



Marginal contrast in loanword phonology: Production and perception

Alexander Martin, Université Paris Cité – CNRS – Laboratoire de Linguistique Formelle, France; University of Groningen, Netherlands, alexander.martin@rug.nl

Marieke van Heugten, The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia, mhvyanheugten@gmail.com

René Kager, Utrecht Institute of Linguistics OTS, Utrecht University, Netherlands, R.W.J.Kager@uu.nl

Sharon Peperkamp, Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS), École Normale Supérieure – PSL University, France, sharon.peperkamp@ens.psl.eu

Though Dutch is usually described as lacking a voicing contrast at the velar place of articulation, due to intense language contact and heavy lexical borrowing, a contrast between /k/ and /g/ has recently been emerging. We explored the status of this contrast in Dutch speakers in both production and perception. We asked participants to produce loanwords containing a /g/ in the source language (e.g., *goal*) and found a range of productions, including a great many unadapted [g] tokens. We also tested the same speakers on their perception of the emerging [k] ~ [g] contrast and found that our participants were able to discriminate the emerging contrast well. We additionally explored the possibility that those speakers who use the new contrast more in production are also better at perceiving it, but we did not observe strong evidence of such a link. Overall, our results indicate that the adoption of the new sound is well advanced in the population we tested, but is still modulated by individual-level factors. We hold that contrasts emerging through borrowing, like other phonological contrasts, are subject to perceptual and functional constraints, and that these and other ‘marginal contrasts’ must be considered as full-fledged parts of phonology.



1. Introduction

1.1. Marginal contrast

Contrast is a central tenet of phonology. Traditionally, two sounds are said to be in a relation of contrast in a given language if they can be used to distinguish minimal pairs in the lexicon. Sounds that cannot be used to distinguish minimal pairs and whose presence is predictable from the phonological context in which they appear are said to be allophonic. But it is not always straightforward whether two sounds in a given language are in a relation of contrast or allophony. Recently, there has been a surge in interest in articulating contrast in a more fine-grained, gradient way (i.a., Ernestus, 2011; D. C. Hall & Hall, 2016; K. C. Hall, 2013; Renwick, 2014; Renwick & Ladd, 2016; Scobbie & Stuart-Smith, 2008; Stevenson & Zamuner, 2017). In particular, the concept of *marginal contrast*¹ has garnered a substantial amount of attention. Phonological contrasts are considered marginal when their distributions lie somewhere between completely separate and totally overlapping (full allophony and full contrast respectively), a scale that can be operationalized probabilistically (K. C. Hall, 2009). Such contrasts are quite common. For example, Scobbie and Stuart-Smith (2008) discuss the phonemic status of /m/ and /x/ in Scottish English. These sounds form minimal pairs (with /w/ and /k/ respectively, e.g., /wɪtʃ/ – *witch* ~ /mɪtʃ/ – *which* or /lɔx/ – *loch* ~ /lɔk/ – *lock*) for many, but not all speakers, and in some, but not all phonological contexts (e.g., only at the beginning of words for the former, but only at the end for the latter). Such minimal pairs, in addition to being subject to inter-individual variation and contextual constraints, are also relatively infrequent. This variability is measurably different from other phonological contrasts in the system (e.g., /p/~/k/) which serve to make many meaningful lexical distinctions in all contexts for all speakers (e.g., *pill* versus *kill* or *lip* versus *lick*). Similarly, contrasts that neutralize in certain contexts may be viewed as marginal. For example, in German (and many other languages), laryngeal contrasts on obstruents are neutralized in word-final position. While such contrasts are clearly part of the phonological system (they make many distinctions at the beginnings of words for all speakers, e.g., /dɔrf/ – *Dorf* – “village” ~ /tɔrf/ – *Torf* – “peat”), their lack of contrast in word-final position means that their distribution is, like the Scottish cases, partly predictable. This semi-predictability of their distribution qualifies these contrasts as ‘marginal,’ and the longstanding debate on the incomplete phonetic realization of this neutralization and its implications for representation only add to the complexity of this case (see Winter & Roettger, 2011, and references therein).

