effect on the underlying distributional pattern of the category.
- Gradual learning / effective language training facilitates the acquisition of new categories or the modification of existing boundaries.
- Dutch learners of /g/ will therefore be able to reach native-like perception and production[7].

- Discrimination is expected to be poor in the initial (inexperienced) state, gradually improving with L2 experience until a boundary is formed - at which point high discriminability is expected.
- Productions such as [kukəl] ('Google') can be explained by initial L1 interference during which [g] is assimilated with the Dutch L1 category /k/.

[6] What's more, the algorithm would also predict advancing in the discrimination of /r/ and /l/ by Japanese learners, which still appears to be quite a controversial point in literature. Literature such as Miyawaki et al. (1975) and Takagi & Mann (1995) suggests that Japanese will never reach native-like levels. On the other hand there is literature that indicates that their perception and production can in fact improve (e.g. Lively et al., 1993, 1994) - which is in line with the concept of gradual learning outlined here - but such claims are still vague in terms of ultimate attainment and/or possible limitations on acquisition. Future research on this specific contrast should prove interesting and may either provide overwhelming evidence for gradual learning and it's simulation, or it may yield claims that the model fails to cope with in a true-to-nature fashion.

[7] Depending on the particular L2, the assumption might be required that it is possible for perception to shift boundaries according to language mode (perception adjusts to the language the listener thinks (s)he hears). Otherwise the modification of existing boundaries may lead to merging of the values for L1 and L2.

Theory-specific:

- According to the Phonological Assimilation Model (Best), the foreign /g/-/k/-contrast will initially be mapped onto the Dutch /k/-category. Assuming a voicing contrast similar to other Dutch stops, the category goodness of foreign /k/ will be strong while foreign /g/ will be perceived as odd or less than native-like (type: single-category assimilation with strong category goodness).
- According to the Phonological Interference Model (Brown), the foreign /g/-/k/-contrast can be detected and thus acquired because the distinctive feature [voice] is present in the Dutch language system.
- According to Equivalence Classification (Flege), /g/ can be acquired, assuming that it is classified as a 'new' sound. Initially /g/ will be perceived as /k/, but a separate category will be formed with gaining L2 experience and learners will be able to reach native-like levels.
- According to the application of Optimality Theory / Gradual Learning Algorithm to L2 learning (Boersma & Escudero), /g/ will initially assimilate with an L1 category. Which category this is will probably depend on cue weighting; if cues on place of articulation are more important to the Dutchman than durational cues, this category will be /k/ (otherwise it might have assimilated with /d/). This is consistent with the observation of productions such as [kukəl] ('Google') and [kol] ('goal'). L2 input will trigger constraint reranking and a new category will be formed to achieve an optimal input-output relation and native-like proficiency is expected.

- All theories with the exception of Brown's PIM predict poor discrimination initially, but fine to native-like discrimination for the highly experienced L2 learner. Because /g/ can (eventually) be acquired as a new category, the discrimination test should show the forming of a boundary between /k/ and /g/ with the gaining of L2 experience.
- These theories explain productions such as [kukəl] ('Google') by initial assimilation of /g/ with L1 category /k/, as supported by all theories – again with the exception of Brown's PIM (which predicts proper discrimination and thus acquisition of /g/).

Hypotheses to directly be tested in the present experiment:

- Main: certain Dutch listeners are unable to discriminate between [g] and [k].
- Sub I: the ability to discriminate [g] and [k] depends on (extensive) foreign language experience.
- Sub II: the location of the VOT boundary (if present) between /g/-/k/ as perceived by native Dutch listeners may differ from that found in other languages that do employ both /g/ and /k/.

An option that should also be considered here, is that 'goal' is interpreted by Dutch listeners as [kol], simply because English speakers generally may actually pronounce it that way: recall the phonological rule that voiced stops in syllable-initial position (followed by a vowel) become only partly voiced or even voiceless. This might have been a fine explanation if this would be the only known case. However, this account seems unlikely for the following (slightly speculative) reasons. It is expected that the underlying form of /g/ in English is [g] ([+voice]). The syllable-initial devoicing is a phonological rule that applies specifically to English, Dutch learners will most likely not be aware of it (unless explicitly instructed). Dutch learners of English will most likely (initially) produce English with the production rules of Dutch, supplemented with the potential knowledge they have about English sounds; it is thus expected that - if Dutch learners would have acquired /g/ as a separate phoneme - they will pronounce [gugəl], unaware of the English devoicing rule, unless the learner is experienced to the extent that (s)he is familiar with that specific rule. Also, Dutch people with little to no English experience probably only hear this word from other Dutch people, never in an actual English context, which eliminates syllable-initial devoicing as a direct argument for this group. If syllable-initial devoicing would be the reason for certain Dutch speakers to perceive and pronounce the initial /g/ as [k], one would expect Dutch learners to pronounce [kugəl], since the devoicing does not apply to the second /g/. However, the observation is that they pronounce [kukəl], indicating that there is something going on with the perception of [g] anyway. Furthermore, this particular rule is specific to English. It cannot account for potential perceptual problems experienced with other languages containing /g/, such as Spanish. And, as theories predict and the results will show, these perceptual problems do occur. Regardless of the 'Google' example, the fact remains that /g/ is not originally in the Dutch language (Abramson & Lisker, 1964) and thus forms an interesting subject for L2 acquisition research. This paper focuses only on the perception of [g]. Reporting the articulation of [g] by Dutch listeners would make for an interesting follow-up study.

IV. METHOD:

As mentioned in the introduction, the aim of this paper is to investigate the perception process in Dutch listeners when processing the non-native phoneme /g/. It is hypothesized that, because Dutch lacks a separate category for /g/, [g] will initially be assimilated with the Dutch category /k/. If a learner has a separate category for each phoneme, discrimination is expected. However, if both sounds are perceived as belonging to the same L1 category, accurate discrimination is not expected, because the learner cannot accurately perceive the difference between both sounds. A discrimination task thus makes for the perfect tool for revealing the underlying categorical structure. The purpose of this experiment is to see whether or not Dutch listeners perceive [g]; to see if and where they place the phoneme boundary. Crucial in understanding the workings of a discrimination task is the phenomenon of categorical perception. Categorical perception holds that listeners, along a given continuum (e.g. voice onset time), are better at discriminating pairs of sounds with a constant VOT difference that cross a phoneme boundary than they are at pairs that are within a phoneme category. If for instance the VOT boundary between /b/ and /p/ is located at 0 ms, categorical perception predicts that a listener will be able to discriminate the sound pair [-10 ms, +10 ms] because (s)he will perceive two different phonemes (i.e. /b/ and /p/ respectively), while (s)he will not be able to discriminate the sound pair [-100 ms, -80 ms], which is both perceived as /b/. In an AX discrimination task, also called a same-different task, the subject has to indicate whether (s)he thinks a sound pair (like the ones mentioned above) are the same, or different. These sound pairs range from one end of a particular continuum to the other, with the VOT difference between the two stimuli in all pairs remaining constant (for example 20 ms). Categorical perception predicts that, if a boundary is present of course, all pairs that fall within a category yield a 'same'-response while pairs that cross that phoneme boundary are perceived as being 'different'. This is exactly the type of task used in this experiment. The AX task was preferred over a labeling test in which possible options are given, because it is undesirable to provide the subjects with active knowledge of possible options when it is precisely these possible options you want to investigate. More details of the AX discrimination task setup follows in section IV.II.

IV.I. Stimuli:

Four productions of /ba/, /pa/, /da/, /ta/, /ga/ and /ka/ were recorded. For the experiment, only pairs of /g/-/k/ (focus of investigation) and /d/-/t/ (distracters) were used. The stimuli are based on naturally produced sounds rather than synthesized speech. They were recorded by a phonetically trained speaker to make sure /g/ was objectively pronounced. This was confirmed by acoustic measurements (table 1). Although the average VOT for /g/ (-49 ms) was shorter than expected, it is still a typical voiced velar plosive and the sound with the most extreme VOT, i.e. -71 ms (which will be named 'reference sound' from here on; the voiced reference sounds are the ones used for modification later on), suffices for the VOT-range used in this experiment, namely -70 ms to +30 ms. One can see that, except for the average VOT of /b/ (which is still a perfectly fine value for a typical /b/) somehow being shorter than it is for /d/, the measurements match those of Abramson and Lisker (1964), given in figure 1, quite well. The slightly unexpected values for /b/ (with respect to /d/) and /g/ might be due to experimental effects, the speaker's dialect (since no explicit instructions on 'type of Dutch' were given) and/or a general lack of explicit instructions (about speech rate, articulatory emphasis, pausing etc.).

	/b/	**/p/**	**/d/**	**/t/**	**/g/**	**/k/**
Reference	-109	+9	-133	+20	-71	+30
Range	-109 : -62	+9 : +9	-133 : -84	+20 : +18	-71 : -22	+30 : + 26
Average	-86	+9	-108	+19	-49	+28

Table 1: VOT values in ms of recorded plosives.

	/b/	/p/	/d/	/t/	/k/
Av.	−85	10	−80	15	25
R.	−145: −50	0:30	−115: −45	5:35	10:35

Figure 1: VOT values in ms for Dutch as measured by Abramson and Lisker (1964: p. 392).

Out of the four productions of each stop, the production that seemed most average of all four (average burst duration, average burst amplitude), was used for the modification of the burst. An ambiguous burst was made per place of articulation so that it could contain no obvious cues as to whether the stimulus is voiced or voiceless. The duration and amplitude of each of these four bursts was measured, and the average burst duration and amplitude between /g/ and /k/ as well as between /d/ and /t/ was calculated. These numbers were then used in the editing of the separate bursts (figure 2.1 for /k/-/g/, figure 3.1 for /d/-/t/) into one ambiguous burst (figure 2.2 for /k/-/g/, figure 3.2 for /d/-/t/) in Praat (Boersma & Weenink), as described in the following. In the case of /k/-/g/, the duration of the chosen burst for /k/ is 36 ms and that of /g/ is 34 ms. A (albeit very small) part that did not contain high energy levels was cut off the /k/-burst until the duration was about 35 ms, i.e. the average duration between /k/ and /g/. The start and end of selections that are cut out was always placed at zero-crossings so that its deletion did not cause abrupt shifts in the oscillogram. Next, to incorporate the average amplitude between the /k/ and /g/ bursts, the amplitude of the edited /k/ burst was multiplied with a factor that is calculated by taking the average amplitude and divide it by the amplitude of the voiceless burst. For /k/-/g/, this factor is .0309 Pascal / .0322 Pascal = 0.96. The resulting burst is now considered to be ambiguous (by the criteria used in this paper). This same procedure was done for /d/ and /t/ and their specific values, as shown in figure 4. After each ambiguous burst an /a/ was inserted (see figure 2.3 for /k/-/g/ and 3.3 for /d/-/t/). For the alveolar pair the /a/ was taken from a /t/ stimulus, and for the velar pair it was taken from a /k/ stimulus. These /a/'s should contain no cues on voicing of the preceding consonant so that the consonant in question can be interpreted as either voiced or voiceless. The formant transitions seen in the first few cycles of /a/ contain cues on place of articulation, hence the separate /a/ per articulatory group. These formant transitions will help keeping the phonemes natural and recognizable. Now that the bursts are ambiguous the manipulating of the voice onset time can commence.