K. C. Hall (2013) laid out the first typology of such marginal relations, identifying the primary ways that can lead the distribution of contrasts to be in between full contrast and allophony. One primary way of defining contrasts as marginal is based on their distribution,

¹ Many terms have been used to describe these ‘intermediate’ contrasts (for a detailed typology, see K. C. Hall, 2013), but we will use the term ‘marginal’ throughout this article.

such that frequency differences, phonotactic restrictions, and neutralizing phonological rule application (among others) can render contrasts marginal. But marginal contrasts can also come about through mergers. For example, in Parisian French, there is an ongoing merger of mid-vowels, rendering variable the difference between words like /epe/ – *épée* – “sabre” and /epe/ – *épais* – “thick” (Fagyal, Hassa, & Ngom, 2002). Over time, such mergers may make previously full contrasts somewhat or entirely allophonic. In the case of the French mid-vowels, close-mid and open-mid variants are entirely predictable based on syllable structure in varieties where the merger is complete.²

While contrasts may become (more) marginal through mergers, marginal contrasts may also come about in cases of intensive language contact. Such cases are only touched upon briefly in K. C. Hall’s (2013) typology, but represent an important piece of the marginal contrast puzzle. When languages are in contact, lexical borrowing is one of the earliest consequences (Thomason & Kaufman, 1988). Such borrowing leads to an interesting phonological situation when words that are brought in contain sounds that do not form contrasts in the borrowing language. These sounds are usually adapted to the phonology of the borrowing language (Winford, 2005). For example, in the English loanword *backpacker*, Dutch speakers pronounce the English /æ/, absent from the Dutch inventory, with the nearest native Dutch category, /ɛ/. Although theories of loanword adaptation differ in whether they consider this process to take place at the phonetic or phonological level (see Kang, 2011; Uffmann, 2015, for reviews), they do cover most of what happens in lexical borrowing: Foreign sounds are adapted to native ones through some mapping function.

But words that are borrowed are not always (fully) adapted to the phonology of the borrowing language (Blevins, 2009; Boretzky, 1991; Cohen, 2019; Eckman & Iverson, 2015; Haugen, 1950; Itô & Mester, 1999, 2001; Kennard & Lahiri, 2020; Lee, 2013; Ussishkin & Wedel, 2003; Zuraw, O’Flynn, & Ward, 2019, among others). This can lead to the emergence of a marginal contrast in the phonology of the borrowing language. Lee (2013), for example, described how the originally context-dependent allophone [ʃ] in Korean now forms minimal pairs with native /s/. Before heavy borrowing, /s/ was produced as [s] in the context of non-high vowels and as [ʃ] in the context of high vowels (a case of full allophony). Through the introduction of words like *show* from English (which contains [ʃ] followed by a non-high vowel), the two sounds became contrastive for some word pairs. This is a situation that Hall terms “mostly predictable, but with a few contrasts” (K. C. Hall, 2013, p. 231).

² It is interesting to note that in Italian, similar mid vowel pairs also form marginal contrasts, though contrary to French, this marginalness appears to be stable diachronically (Renwick & Ladd, 2016). Inter-individual variability and low functional load need not then necessarily lead to merger, though they may be potential factors (e.g., Wedel, Kaplan, & Jackson, 2013).

One approach to understanding this type of phenomenon considers the lexicon as a hierarchical set of strata. For example, Itô and Mester's (1999, 2001) approach organizes the lexicon into strata that may be closer or further away from the native core, based on the grammatical constraints that operate on different kinds of lexical items. As the strata are hierarchically organized, certain grammatical constraints may apply to a certain subset of the lexicon while not affecting the lexicon as a whole. Cases of non-adaptation, like the emergence of /ʃ/ in words like *show* in Korean, are then treated as more peripheral than “assimilated foreign items”³ since they may contain segments or structures that are illicit in more core strata of the lexicon. The alternation between [s] and [ʃ] does not become completely inactive for Korean speakers when words like *show* are borrowed; rather, the alternation applies only to a subset of the lexicon, namely core native words, while the new contrast applies only for loanwords (for more discussion on the intermediate steps, see Simonović, 2015).