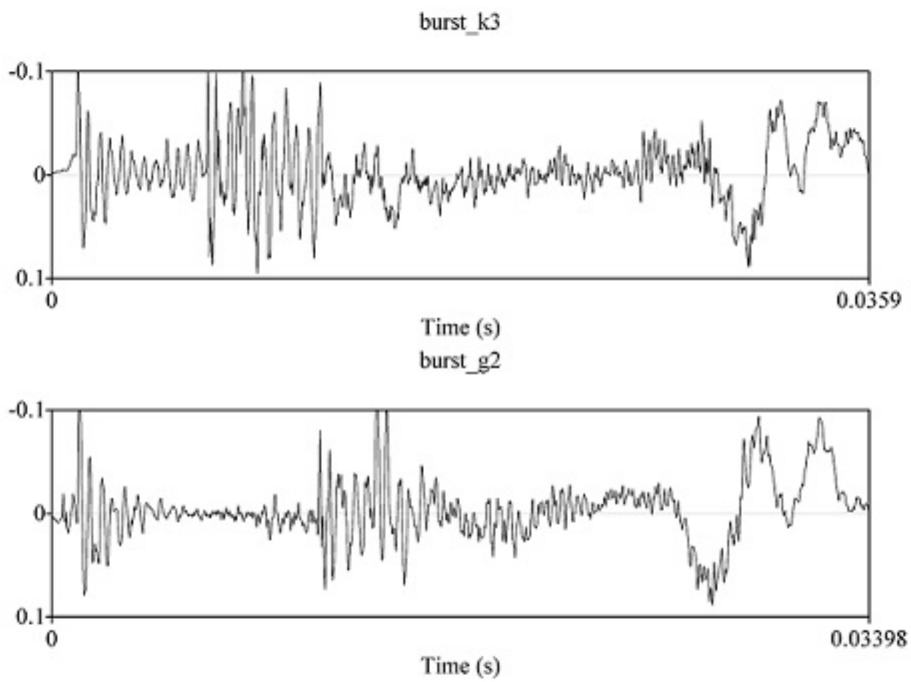

Figure 2.1: Original bursts before editing (/k/ above and /g/ below).

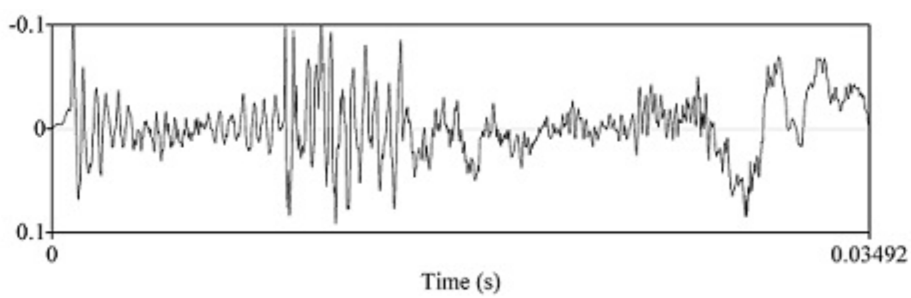

Figure 2.2: Ambiguous /g/-/k/ burst.

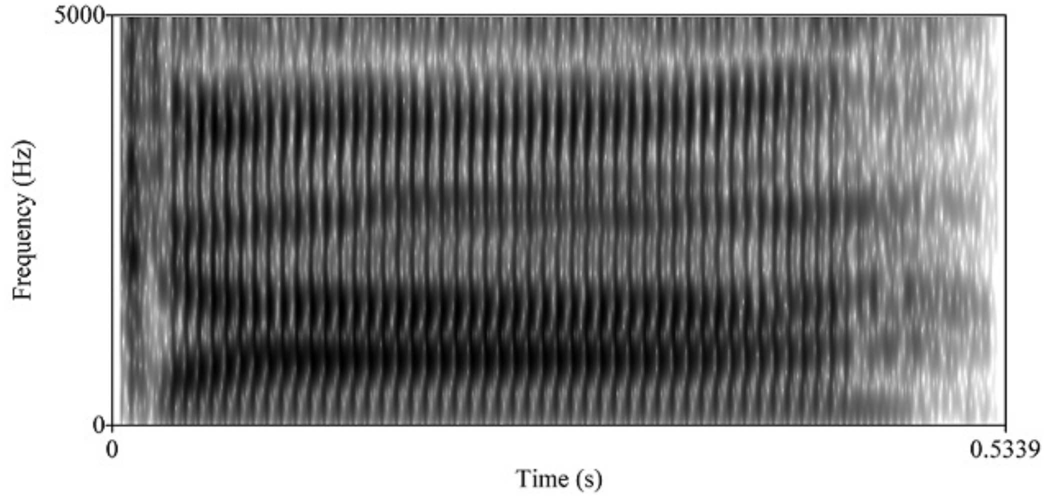

Figure 2.3: Spectrogram of ambiguous /g/-/k/ burst with /a/ attached (formant transitions included).

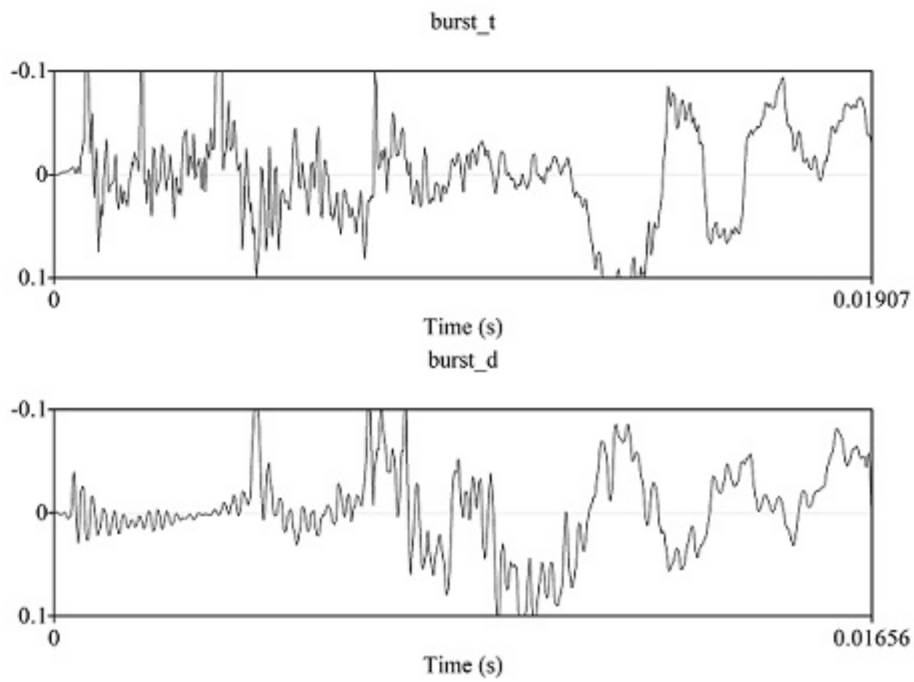

Figure 3.1: Original bursts before editing (/t/ above and /d/ below).

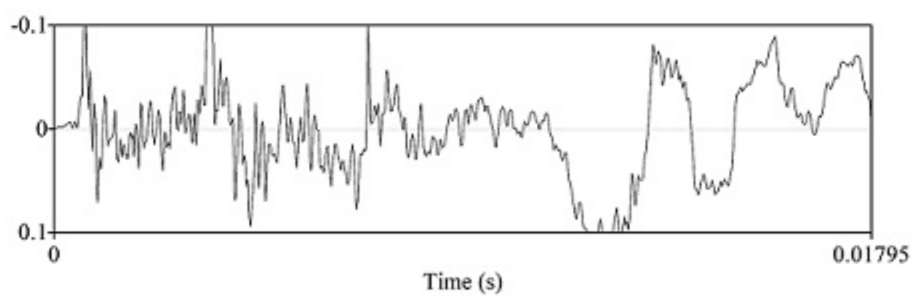

Figure 3.2: Ambiguous /t/-/d/ burst.

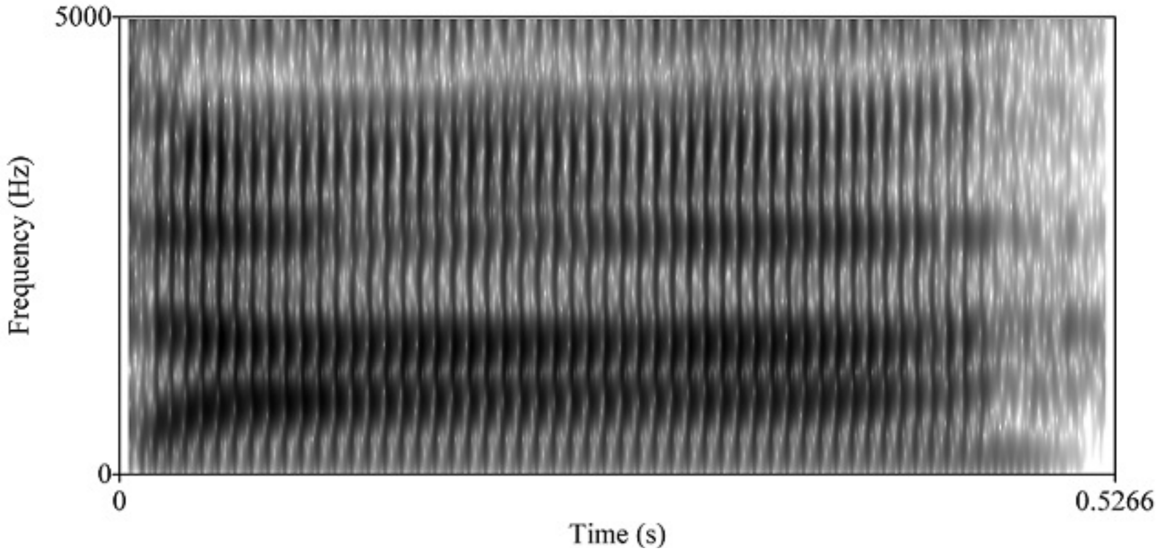

Figure 3.3: Ambiguous /t/-/d/-burst with /a/ attached (formant transitions included).