The remainder of this paper will address the psychological reality of marginal contrast. How is a contrast that is defined as marginal according to the aforementioned distributional criteria represented in speakers' minds and what effects does that have in speech production?

1.2. Cognitive constraints on contrast emergence

A stratal analysis as described above can account for observed distributional differences. This, however, is only one part of the story. What can *perception* tell us about the representation of phonological contrast? Does perception strictly track lexical organization, with more marginal contrasts being perceived less well than more core ones? Loanword non-adaptation is an ideal topic to explore such questions. Here, we will consider just such a case, where a new contrast is emerging after intensive lexical borrowing, causing two sounds which were previously completely allophonic to form contrasts between native and borrowed words. We propose that the evolution of such marginal contrasts may be dependent on the link between speakers' ability to produce and perceive the borrowed sound relative to native categories.

Evidence from perceptual studies clearly shows that listeners are sensitive to whether a pair of sounds is contrastive in their language or not (Babel & Johnson, 2010; Boomershine, Hall, Hume, & Johnson, 2008; Hume & Johnson, 2001; Whalen, Best, & Irwin, 1997, among many others). Boomershine et al. (2008), for example, demonstrated how sounds that are in a relationship of allophony are perceived as more similar to each other than sounds that are contrastive. They tested the perceived similarity of the sounds [d], [ð], and [r] in populations whose native language uses these sounds differently. Spanish speakers, for whom [d] and [ð] are allophonic, perceived these two sounds to be highly similar, but found [d] and [r], which correspond to different phonemes

³ Itô and Mester (1999) importantly note that a lexical item's history does not automatically determine its position in the lexicon, as native items may well be on the periphery while some borrowed items may be much closer to the core.

of Spanish, to be relatively dissimilar. English speakers, on the other hand, for whom [d] and [ð] correspond to two different English phonemes, perceived these two sounds to be relatively dissimilar, while [d] and [r], which are both allophones of English /d/, were perceived as much more similar. Thus, the status of a contrast has clear effects on perception.

Beyond the allophonic versus contrastive dichotomy, listeners have been shown to be sensitive to the level of distinctiveness of different contrasts, even if those contrasts would all be considered part of the core native phonology (i.e., contrastive, not allophonic). For instance, French listeners have been shown to be better at recognizing nouns whose initial consonant is mispronounced in voicing than if the mispronunciation involves manner or place of articulation (e.g., listeners can recognize the non-word *pagnole* as the real word *bagnole* – “car” [voicing difference], but have a much harder time with the non-words *dagnole* [place difference] or *vagnole* [manner difference]), indicating that manner and place of articulation are somehow more important for lexical access in French (Martin & Peperkamp, 2015). A follow-up study showed that the functional load of these contrasts could explain part of this pattern (Martin & Peperkamp, 2017). In a similar vein, K.C. Hall and Hume (2013) focussed on how listeners’ impressions of similarity vary according to the functional load of the contrast in question. They found that contrasts with a lower functional load (i.e., those that are used to distinguish fewer pairs of words in the lexicon) were perceived as more similar than contrasts with higher functional loads. An entropy-based translation of predictability of distribution in different phonological environments has also been shown to be predictive of listeners’ perceptions (K. C. Hall, Letawsky, Turner, Allen, & McMullin, 2015). That study tested the discriminability of the English fricatives [f], [s], [ʃ], and [h] in onset, intervocalic, and coda positions and in the context of different vowels. They found that sound pairs with a less predictable distribution in English were better discriminated than those with a more predictable distribution. Thus, clearly perception is reflective of the gradient nature of phonological contrast.