With the ambiguous burst and vowel taken care of, the next step is to create a gradual VOT continuum. As mentioned above, the range of this VOT continuum will be -70 ms to +30 ms. This range covers the shift in VOT associated with the shift of place of articulation, so that it is applicable for both /d/-/t/ and /k/-/g/. The stimuli vary in steps of approximately 10 ms, resulting in eleven stimuli, i.e. -70, -60, -50, -40, -30, -20, -10, 0, +10, +20 and +30.

stimulus	-70	-60	-50	-40	-30	-20	-10	0
actual	-69	61.2	51	40.3	31.6	21.8	10.6	1.3
steps		-8.8	-10.2	-10.7	-8.7	-8.8	-9.2	-9.3

Table 2: Stimulus times and VOT-steps in ms.

Starting with a description of the /g/-/k/ continuum: the reference VOT of /g/ is -71 ms. As mentioned before, selections that are cut out are always cut at zero-crossings. The smallest possible part, approximately 2 ms, was cut out of the prevoicing, so that the actual VOT becomes -69 ms. With each step, one cycle of prevoicing was removed from the middle of the prevoicing, averagely corresponding to 9.4 ms. Unfortunately, in this case the cycles did not conveniently coincide with the intended 10 ms steps. The actual VOT values are shown in table 2. For the positive VOT-stimuli, the remaining 1.3 ms of voicing was removed from the 0-stimulus, after which a noise period of exactly 10 ms was inserted after the burst. This noise was taken from the original /ka/ just after the burst and before the voicing of /a/. The +10-, +20- and +30-stimuli therefore match their exact values. The reference VOT of /d/ is -84 ms. Half a cycle (approximately 5 ms) was cut off the beginning of the prevoicing and one cycle of slightly more than 8 ms was cut off at the end. The remaining periods conveniently correspond to approximately 10 ms and again, with each step one period was removed from the middle for voiced stimuli, and 10 ms breaks inserted after the burst for the voiceless stimuli. The stimuli are now ready for testing. To give an impression of the continuum, figure 4 below shows the spectrogram for various stimuli from the /g/-/k/ continuum, namely -70 ms, -30 ms, 0 ms and +30 respectively. The green lines underline prevoicing, the orange line indicates the inserted noise. The differences between these stimuli are clearly visible.

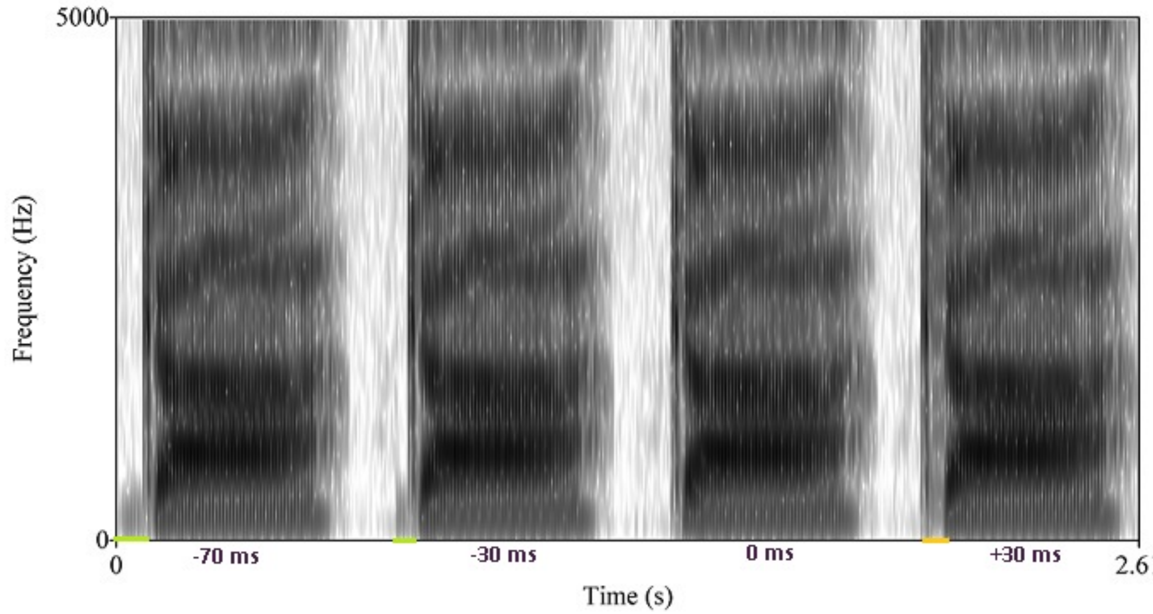

Figure 4: Spectrogram of various stimuli from the /g/-/k/ continuum.

IV.II: AX discrimination task.

The AX pairs presented in the experiment are two-step, i.e. presented in steps of 20 ms (figure 5). This specific number has been found to be the minimal VOT-difference required for a listener to actually potentially perceive a difference between two stimuli (Abramson & Lisker, 1970; Liberman, 1957; Zinoviadou, 2012 (see figure 6)). For each continuum, nine stimulus pairs are formed out of the eleven stimuli (-70 with -50, -60 with -40, ..., and finally +10 with +30).

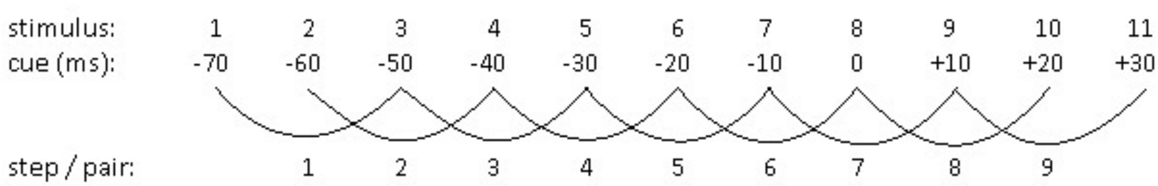

Figure 5: Two-step stimulus pairs.

The effect of categorical perception as visualized by discrimination task results is illustrated by Dutch data for /ba/ and /pa/ from Zinoviadou (2012) in figure 6 on the page below. Subsequently, these data demonstrate the effect of varying VOT-steps in stimulus pairs. Identical pairs (differing by 0 ms) are recognized in virtually all cases; the scores are well above chance. However, when the stimuli in a pair differ by 10 ms (one-step), discrimination rates drop way below chance level (under 50%). The information then cannot be used to make solid claims about discrimination, for the answers can consist merely of guesses. When applying a two-step difference, a precise peak forms - far above chance level. This is no coincidence; the listener perceives a difference between phonemes. This is a typical illustration of categorical perception. It shows that the boundary between Dutch /b/ and /p/ lies at about 0 ms, and that pairs crossing this boundary are well-discriminated while pairs that belong to the same category are not (scores are below chance). A three-step difference on the other hand is somewhat too easy, leading to higher discrimination rates and a broad, less informative peak, which does not show the boundary location as precise as two-step does. Two-step distances of 10 ms is therefore considered the best option for the present AX test, which analogous to Zinoviadou (2012) investigates the use of VOT in the discrimination of Dutch plosives.

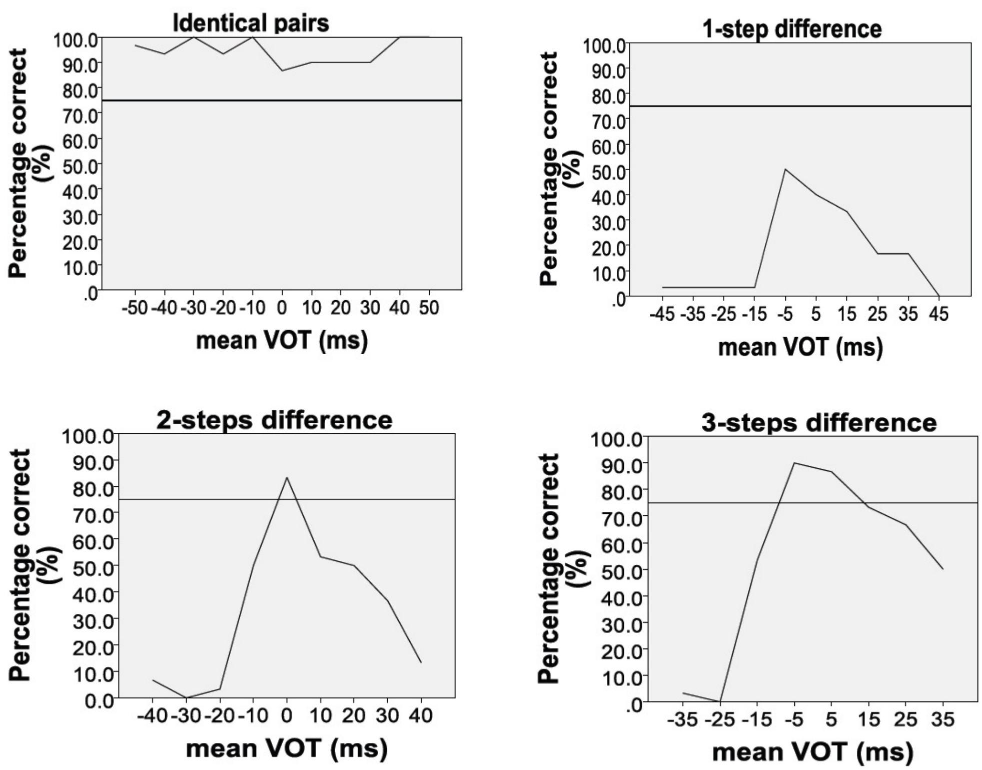

Figure 6: Discrimination rate for Dutch stimulus pairs per x-step difference as tested by Zinoviadou (2012).

The members of the AX pairs described above all differ by 20 ms, and thus are in fact all AB pairs. The correct answer for all pairs is therefore 'different'. However, categorical perception predicts that listeners will actually choose 'same' for all pairs, except for those that cross a phoneme boundary. Theoretically, if a listener has no category boundary, only 'same'-responses are expected. For these reasons, 12 distracter pairs were presented in addition to the original 18 pairs (table 3). Since these 18 original pairs (9 x /g/-/k/ and 9 x /d/-/t/) are in fact different (namely by 20 ms-steps), 7 identical pairs were added. On the other hand, since the lion's share of all pairs are expected to be perceived as 'same' as an effect of categorical perception, 5 pairs with stimuli that differ more than 20 ms were added.

	Identical				Different		
/d/-/t/	-60, -60	-20, -20	+10, +10		-20, +30	-50, +10	
/g/-/k/	-60, -60	-20, -20	+10, +10	0, 0	-20, +30	-70, +10	-70, +30

Table 3: Distracter pairs (in ms) added to the AX experiment: 5 extra /d/-/t/-pairs, 7 extra /g/-/k/-pairs.

The extreme /g/-/k/-pairs [-70, +10] and [-70, +30] ms were added only after giving it a lot of thought. On the one hand, these pairs might hint the listener at the presence of a sound contrasting with /k/, which is undesirable considering the hypothesis that inexperienced listeners will hear only /k/ on the

/g/-/k/ continuum. On the other hand, these distracter pairs form the ultimate test: if a listener really is unable to discriminate between /g/ and /k/, (s)he will also respond 'same' to these extreme pairs. This would provide overwhelming evidence for the assimilation of non-native /g/ to native /k/. For this reason the decision was made to include the pairs in the experiment. It is possible that subjects who do not show a consistent boundary between /g/ and /k/ (i.e. choose 'same' for all original stimuli), do hear these extreme stimuli as 'different'. Firstly, this could be because the acoustic distance is just that great: 80 ms for [-70, +10] and 100 ms for [-70, +30]; while a listener can potentially discriminate between sounds that differ in only 20 ms. Secondly, recall that the edited voiceless stimuli of +20 ms and especially +30 ms had something slightly unnatural about them. Especially when contrasted with a stimulus as extreme as -70 ms, it might be that listeners respond 'different' because the stimulus of +30 ms perhaps might just be the 'odd one out'. The pair [-70, +10] ms should suffer less from the arguments described above and therefore might be less discriminable than [-70, +30] ms. The pairs are considered to be interesting test and distracter items. Furthermore, each subject is explicitly instructed that the stimuli are in a completely random order and that the answers are not per se divided 50/50 - which could for example mean that they only choose 'same' throughout the experiment, and that they should definitely not be distracted by that. If a subject's responses do seem random or extremely odd, the data of that subject will be considered unusable. Such oddness is expected to manifest itself through consistently identifying identical pairs as different while identifying different pairs (≥20 ms) as identical, as well as showing a widespread range of discrimination (e.g. discrimination of multiple pairs anywhere on the continuum) for both /g/-/k/ and /d/-/t/ (which might indicate that the subject was either not concentrated or perhaps even 'paranoid'; clicking 'different' only because (s)he thinks the answers *have to* alternate between 'same' and 'different'). The criteria for exclusion are as follows: the subject consistently identifies at least one identical pair as different while at the same time consistently identifying at least one different (≥20 ms) pair as identical *and also* consistently discriminates 3 or more pairs (pairs containing [+20] and [+30] ms excluded, see V.I) on both the /g/-/k/- and /d/-/t/ continuum.

IV.III: Pilot.

Before using the continua in the actual experiment it is wise to test the stimuli in advance to see what perceptual results they evoke. Surely it is desirable to know whether subjects perceive for instance an obvious /d/ of say -60 ms as an actual /d/. Therefore a labeling test was conducted in this pilot which should give a clear overview of the phonological categories the subjects assign to the stimuli. Participants with a (in most cases) monolingual native Spanish background made suitable subjects for this test because this language include /g/ and the average VOT productions are quite similar to those of Dutch (Abramson & Lisker, 1964). All stimuli were presented in random order and each stimulus occurred three times. Since seven out of the eleven stimuli are voiced, versus the four tenuis or voiceless stimuli, the +10-stimuli of both /t/ and /k/ were added an extra three times each so that there was a slightly more balanced division of voiced and voiceless stimuli. This resulted in 72 stimuli which were to be labeled as either /d/, /t/, /g/ or /k/ by the subject. The script that defined this experiment was written and executed in Praat (Boersma & Weenink). Eleven subjects participated in the pilot, both male and female, of whom none were bilingual. Most subjects did have some experience with English and/or Dutch, but showed only low proficiency. They were explicitly told that the stimulus order was completely random. They were then instructed to, after each presentation of a stimulus, select the consonant that was most similar to the sound they had just heard.

After the first four participants, a quick look into the results revealed some seemingly undesired data. As it turned out, three out of four participants consistently chose /k/ whenever the target item was /t/. This seemed rather odd and was taken to be due to editing effects. With the next two participants, a set with new /t/-stimuli was tested. In these /t/-stimuli a new ambiguous burst and noise was inserted. This did, however, not change the results, and the original /t/-stimuli were put back into the experiment for the rest of the subjects. Some speculations concerning these particular results will be considered here. A possible explanation may be found Spanish. What was not considered beforehand, is that Spanish has a dental /t/ (Martínez-Celdrán, 2003), whereas Dutch has an alveolar one. This can have consequences for the interpretation of VOT, since VOT shifts along places of articulation: the further back the place of articulation (towards the glottis), the more positive the VOT with respect to previous places. This can be exemplified by the measurements of Abramson and Lisker (1964), given in figure 7. One can see that the average VOT values for Dutch voiceless plosives shift from labial: +10 ms to alveolar: +15 ms to velar: +25 ms. The same shift is true for voiced plosives, in case of Spanish this goes from labial: -138 ms to alveolar: -110 ms to velar: -108 ms. Please note that these are averages (of only one or two speakers).

Voice Onset Time in Msec: Dutch
(1 speaker)

	/b/	/p/	/d/	/t/	/k/
Av.	−85	10	−80	15	25

Voice Onset Time in Msec: Spanish
(2 speakers)

	/b/	/p/	/d/	/t/	/g/	/k/
Av.	−138	4	−110	9	−108	29

Figure 7: Measurements Dutch vs. Spanish (Abramson & Lisker, 1964)

One can see that the average VOT for Dutch /t/ is +15 ms, while this is +9 ms for Spanish /t/. Spanish speakers may interpret a longer (i.e. a more positive) VOT as a cue for /k/, because it is just unlikely to belong to /t/. To further investigate this matter, seven Dutch speakers participated in the very same pilot test. Assuming for the sake of convenience that all voiceless stimuli of the /d/-/t/ continuum should be perceived as /t/, the stimuli of +10 ms, + 20 ms and +30 ms are called /t/-items here. Results showed that five out of seven participants, except for a few /d/-occurrences, choose /t/ whenever /t/ was intended; /k/ or /g/ is selected in only 2 out of the total of 60 /t/-items: in contrast to many Spanish subjects. This might possibly demonstrate the effect of the articulatory difference between Dutch and Spanish /t/. The remaining two Dutch participants did however show differing results. Together, they chose /g/ or /k/ in 9 out of 24 /t/-items. Not a single /t/ was selected; they perceived the remaining 15 /t/-items as /d/. It would be interesting to know what might cause these individual differences, however, possible explanations are again limited to the level of speculation. A factor to be considered might definitely be (second) language background, something that unfortunately has not been investigated during the pilot. Since over half of the Spanish speakers do perceive /t/-items as /t/ (or /d/), one can wonder what the difference between these subjects and those that do not consistently perceive

/t/ might be. Second language proficiency, for example, might be a factor of interest. However, because this information about the subjects was not recorded, a clear claim cannot be made at this point. The differing results of the aforementioned two Dutch participants suggest that L2 background might play a role. These are young males who have a very high level of proficiency in English. They are exposed to and actively concerned with English most part of the day - in contrast to the other Dutch participants, who have not had such notable L2 experience. If anything, it is an intriguing question which cannot be answered at this point but would make for interesting future research. In the actual AX experiment of which the results will be discussed in the chapter below, background information of the participants will be recorded and hopefully will be able to explain any inter-subject discrepancies that may be found in the results.

The following remarks conclude the speculations on the perception of /t/. The positive VOT-stimuli (+20 ms, but especially +30 ms), were rather difficult to create. Somehow these sounds did not sound exactly natural. At some point, the stimuli even tended to sound like /kra/ and /tra/. Many different types of post-burst noise were inserted to alter the sounds and eventually the best sounding one (subjectively, as an L1 speaker of Dutch) was selected. Although these stimuli definitely sound like /t/ and /k/ (as judged by various experienced listeners), one might say that there is still something just a little odd about them, in which case it might perhaps have some influence on perception. Also, one could argue that +30 ms is just somewhat to extreme a VOT for /t/, also for Dutch. As figure 7 shows, an average VOT of +25 ms is normally associated with /k/, possibly causing the discussed perceptual results. Whatever the case, this specific matter does not have any influence on the issue to be researched, namely the /g/-/k/-continuum and its possible boundary.

The pilot revealed another issue concerning the VOT boundary between /g/ and /k/. The global VOT boundary appears to be more on the left side of the continuum than expected, i.e. there appears to be a /k/-bias. As mentioned above, each stimulus appeared three times in the experiment. The boundary in this case was set at the (VOT) point from which on a subject consistently chooses the target item in two or three out of three cases. In the example below (table 4) the boundary was placed at -40 ms (because -50 ms, -60 ms and -70 ms were correctly perceived in ≥2 out of 3 tokens). Half of the boundaries found in the pilot results are ≥-30 (see table 5)[8], which is really much more on the left side of the continuum than expected; especially considering that known data (Abramson & Lisker, 1973) indicate that the boundary for plosives is generally 0 ms or higher (i.e. more positive).

Stimulus:	-70	-60	-50	-40	-30	-20	-10	0	+10	+10	+20	+30
1st	g	g	g	k	k	k	k	k	k	k	k	k
2nd	g	g	g	g	k	k	k	k	k	k	k	k
3rd	g	g	g	k	g	k	k	k	k	k	k	k
Correct:	100%	100%	100%	33%	33%	0%	0%		100%	100%	100%	100%

Table 4: Example results table of subject p8 from the pilot study. Boundary at -40 ms.

Subject:	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11
/d/-/t/	-20	-40	-30	-10	-10	-20	-30	0	-20	-30	0
/g/-/k/	-30	-20	-40	-10	-20	x	-40	-40	-30	-50	-20

Table 5: Apparent category boundaries (VOT in ms) for all eleven subjects according to the pilot results.

[8] For p6, no consistent boundary was found at all for /g/-/k/.