The studies described above focus on native contrasts, but perception may tell us even more about marginal contrasts that arise through borrowing. We will refer to this process of borrowing that results in a change to the phonology of the borrowing language as *non-adaptation*. Perception has been proposed as a key factor in determining the way words are adapted into the borrowing language (i.a., Boersma & Hamann, 2009; Davidson, 2007; de Jong & Cho, 2012; Kang, 2003; Peperkamp & Dupoux, 2003; Peperkamp, Vendelin, & Nakamura, 2008). Peperkamp and Dupoux (2003), for example, argue that ‘phonological deafness’ drives certain patterns of loanword adaptation. According to this account, French listeners cannot perceive lexical stress contrasts (since they are not present in French) and thus adapt loanwords according to how they do perceive them (namely, without lexical stress). Thus, even if French were to borrow very extensively from a language like English, it is unlikely that minimal pairs based on a stress contrast would come to exist in the language, simply because French listeners cannot perceive the stress contrasts to begin with.

But what of phonetic contrasts that exist in a borrowing language, but do not distinguish word pairs (i.e., allophones)? It has been shown that allophones may actually be easier for listeners to perceive and produce if they are presented outside of their licensing environments (Peperkamp, Pettinato, & Dupoux, 2003; Whalen et al., 1997). For instance, the French phoneme /r/ is realized as [χ] when it is adjacent to a voiceless consonant and otherwise as [ʁ]. Thus, French speakers have experience with the production and perception of both sounds. Interestingly, Peperkamp et al. (2003) found that while French listeners have difficulty perceiving the allophonic [ʁ]~[χ] contrast when the sounds precede another consonant (precisely where they alternate in the language), this difficulty disappears when they occur in final position. This could be an important factor in cases of non-adaptation: Sounds that are present allophonically in a borrowing language may be more likely to be borrowed and lead to a marginal contrast because listeners are better able to perceive them than sounds that are completely absent from their language (though perhaps only outside of their traditional licensing environment). It might be possible, then, for a contrast between [ʁ] and [χ] to emerge through borrowing in French, but this might be restricted (at least initially) to contexts where the two sounds do not alternate allophonically.

Another factor that may play a role in loanword (non-)adaptation is the abstract representation of languages' sound inventories. A non-adaptation that leads to a contrast that recombines features already present in the system (i.e., filling a gap in the inventory) adheres to the principle of *feature economy* (Clements, 2003). And indeed, if inventories tend to use featural distinctions maximally, non-adaptation might be facilitated in cases where a foreign sound recombines features (or gestures) already present in the borrowing language (Cohen, 2019; Ussishkin & Wedel, 2003). In this paper, we will consider a case study of a marginal contrast that is emerging in Dutch through heavy lexical borrowing and non-adaptation, involving two previously allophonic sounds whose representation recombines phonological features already active in the Dutch inventory. We return to the role of allophonic perception and feature economy in contrast emergence below.

1.3. Emergence of a marginal contrast in Dutch

Our case study concerns the emergence of a marginal contrast in Dutch, a West Germanic language spoken by some 22 million people mainly in the Netherlands and the Flanders region of Belgium (Lewis, Simons, & Fenning, 2016). Dutch, like languages all around the world, has borrowed extensively from other languages, including English, French, Italian, Indonesian, and many others. For example, in the 20th century alone, over 1,300 words were borrowed into Dutch from English (van der Sijs, 2002). A quick analysis of the complete loanword dictionary compiled by van der Sijs shows that these words come mostly from the domains of communication (e.g., information and communications technology, media), social life, sports, consumption, and

science, and much more rarely concern domains like religion, geography, or time. Simply put, words borrowed in the last century pertain to our modern lifestyles. Many of these words contain sounds that do not exist natively in Dutch. Our specific interest here will be the sound /g/ such as occurs in words like *gate* (borrowed from English) or *spaghetti* (borrowed from Italian).