A plausible explanation for this tendency may be found in the bursts of the stimuli. As described in the method section, these bursts were made ambiguous qua duration and amplitude. However, it seems that the intensity of the bursts is still somewhat high; possibly containing a cue for voiceless sounds. This might well explain why the results appear to be biased towards voiceless plosives. Whatever the reason, this effect will be considered in the results - especially when comparing to known data - and the discussion.

IV.IV: Participants.

A total of 51 native Dutch subjects participated in the perception experiment. All participants live in the province of Noord-Holland or Zuid-Holland. The age range is 12 to 80 years (see table 6), with the modus being 26 years and the median 43 years. The average age is 43,2 years.

Age:	10-19	20-29	30-39	40-49	50-59	60-69	70-80
Subjects:	2	20	2	5	8	11	3

Table 6: Age range of the 51 participants.

Every subject filled in a questionnaire about foreign language background, since this is a factor which is expected to account for any individual differences and the possible perception of [g] in general. Most of the foreign language experience is gained during high school years. In the Netherlands it is common to have English, German and French as high school courses, as is the case for 67% of the subjects. The group that has educational experience with at least one of these three languages makes for 82% of all subjects. Out of the 51 subjects, there were only 3 that never learned English (6%). One of the questions on the questionnaire was what grade the participants would give themselves for each foreign language on a scale of 1 (poorly) to 10 (fluently); which is of course highly subjective, yet interesting (especially from a social-psychological view). Only 6 people gave themselves a 6 or higher for French. For German, this were 22 subjects. English yields the best self-claimed results; 43 subjects gave themselves a 6 or higher. Therefore English is used here as the main indicator of foreign language experience. Another question was about the length of exposure to the foreign languages, which was explained to the participants as the number of years during which they had been acquiring the particular language and/or have been in substantial contact with it (for example through work or studies). Despite maintaining this definition, responses were still susceptible to subjectivity (i.e. how does a subject define substantial and how accurate is the estimation of number of years). Table 7 gives the data for length of exposure to the most common L2: English (94% of the participants).

Length of exposure:	0-2	3-6	7-9	≥10
Subjects:	0	21	2	25

Table 7: Length of exposure (in years) to English.

There are two notable peaks in the data. The first one, 3-6 years, roughly corresponds to the number of years subjects would have had English class during high school. This suggests that these 21 subjects have never been in substantial contact with English since high school (which for many subjects is a long time ago). On the other hand, 25 subjects report that their length of exposure to English is 10 years or longer.

Interestingly, 17 of the 25 subjects with over 10 years of exposure to English are in the age category of 20 to 29. This suggests that these subjects have generally been using English ever since they have learned it - which in their case is often already during the last years of primary school. This particular group is expected to score very well on discrimination between [g] and [k]. They, along with the other subjects with a length of exposure of 10 years or longer, will form the control group (25 participants), with the exclusion of 4 participants that through other questions did not indicate that they are actively involved with English (i.e. they only use English on vacations while the other participants use it for work, study or an active hobby), resulting in a control group of 21 participants. The 24 subjects that were substantially exposed to English only during high school or not at all are of particular interest to this experiment, for it is presumably this group of subjects that is potentially unable to discriminate between [g] and [k], due to a lack of substantial or active experience with [g] (in any language). They, together with the remaining participants, form the test group (30 participants).

V. DATA:

In this section, the distracter pairs (i.e. the identical pairs and pairs in which the members differed more than 20 ms) are not considered. The only pairs of interest here are those that differ in 20 ms steps, to which, as explained in the previous section, the correct answer is always 'different'. Because of the phenomenon of categorical perception, the expectation is that subjects will in fact always answer 'same', unless a pair crosses a phoneme boundary, in which case they will respond 'different'. Plotting stimulus pair against percentage of correct responses to that pair (i.e. 'different') yields the discrimination rate, and a clear peak should become visible that indicates where the subject locates the VOT boundary for a particular contrast. The percentage correct is calculated as the amount of 'different' responses per pair (maximal 3 per subject *s* per pair, since each stimulus pair is presented 3 times) divided by the total number of possible 'different' responses per pair (3 * *s*). Negative reaction times are subtracted from the total number of possible 'different' responses per pair. For /g/-/k/, the number of negative reaction times is 5 ([-60], [-50], 2 times [-30] and [+10]). For /d/-/t/ this is only 1 ([-20]).

V.I. RESULTS:

Table 8 shows the results for /g/-/k/ per subject and per pair. The number of 'different' responses per pair per occurrence is marked by 'x'. The bright orange lines correspond to subjects that did not perceive any difference at all; the light orange lines correspond to subjects that had a few 'different' responses, but do not discriminate consistently (i.e. in two or three out of three tokens). Consistent discrimination is indicated by the blue boxes, the sum of these for each pair is given at the bottom of each table. The pink boxes represent subjects from the control group (as described in IV.IV). This legend also applies to table 9, which shows the results for /d/-/t/ per subject and per pair. These tables give a clear visual overview on the individual scores as well as the difference in discrimination of /g/-/k/ and /d/-/t/. There are 4 participants, namely C10, T23, T24 and T25, who show discrimination of /g/-/k/ based solely on a stimulus pair containing the +20- or +30-stimulus. As discussed in the method section, there is something slightly odd about the +20- and +30-stimuli as a result of sound editing. These modification -

Pair:	-70/-50	-60/-40	-50/-30	-40/-20	-30/-10	-20/0	-10/+10	0/+20	+10/+30
s	**-60**	**-50**	**-40**	**-30**	**-20**	**-10**	**0**	**+10**	**+20**
T1									
T2									
T3									
T4									
T5									
T6									
T7									
T8*									
T9									
T10									
T11									
T12									
T13									
C1									
C2									
T14		x	x	x	x				
T15		x						x	x
T16*					x				
T17*			x		x				
T18	x		x						x
T19	x							x	
T20	x								
T21*					x				
T22*								x	
C3						x			
C4					x				
C5			x			x			
C6			x	x				x	
C7							x	x	
C8	x					x			x
C9					x				
C10	x	x	x		x	x	x	xx	
T23	x				x			x	xx
T24					x	x	x	xxx	xx
T25*								xxx	xx
T26*		xx	xx	xxx	xx		x		
T27				x	xxx	xx	x		
T28*	x	x		xx	xx	xxx	x	xxx	
T29				x	xx			x	
T30		x	x	x	xx	x			x
C11	xx			xx			x	xx	x
C12			xx	x	xxx				
C13	x	x		xx	xx	xx	x	x	x
C14			xx	xxx	x	x			
C15		xx	xxx	x	x	x	x	x	
C16	xx		x	xx	x	x			x
C17			x		xxx			x	
C18			x	xxx	x	x		xxx	
C19				x	xx	x			
C20				x	xxx				
C21					xx	x		xx	
	2	2	4	7	11	3	0	(7)	(3)

Table 8: All results for /g/-/k/.

Pair:	-70/-50	-60/-40	-50/-30	-40/-20	-30/-10	-20/0	-10/+10	0/+20	+10/+30
s	**-60**	**-50**	**-40**	**-30**	**-20**	**-10**	**0**	**+10**	**+20**
T21*									
T8*				x					
T16*					x	x			x
T17*									x
T22*							x		
T25*					x		x		x
T1				x		x	xxx	x	
T2				x	x		xxx	xx	
T3							xx	x	x
T4						x	xxx		x
T5	x	xx	xx	xx	xx		xx	x	
T6		x	x	xx	xx	x	x	xx	
T7							xx	x	
T9	x	x	x	xx	x		xx	xxx	
T1O			x				xxx	xx	x
T11			x	xx					
T12						x	xxx	xxx	
T13					x	x	xx	xx	
T14	x	x	xx	xx	x	x		x	
T15						xx	x	xx	
T18							xx	xxx	xx
T19		x	x	xx			xx		
T20						x	xxx	x	
T23	xx				x	x	x		x
T24	xxx	xx	xxx	x	xx	xx		xxx	
T26*			xx	xx	xx	x	x	x	
T27	xx	xx	xxx	xxx	x	x	xx	x	
T28*	xx	xx	x	x	xx	xxx	xx		x
T29							xxx	xxx	
T30				xxx	x	xxx	xxx	xx	x
C1				x	xx	x	xxx	xx	
C2							xx	xx	
C3			xxx	xxx	x	xx	xxx	xx	
C4							xx	xx	
C5			xx		x			x	
C6			xx	xx			xx	xxx	x
C7					x		xxx	x	x
C8	xx	xx	xxx	xxx	xxx	xxx	xx	xxx	xx
C9				x	x	x	xx	xx	x
C10						x	xxx	xxx	x
C11		xxx		xx	x	x		x	xx
C12	x					xx	xxx	xxx	x
C13						x	xxx	xxx	
C14			x	x	xx	xx	xx		
C15					x	x	xxx	xxx	x
C16					xx	x	xxx	x	
C17				x	x		xxx	xx	
C18			x	x	x	x	xxx	xxx	
C19		x		x		x	xxx	xxx	
C20				x		x	xxx	xxx	
C21			x				xxx	xxx	x
	5	6	9	13	9	8	36	(27)	(3)

Table 9: All results for /d/-/t/.

artifacts may be the reason for these subjects to answer 'different', instead of discrimination of this pair indicating an actual perceptual boundary, which is also supported by the fact that all other discriminating subjects show improved discrimination around -20 ms. Therefore the choice was made to ignore discrimination based on the pairs [+10] and [+20]. The same holds for /d/-/t/. The abovementioned 4 participants are considered not to discriminate between [g] and [k]. The results then show that as many as 35 out of the 51 participants do not consistently discriminate between [g] and [k], in contrast to /d/-/t/; which 45 out of 51 subjects do discriminate. This means that 6 subjects did not discriminate between [d] and [t] consistently, namely T8*, T16*, T17*, T21*, T22* and T25*, which is odd because the /d/-/t/ contrast is present in Dutch and thus discrimination is expected for all subjects. These subjects did not discriminate between [g] and [k] either. One might argue that the /g/-/k/ results of these subjects are not useable because they do not discriminate between [d] and [t], an odd outcome which might be due to experimental effects. Therefore, the results of these 6 subjects (unfortunately including mother) are discarded in any further analyses. At this point the data of another two participants, namely T26* and T28*, will be excluded on the basis of the criteria mentioned in IV.II (their data appears to be odd or random). The new score is thus 29 out of 43 subjects for /g/-/k/ that do *not* discriminate (and 14 subjects that *do*, which is 32,6%), while 43 out of 43 subjects do discriminate /d/-/t/ (which is 100%). This score suggests there is something going on with the perception of [g], confirming the main hypothesis. Looking at the number of consistent discriminations per pair, stimulus pair [-20] yielded the highest score for /g/-/k/. This indicates that the perceptual boundary for /g/-/k/, for those who discriminate, lies at about -20 ms, as can be seen in figure 8 which shows the discrimination rate for the group as a whole. The numbers on the x-axis (stimulus pair) represent the intermediate VOT for each stimulus pair: [-70, -50] ms is [-60], [-60, -40] ms is [-50], etcetera.

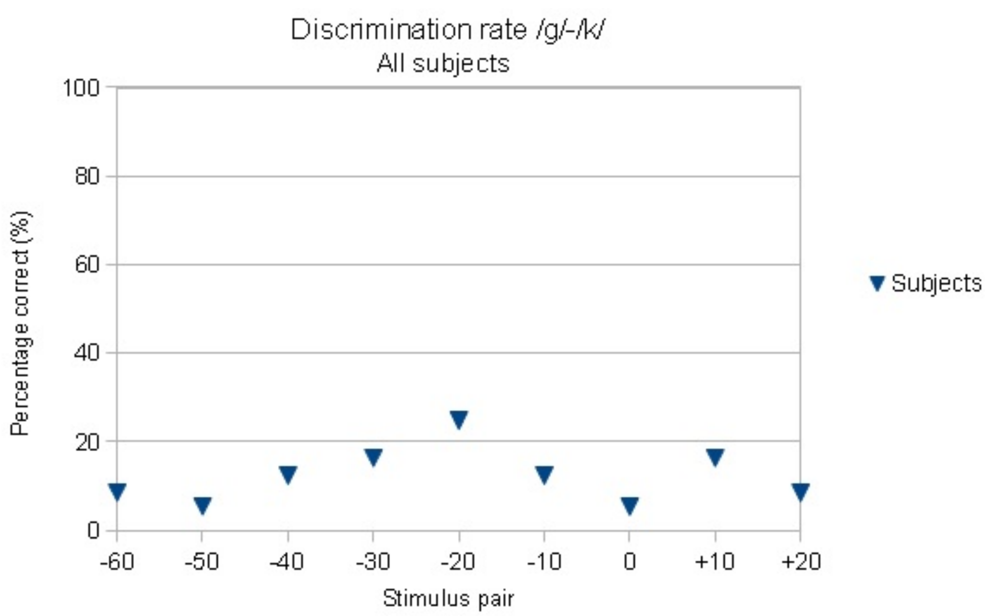

Figure 8: Discrimination rate for all subjects for /g/-/k/.

This does not quite match data in existing literature (e.g. Zinoviadou, 2012): if the boundary for the Dutch /b/-/p/ contrast is about 0 ms, the boundary for /g/-/k/ is expected to be a slightly more positive number (recall that VOT shifts with place of articulation). To be able to say more about the found discrimination, it will be interesting to compare it to the /d/-/t/-data. Considering the choice to ignore discrimination of [+10] and [+20], the VOT boundary between [d] and [t] seems to be located quite unequivocally around 0 ms, which is also visualized by the discrimination rate for all subjects in figure 9.

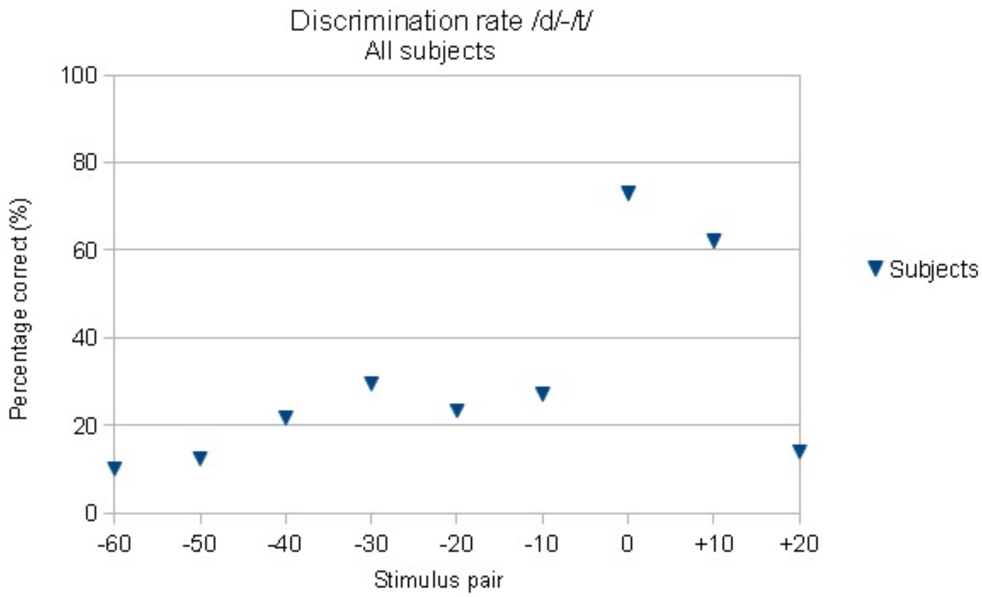

Figure 9: Discrimination rate for all subjects for /d/-/t/.

This is interesting with respect to the boundary found for /g/-/k/, since the /g/-/k/ boundary is expected to be located on the right of the /d/-/t/ boundary, but this finding is the converse. This discrepancy will be discussed in the following sections. The expectation is that foreign language experience is the main predictor for the ability to discriminate between [g] and [k]. The graph in figure 10 shows the discrimination rates for both the test group and control group for /g/-/k/.

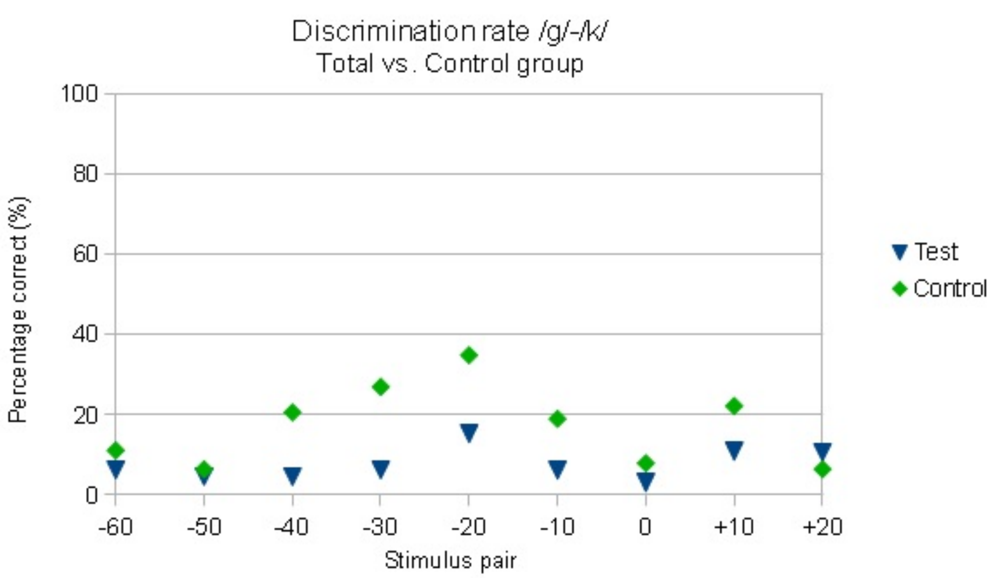

Figure 10: Discrimination rate test vs. control group for /g/-/k/.

The next step is to compare the data for /g/-/k/ of the test (T) and control (C) group in order to find out whether or not the control group scores significantly better on the discrimination of /g/-/k/ than does the test group. This is done by giving each subject a personal score for [-20], according to the number of 'different' responses to this specific pair. This results in the following scores:

T (22)	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	3	2	2
C (21)	0	0	0	1	0	0	0	0	1	1	0	3	2	1	1	1	3	1	2	3	2	

Table 10: Individual scores for [-20] on the /g/-/k/ continuum.

An independent samples *t*-test (df = 41, *t* = -2.009) gives a *p*-value of 0.051. Assuming significance at $p \leq 0.05$, the found *p*-value is on the verge of significance but not quite so. The reason for this lies in the group division. Table 10 shows that the control group scores way below expectation. Based on their foreign language experience, the expectation is that all these subjects are able to discriminate between [g] and [k]. However, only 6 out of 21 subjects seem to do so consistently. This statement is not entirely fair though, for table 8 indicates that subjects that do not show a boundary at [-20], tend to do so at [-30]. Discrimination of pairs other than [-20] and [-30] is considered as noise in the data. Therefore, new individual scores were assigned to the subjects for both [-20] and [-30], as presented in table 11.

T (22)	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	1	4	3	3
C (21)	0	0	0	1	0	1	0	0	1	1	2	4	4	4	2	3	3	4	3	4	2	

Table 11: Individual scores for both [-20] and [-30] on the /g/-/k/ continuum.

The lighter aqua boxes indicate that one of both pairs was consistently discriminated and the darker aqua box indicates that both pairs were consistently discriminated. These scores do yield a significant result from the *t*-test (df = 41, *t* = -2.835): *p* = 0.007. On and around the boundary the control group does seem to do better on discrimination than the test group. This seems to confirm the role of foreign language experience. However, the fact remains that 11 out of the 21 subjects in the control group do *not* consistently discriminate between [g] and [k], which can be taken to be evidence against the role of foreign language experience. The data thus cannot make solid claims about this subhypothesis at this point. This issue will be elaborated in the conclusion and discussion. Evidence for the main hypothesis, namely that perception of [g] poses a problem for some Dutch subjects, was found in the discrepancy between the discrimination scores of /g/-/k/ (32,6% of the subjects) as compared to those of /d/-/t/ (100% of the subjects). To exemplify this discrepancy the interaction effect is measured. This is done by subtracting the individual scores of /g/-/k/ from the individual scores of /d/-/t/. The higher the remaining score, the more difficulties the subject appears to have with discriminating between [g] and [k]. The individual scores for /d/-/t/ pair [0] are given in table 12.

T (22)	3	3	2	3	2	1	2	2	3	0	3	2	0	1	2	2	3	1	0	2	3	3
C (21)	3	2	3	2	0	2	3	2	2	3	0	3	3	2	3	3	3	3	3	3	3	

Table 12: Individual scores for [0] on the /d/-/t/ continuum.