The native Dutch stop inventory includes a voicing contrast between prevoiced and voiceless stops at both the labial and coronal places of articulation, creating minimal pairs like *paard* – “horse” ~ *baard* – “beard” and *tak* – “branch” ~ *dak* – “roof.” The voicing contrast at the velar place of articulation, however, was lost, with West Germanic */g/ spirantizing to a fricative in Old Dutch (Goossens, 1974), leaving voiceless /k/ without a voiced counterpart. The Germanic past tense prefix *ge-* is therefore produced as [χə-]⁴ in modern Dutch words like *gedaan* – “done.” The phone [g] is not completely absent from Dutch, though. In regressive voicing assimilation, underlying /k/ may be rendered as [g] before a voiced stop, such as in *i[g] ben* – “I am” (Booij 1999). This assimilation also occurs across stem boundaries in compound words like *plakband* – “adhesive tape” formed from *plak* and *band*.⁵ Dutch speakers are therefore used to producing the sound in such contexts, and also to being exposed to it auditorily, with the mapping [g] → /k/. Thus, although the phone [g] is present, it does not contrast in the native lexicon with [k]. Rather, it represents one end of a continuum that maps onto the category /k/. Given that [g] is both an allophone (thus potentially perceptible to Dutch speakers outside of its licensing environment) and that its representation recombines features already present in the inventory, it is a prime candidate for non-adaptation. What then happens when words containing this sound are borrowed into Dutch? Previous research has revealed several outcomes.

On the one hand, foreign /g/ can be adapted and pronounced with a native Dutch sound. Since Dutch shares an orthographic system with most of the languages it borrows from (namely the Latin alphabet), one such possibility is for Dutch speakers to simply produce the sound represented by the grapheme <g>, namely [χ] (or the corresponding regional fricative). Indeed, orthographic strategies of adaptation are not uncommon, even in the presence of other strategies (Daland, Oh, & Kim, 2015; van Coetsem, 1988; Vendelin & Peperkamp, 2006). Alternatively, the sound can be adapted while still approximating the way it is pronounced in the source language (thus produced with the phonetically closest Dutch sound, [k]⁶). On the other hand, these loanwords may be produced with a new sound [g], which contrasts in voicing with native [k], leading to a structural change in the phonology of Dutch. And indeed, through borrowing and non-adaptation, [g] has come to form minimal pairs with its native voiceless counterpart

⁴ The sound corresponding to the Dutch grapheme <g> has many regional variants, all of which are fricatives. We will use the symbol [χ] to represent these sounds and will not discuss this variation further here.

⁵ While this assimilation is sometimes presented as being complete and categorical, acoustic evidence shows that it is in fact a low-level gradient pattern that does not yield the voiced phone systematically (Jansen, 2007).

⁶ This is the closest Dutch sound to any velar stop, be it prevoiced or not.

[k]: *kool* – “cabbage” ~ *goal* – “goal (sports)”; *manco* – “deficiency” ~ *mango* – “mango”; *keet* – “shack” ~ *gate* – “gate (at an airport)”; *lekkend* – “to leak” ~ *laggen*⁷ – “(for a computer) to lag.”

It is noteworthy that these minimal pairs follow the same phonetic pattern as the voicing contrast that distinguishes *paard* from *baard* and *tak* from *dak*: namely prevoicing versus short-lag VOT. Dutch voiced stops have been reported to have a mean VOT of -85 ms for labials and -80 ms for coronals, while the corresponding voiceless stops have VOTs of 10 ms and 15 ms, respectively (Lisker & Abramson, 1964). When Dutch speakers contrast the native word *kool* (mean VOT of 25 ms) with the loanword *goal* (producing prevoicing on the /g/), they are implementing this new voicing contrast within the Dutch phonetic space. As we will see, this implementation is not sensitive to the phonetics of the voicing contrast of the source language. Indeed, in English, the initial /g/ is often produced with short-lag VOT, thus very close to Dutch /k/. Minimal pairs like *kool*~*goal* highlight how this contrast has become meaningful in the Dutch phonological system, beyond reference to a phonetic contrast in a source language.

The production of this new phoneme /g/ in Dutch has received some attention in three previous studies (Hamann & de Jonge, 2015; van Bezooijen & Gerritsen, 1994; Van de Velde & van Hout, 2002). In two of these, /g/ was considered as part of greater surveys concerning the pronunciation of loanwords in Dutch, and specifically the realization of various sounds that are not distinctive natively in the language. First, van Bezooijen and Gerritsen (1994) examined the pronunciation of loanwords by asking 75 women from various regions of the Netherlands and Belgium to read them aloud. Two of the words contained /g/: *goal* and *drugs*. The two differ though, as the latter is subject to final devoicing; the production of a voiced stop in such a case would thus involve an extra degree of non-nativeness. They found evidence for all three phonetic realizations of /g/ for these words: [g], [k], and [χ], though [g] occurred most often, at least in *goal* (49% of tokens of this word). Van de Velde and van Hout (2002) surveyed a wider range of regional populations in both the Netherlands and Belgium. They asked school teachers from four regions in the Netherlands and four regions in Belgium to read a list of loanwords aloud according to standard Dutch pronunciation. Their list included a few more words with /g/ (*goal*, *guillotine*, *goulash*, *buggy*, *mango*), and they found similar results to van Bezooijen and Gerritsen (1994), with unadapted [g] being rather common, but with the other two variants also being used. The third study used an adaptation of the map task (Brown, Anderson, Shillcock, & Yule, 1985), which the authors called the ‘menu task,’ to elicit spontaneous productions of three food-related loanwords containing /g/ (*gorgonzola*, *spaghetti*, and *mango*) by an age-stratified sample of dyads (Hamann & de Jonge, 2015). Again, there was clear variation in the pronunciation of the words. In addition, they noticed slight differences in the way the younger and older groups pronounced *spaghetti*. Younger people in their sample realized this word with

⁷ Note, as mentioned above for the word *backpacker*, that the English /æ/ is adapted to Dutch /ɛ/.

a fricative (following the orthographic pronunciation) more often than older individuals. The authors suggest word frequency and word age as possible explanatory factors for their findings.

While clearly establishing that there is great variation in the (non-)adaptation of loanwords containing /g/ and showing strong evidence for the use of the sound [g] outside of its usual allophonic context, none of these studies considered listeners' ability to perceive the emerging contrast. As discussed above, perception plays a key role in loanword (non-)adaptation. In the present case, it is therefore important to consider individuals' ability to perceive the new sound [g] as different from their native [k], and to test if this is related to the way they produce the critical loanwords. Are the speakers that produce the most [g] tokens also the ones who can perceive the contrast between [k] and [g] best? Previous work has shown that the relation between perception and production at an individual level may follow complex trajectories (e.g., Pinget, Kager, & Van de Velde, 2020), and consideration of perception alongside production is thus a central concern.

To address this issue, this paper details an experimental investigation of Dutch /g/ usage effects in production and in perception, and explores the relation between the two. Focussing on university students, we assess the production and perception of /g/ by means of an elicitation and a discrimination task respectively; this allows us to consider the relation between individual speakers' productions and their ability to perceive this new contrast. To preview, we find that while there is a range of variability in production, participants' perception of the emerging contrast is very good, though not quite as good as of a contrast that is established in Dutch. We also explore the idea that speakers who produce more [g] tokens also tend to be better at perceiving the emerging [k]~[g] contrast, but our data do not provide strong support for this.

2. Method

Our experimental investigation relies on a methodology which compares performance within a single group of participants. Specifically, we test production and perception in the same individuals. We used a production task (involving sentence and word list reading) and a perception task (comparing perception of an established contrast to the emerging one between [k] and [g]). We first measured /g/ production variability in a way that allowed us to explore whether there is a link between the phonetic realization of the sound and speakers' phonological representation of it. Additionally, we measured the variability in participants' *perception* of the emerging [k]~[g] contrast. Finally, we tested if, at the level of the individual, listeners' ability to perceive the contrast correlates with their personal use of the unadapted variant [g] in production.