There is no significant difference in discrimination of /d/-/t/ between the test- and control group (df = 41, *t* = -1.571, *p* = 0.124). The individual scores for /g/-/k/ are based on [-20], except for when a subject showed a boundary at [-30] instead of [-20]; for those subjects the score for [-30] is taken instead of the score for [-20], so that discrimination of both pairs is represented in one array. See table 13:

T (22)	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	3	2	2
C (21)	0	0	0	1	0	1	0	0	1	1	2	3	2	3	1	2	3	3	2	3	2	

Table 13: Individual scores for [-20] or [-30] on the /g/-/k/ continuum.

The /g/-/k/ scores are now subtracted from the /d/-/t/ scores, resulting in the following numbers:

T (22)	3	3	2	3	2	1	2	2	3	0	3	2	-1	1	2	2	3	0	-1	-1	1	1
C (21)	3	2	3	1	0	1	3	2	1	2	-2	0	1	-1	2	1	0	0	1	0	1	

Table 14: Individual scores interaction effect /d/-/t/ minus /g/-/k/.

The numbers can range from -3 to +3. The former indicates that a subject did discriminate /g/-/k/ but not /d/-/t/ while the latter indicates that the subject did discriminate /d/-/t/ but not /g/-/k/.The positive numbers (and especially the higher numbers 2 and 3) are therefore of particular interest and are marked by pink boxes. To the eye, the test group seems to have slightly more trouble discriminating /g/-/k/ (17 out of 22 subjects score worse on /g/-/k/ discrimination than /d/-/t/ discrimination) than the control group (14 out of 21 subjects score worse on /g/-/k/ than /d/-/t/). A *t*-test (df = 41, $t = -1.224$) on the scores given in table 14 gives a *p*-value of 0.228, which indicates that there is no significant difference between the scores of the test group versus the control group. Again, this shows that the subhypothesis concerning foreign language experience cannot be confirmed by the data. In general, 31 out of 43 subjects scored worse on /g/-/k/ discrimination than they did on discrimination of /d/-/t/, which is a nice illustration of the apparent problems with the discrimination between [g] and [k]. A one-sample *t*-test (test value = 0, df = 42, $t = 6.114$) on these whole group scores yields a significance of $p < 0.001$.

V.II. CONCLUSION:

The data provide evidence in favor of the main thesis that some Dutch listeners are unable to perceive a contrast between [g] and [k]. They apparently perceive both [g] and [k] as belonging to the same category, indicating that non-native phone [g] assimilates to native /k/. This inability is specific to /g/-/k/, for no such effect is found for the discrimination of /d/-/t/: participants that did not discriminate between [g] and [k], did discriminate between [d] and [t]. The division into groups based on foreign language experience did not correctly predict discrimination of /g/-/k/ by those groups; this factor thus was unable to define exactly who 'some Dutch listeners' are, and the subhypothesis that the perception of non-native contrast /g/-/k/ is influenced by foreign language experience cannot be confirmed by these data. It cannot fully be rejected either, for a deeper look into the results shows that subjects with non-substantial L2 experience did, as expected, not discriminate /g/-/k/ (except for 3 subjects). It is the group of experienced L2 speakers that surprises: 10 subjects that indicate having substantial and active L2 experience did in fact not discriminate /g/-/k/, contrary to expectation. Perhaps there might be other factors explaining those results. This matter will be elaborated in the discussion in chapter VI. The data (see figure 8 and 9) clearly show the VOT boundaries for both /g/-/k/ and /d/-/t/. The boundary for /g/-/k/ is located differently than would be expected from existing literature, especially with respect to the

/d/-/t/ boundary, which is on the right side of the /g/-/k/ boundary while in natural language it is always on the left (see for example Abramson & Lisker, 1964). This can be interpreted in two ways: since the bursts of both continua were edited in the same way, an eventual voiceless-bias (as described in the pilot section) would be expected to be found for both continua. However, this appears not to be the case (and if it is the case, it is far more extreme for /g/-/k/ than for /d/-/t/). One might conclude that the unexpected boundary for /g/-/k/ is due to the non-nativeness of /g/ and possibly visualizes the emergence of a new category, or even that the values for /g/ and /k/ are merged. On the other hand, the chosen burst for /d/-/t/ perhaps contains fewer cues on voicelessness than does the burst for /g/-/k/. In this case, possible voiceless-biases and the locations of the boundaries are not directly comparable. Because a conclusive answer as to why exactly the boundaries for both contrasts differ (as methodological effects cannot be excluded), a second subhypothesis cannot be confirmed (nor rejected); namely that the location of the VOT boundary between /g/ and /k/ might be located differently with respect to other languages that do employ this contrast. The non-native contrast /g/-/k/ appears to be acquirable, in this case through the forming of a new category for [g], as demonstrated by the discrepancy in subjects that did, and subjects that did not discriminate between [g] and [k]. If the location of the VOT boundary for this contrast (i.e. -20 ms) would not be due to methodological issues, it would be an odd boundary location in comparison with languages that originally have /g/ and /k/ in their phoneme inventory (which most likely place the boundary at least at a positive VOT), leading one to wonder whether this found boundary could yield L2 native-like perception, and production accordingly. However, this paper merely investigates the overall ability of Dutch listeners to perceive [g], it cannot make claims about the (native-like) level of perception. Neither does it investigate production. Nonetheless, if Dutch speakers use the articulatory patterns of /k/ for [g] as a result of the initial assimilation of non-native [g] to L1 /k/, production examples such as [kukəl] can be accounted for.

Recall from the theoretical background that the Phonological Interference Model as posed by Brown (2000) was the only theory to predict that no assimilation of non-native [g] to L1 /k/ would take place. Because the phonetic feature distinguishing [g] from [k], namely [+voice], is present in Dutch, listeners are expected to discriminate between [g] and [k] effortlessly. Results from the present experiment seem to reject this model, for 29 out of 43 participants do not show consistent discrimination. The results seem to be in favor of the other discussed theories then (namely the Phonological Assimilation Model (Best, 1994), Equivalence Classification (Flege, 1987:2) and the L2 learning adaptation of the Gradual Learning Algorithm based on stochastic Optimality Theory (e.g. Escudero & Boersma, 2004)), which predict interference of L1 phonology in the form of assimilation of non-native [g] to L1 /k/, as, again, seems to be the case for those 29 participants. The remaining 14 participants that did discriminate support the subsequent prediction of these theories that it is possible for native Dutch speakers to acquire the non-native phone [g]; these subjects seem to have established a separate category for /g/. These results indicate that, in terms of Flege's Equivalence Classification, [g] is a so called new sound for Dutch. If [g] had been classified as a so called similar sound, acquisition of a separate category would not have been possible. In short, all theories with the exception of Brown's PIM are capable of explaining productions such as [kukəl] and do so by stating that non-native [g] initially assimilates to L1 /k/ as a result of L1 interference.

VI. DISCUSSION:

The first issue that deserves further explanation is the influence of L2 language experience. As described above, extensive foreign language experience did not seem to guarantee successful discrimination of /g/-/k/, something that is in fact predicted and proposed by existing literature such as that discussed in the theoretical background. One might argue that the failure of the foreign language experience factor as an indicator for discrimination is unexpected and perhaps can be explained away by other indirect factors. For instance, the highly subjective nature of the questionnaire is undesirable when conducting academic research. An objective language test of some sort would be preferable, although this on the other hand requires more time and resources. Although the questionnaire is expected to generally reflect language background quite well, it is susceptible to errors. For example, if someone misunderstood the question about length of exposure, it might well be possible that this subject fills in a number that is much greater than the number that was actually intended by this question, thus selecting him- or herself for the control group while this is not per se justified. In order to prevent as many cases of this type of error as possible, an extra criterion was added to the control group selection, namely that the subject has to be actively occupied with the L2 (through work, study or active hobbies). This excluded 4 participants from the original control group (which all indeed turned out not to discriminate between [g] and [k]). However, of the remaining 21 participants in the control group, who as far as the questionnaire tells us belong to that group rightfully, as many as 10 subjects did not discriminate between [g] and [k]. About half of these subjects may have met the criteria for the control group inadvertently, as judged by something as unscientific as the experimenter's intuition (since she is acquainted with each of the subjects). However, the other 5 continue to pose a problem: these participants do not discriminate /g/-/k/ but are legitimately experienced with (in this case) English, as again judged by intuition and supported by all questions on the questionnaire. They have been exposed to English for over 10 years, the average age of onset is 9 years for 4 participants – for one subject the age of onset is even 0 (although not raised bilingually), each subject is actively occupied with English through work and/or study and what's more: they hear and use it on a daily basis. They averagely indicate that 75% of the spoken input through media on an average day consists of their L2. A noisy environment will not suffice to account for the non-discrimination of /g/-/k/ by these participants, for they did well on discrimination of /d/-/t/. A speculative argument could be found in language mode. It has been found in the literature (e.g. Zinoviadou, 2012) that experienced listeners are in fact able to shift their phoneme boundaries according to the language they think they hear. Perhaps this is true for these highly experienced subjects: all instructions (both on screen and given by the experimenter) were in Dutch, possibly activating the Dutch language mode. However, this remains purely speculative and it cannot be claimed that this phenomenon is even applicable to this particular situation. If an appropriate explanation (such as methodological effects) for the performance of this specific group cannot be found, the results will have to be hold as evidence against language experience as a factor responsible for discrimination of non-native phones. On the other hand it might be interesting to take a closer look at the data of the discriminating group. Of the 14 /g/-/k/-discriminating subjects, 11 come from the original control group (over 10 years of experience and actively occupied with the L2). The 3 remaining (test group) participants have less than 10 years of experience with English (at least that is what they filled in on the questionnaire) but are actively occupied with English through work, study or active hobby. It is perhaps not that surprising that these subjects did discriminate between [g] and [k]. Certain participants were not selected for the control group due to over-modesty on the questionnaire, subsequently influencing the experiment because these participants can discriminate /g/-/k/ while this is not expected based on the questionnaire. The opposite is also true: quite a few participants gave themselves more credit than appropriate and were placed in the control group while their foreign language experience was in fact not that substantial. If we can conclude that some participants from the control group should not have been placed there and conversely that some participants from the test group should actually have been placed in the control group, this could explain why the division into groups based on the questionnaire did not quite work out and also, why the percentage of correct responses per group is never that high: certain results may cancel each other out. In summary: certain experienced listeners did unexpectedly not discriminate between /g/-/k/, possibly indicating that foreign language experience plays no significant role in the perception of non-native phones. Conversely, the opposite never seems to be the case: virtually all of the subjects that did discriminate but were not in the control group had quite substantial experience with L2; they just did not meet the criteria for the control group. This seems to support the role of foreign language experience. Unless proper explanations can be found for the non-discrimination of /g/-/k/ by highly experienced subjects, the data at this point can neither confirm nor reject this hypothesis.

It should have become clear at this point that the questionnaire as composed for this project was not the best option. The methodology leaves room for further improvement. For instance, instructions of all sorts could have been even more explicit; instructions for the recording of the stimuli, as well as instructions for the experiment. Although the latter was already quite explicit, it might for instance have been beneficial to instruct the subjects to click ‘different’ only when they hear a different sound (i.e. phoneme), and not just any difference - as appeared to be the case for the +20- and +30-stimuli. The results for these specific stimuli and the results from the pilot (indicating a voiceless-bias) imply that the stimuli should have been modified in greater detail. More ways of making the burst ambiguous should have been employed and the noise insertion in the stimuli with a positive VOT should have been replaced by a more appropriate one. Because of these methodological shortcomings, at least to the extent that they might have an effect, the data cannot be interpreted as 100% pure, and it thus possibly limits the claims that can be made on the basis of these data (for instance with respect to the VOT boundary location). Finally, I would like to conclude this section on a positive note. The experiment elicited great results concerning the non-discrimination of [g] by Dutch listeners. The numbers are even greater than expected: 29 out of 43 participants do not consistently distinguish [g] from [k], thus justifying the purpose of this thesis.

VII. REFERENCES:

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

Abramson, A.S. & Lisker, L. (1970). Discriminability along the voicing continuum: Cross language tests. In: *Proceedings of the 6th International Congress of Phonetic Sciences,* Prague*,* p. 569-573.

Abramson, A. S. & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. In: *Journal of Phonetics*, *1*, p. 1-8.

Aliaga-García, C. (2007). The role of phonetic training in L2 speech learning. In: *Phonetics Teaching & Learning Conference (PTLC)*.

Aslin. R. & Pisoni. D. B. (1980). Some developmental processes in speech perception. In: *Child Phonology,* Vol. 2. Perception*,* p. 67-96. New York: Academic Press.

Best, C. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In: H. Nusbaum, J. Goodman and C. Howard (eds.), *The Transition from Speech Sounds to Spoken Words: The Development of Speech Perception*, p. 167-224. Cambridge: MIT Press.

Boersma, P. & Weenink, D. (1992-2013). Praat: doing phonetics by computer (version 5.3.53) [Computer program]. Online: *http://www. praat. org*.

Boersma, P. (1997). How we learn variation, optionality, and probability. In: *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, vol. 21, p. 43-58.

Boersma, P. & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. In: *Linguistic inquiry*, 32 (1), p. 45-86.

Boersma, P. & Escudero, P. (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In: *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics*, Vol. 8, No. 71.85.

Boersma, P. & Escudero, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. In: *Studies in Second Language Acquisition*, 26 (4), p. 551-585.

Brown, C. & Matthews, J. (1997). The role of feature geometry in the development of phonemic contrasts. In: *Focus on phonological acquisition* (16), p. 71.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In: *Second Language Acquisition and Linguistic Theory*, p. 4-63. Oxford: Blackwell.

Flege, J. E. (1981). The Phonological Basis of Foreign Accent: A Hypothesis. In: *Tesol Quarterly*, 15 (4), p. 443-455.

Flege, J. E. (1987:1). A critical period for learning to pronounce foreign languages? In: *Applied Linguistics*, 8 (2), p. 162-177.

Flege, J. E. (1987:2). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. In: *Journal of phonetics*, 15 (1), p. 47-65.

Flege, J. (2009). Give input a chance! In: *T. Piske and M. Young-Scholten (Eds) Input Matters in SLA*. Bristol: Multilingual Matters, p. 175-190.

Kuhl, P. K. (2000). A new view of language acquisition. In: *Proceedings of the National Academy of Sciences*, 97 (22), p. 11850-11857.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. In: *Nature reviews neuroscience*, 5 (11), p. 831-843.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: implications for the" Critical Period". In: *Language Learning and Development*, 1 (3-4), p. 237-264.

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Martínez-Celdrán, E., Fernández-Planas, A. M. & Carrera-Sabaté, J. (2003). Castilian Spanish. In: *Journal of the International Phonetic Association*, *33* (02), p. 255-259.

Liberman, A.M. et al., (1957). The discrimination of speech sounds within and across phoneme boundaries. In: *Journal of experimental psychology*, 54-5, p. 358-368.

Lively, S. E., Logan, J. S. & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. In: *The Journal of the Acoustical Society of America*, 94 (3 Pt. 1), p. 1242-1255.

Lively, S. E. et al. (1994). Training Japanese listeners to identify English/r/and/l/. III. Long-term retention of new phonetic categories. In: *The Journal of the Acoustical Society of America*, 96 (4), p. 2076-2087.

Miyawaki, K. et al (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. In: *Perception & Psychophysics*, 18.5, p. 331-340.

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

Ruben, R. J. (1997). A time frame of critical/sensitive periods of language development. In: *Acta oto-laryngologica*, 117 (2), p. 202-205.

Schwartz, B. D. & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. In: *Second language research*, 12 (1), p. 40-72.

Takagi, N. & Mann, V. (1995). The Limits of Extended Naturalistic Exposure on the Perceptual Mastery of English/r/and/l/by Adult Japanese Learners of English. In: *Applied Psycholinguistics*, 16 (4), p. 379-405.

Werker, J. F. & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. In: *Infant behavior and development*, 7 (1), p. 49-63.

Zinoviadou, S. (2012). *Language mode effects in the perception of voicing*. Tutorial paper, UvA.

Appendix I: Praat experiment file.

```
"ooTextFile"
"ExperimentMFC 5"
stimuliAreSounds? <yes>
stimulusFileNameHead = "[location]"
stimulusFileNameTail = ".wav"
stimulusCarrierBefore = ""
stimulusCarrierAfter = ""
stimulusInitialSilenceDuration = 1 seconds
stimulusMedialSilenceDuration = 0.8 seconds
numberOfDifferentStimuli = 30
      "dt-70,dt-50"     ""
      "dt-60,dt-40"     ""
      "dt-50,dt-30"     ""
      "dt-40,dt-20"     ""
      "dt-30,dt-10"     ""
      "dt-20,dt0"       ""
      "dt-10,dt+10"     ""
      "dt0,dt+20"       ""
      "dt+10,dt+30"     ""
      "dt-60,dt-60"     ""
      "dt+10,dt+10"     ""
      "dt-20,dt-20"     ""
      "dt-50,dt+10"     ""
      "dt-20,dt+30"     ""

      "kg-70,kg-50"     ""
      "kg-60,kg-40"     ""
      "kg-50,kg-30"     ""
      "kg-40,kg-20"     ""
      "kg-30,kg-10"     ""
      "kg-20,kg0"       ""
      "kg-10,kg+10"     ""
      "kg0,kg+20"       ""
      "kg+10,kg+30"     ""
      "kg-60,kg-60"     ""
      "kg+10,kg+10"     ""
      "kg-20,kg-20"     ""
      "kg0,kg0"         ""
      "kg-70,kg+10"     ""
      "kg-20,kg+30"     ""
      "kg-70,kg+30"     ""
```

```
numberOfReplicationsPerStimulus = 3
breakAfterEvery = 30
randomize = <PermuteBalancedNoDoublets>
startText = "Welkom bij dit luister-experiment.

Een paar bestaat uit twee geluiden.
Deze twee geluiden zijn ofwel identiek aan, ofwel verschillend van elkaar.

Als u na het horen van een paar denkt:
-dat de geluiden identiek waren klikt u 'Hetzelfde' aan.
-dat de geluiden anders waren klikt u 'Anders' aan.

Let op: de geluidsparen staan in compleet willekeurige volgorde
en de antwoorden zijn niet per se gelijk verdeeld.

Dit zou bijvoorbeeld kunnen betekenen dat
u zomaar 10x hetzelfde antwoord achter elkaar invult,
of dat u zelfs alleen maar hetzelfde antwoord invult.

Laat u hierdoor vooral niet afleiden! Alles kan.

Klik om te beginnen."
runText = "Zijn deze geluiden hetzelfde of anders?"
pauseText = "U kunt nu eventueel een pauze nemen. Klik als u verder wilt
gaan."
endText = "Einde van het experiment.

Vergeet niet de enquête in te vullen.

Hartelijk dank voor uw medewerking!"
maximumNumberOfReplays = 5
replayButton = 0.4 0.6 0.01 0.07 "Herhaal" ""
okButton = 0 0 0 0 "" ""
oopsButton = 0 0 0 0 "" ""
responsesaAreSounds? <no> "" "" "" "" 0 0

numberOfDifferentResponses = 2
      0.1 0.4 0.3 0.8 "Hetzelfde" 50 "" "Hetzelfde"
      0.6 0.9 0.3 0.8 "Anders" 50 "" "Anders"

numberOfGoodnessCategories = 0
```

Appendix II: Questionnaire.

Leeftijd: ____________________

Geslacht: ____________________

Woonplaats: ____________________

Opgegroeid in: ____________________

Moedertaal: ____________________

Welke talen spreekt u naast uw moedertaal nog meer? Kunt u indien van toepassing per taal aangeven wat voor niveau u daarin denkt te hebben (met een cijfer van 1 (zeer matig) tot 10 (vloeiend)), hoe oud u ongeveer was toen u met deze taal in contact kwam en hoe lang (in jaren) u substantieel met deze taal in contact bent (geweest)?

Tweede taal: ____________________
- -niveau: ____________________
- -beginleeftijd: ____________________
- -contactduur: ____________________

Derde taal: ____________________
- -niveau: ____________________
- -beginleeftijd: ____________________
- -contactduur: ____________________

Vierde taal: ____________________
- -niveau: ____________________
- -beginleeftijd: ____________________
- -contactduur: ____________________

Hoe heeft u deze talen geleerd?

	School	Zelfstudie	Sociaal contact	Cursus
Taal (2,3,4):				

Anders: ____________________

Wanneer gebruikt u deze talen?

	Werk	Vakantie	Studie	Sociaal contact	Hobby
Taal (2,3,4):					

Anders: ____________________

Hoe vaak gebruikt u deze talen?

	Dagelijks	Regelmatig	Vakantie	Sporadisch	Nooit
Taal (2,3,4):					

Wanneer hoort u deze talen?

	Dagelijks	Regelmatig	Vakantie	Sporadisch	Nooit
Taal (2,3,4):					

Hoeveel uur per dag schat u dat u gesproken input krijgt via media (tv/computer/radio/...): ____________________

Hoeveel procent daarvan is, schat u, in een andere taal dan uw moedertaal? ____________________

Hartelijk dank voor uw medewerking!