All stimuli and anonymized data⁸ along with a Python analysis notebook are available for download at the following OSF repository: <https://osf.io/ec4yq/>.

⁸ Only coded data is available as we do not have ethics approval to share the recordings from our production task.

2.1. Participants

We recruited 55 participants at Utrecht University (43 women, 12 men, mean age 23.2 ± 5.5 years). Utrecht attracts students from various parts of the language area, making it an ideal testing site. We recruited participants with a range of regional backgrounds, with an attempt to draw equally from more northerly and more southerly regions, specifically north and south of the *Grote Rivieren*⁹ dialect boundary (Donaldson, 1983). The hometowns of our participants are visualized in **Figure 1**, with Utrecht indicated by a red cross.



Figure 1: Hometowns of our participants in the Dutch language area of Europe (the Netherlands and Flanders in Belgium). The size of points is relative to the number of participants from that town. Our testing site, Utrecht, is shown as a red cross and the broken line traces the *Grote Rivieren* boundary.

⁹ The *Grote Rivieren*, or *Great Rivers*, form a large convergence as they merge into the Rhine and Meuse deltas. This division has been of cultural importance in the history of the Netherlands, traditionally separating the Protestant north from the Catholic south.

Four participants were excluded from data analysis: two who reported English being spoken at home growing up, one whose parents were both born outside of the Dutch language area of Europe (participants who had at least one Dutch- or Belgian-born parent were not excluded), and one due to experimenter error leading to the loss of the participant's recordings. The remaining 51 participants all reported Dutch as the main language of the household when they were growing up.

On a composite scale including self-reported comprehension ability and use, all participants reported good command and regular use of English. Many participants also reported good knowledge of French and/or German, but knowledge of other languages our target words were borrowed from (Italian, Indonesian, and Hungarian) was rarely reported.

2.2. Production

We conducted a sentence and word list reading task. Sentence-reading seemed to us to be the best way to induce participants to produce the words (given their domain-specific usage), while preserving some amount of ecological validity (i.e., without explicitly revealing what we were looking for). We were careful to have a high ratio of filler items to target words (4 : 1), as detailed below. We additionally included word list reading, as this allowed certain tokens to appear in a different phonological context (e.g., without a preceding vowel).

2.2.1. Stimuli and procedure

Sixty sentences were prepared containing loanwords that have been borrowed into Dutch from various source languages (English, Italian, German, French, Indonesian, and Hungarian). Twelve of the sentences were target sentences that contained words with /g/ as syllable onsets, in word-initial (e.g., *guillotine*) or in word-medial position (e.g., *schlager*) in their source language (see Appendix A).¹⁰ Two of these sentences contained words with two occurrences of /g/: *Google* and *Go Ahead Eagles* (a Dutch soccer team), such that there was a total of 14 target /g/s. The 48 distractor sentences also contained loanwords, but without /g/ in the source language, so that the target sentences would not stand out. During debriefing, many participants remarked on the fact that the sentences contained loanwords; none mentioned noticing the /g/ focus of our experiment. Participants were seated in a soundproof booth and wore a Shure SM10A dynamic head-mounted microphone while stimuli were presented on a computer screen positioned approximately at eye-level. After reading the sentences, participants read the individual loanwords extracted from the sentences, presented one at a time. All targets were therefore read in two contexts: embedded within a sentence and in isolation.

¹⁰ Note that the words in our study were well-established loanwords, known to our participants. For example, they are subject to Dutch morphology (e.g., *ik heb het al gegoogeld* – “I have already Googled it”).

Recordings were automatically controlled by a MATLAB script, beginning when the sentence or word was displayed and ending when the participant stopped speaking. After each item, participants were asked either to press a “continue” button or to rerecord the previous item, if for example they had stumbled while reading, or the recording had been cut off prematurely. If participants rerecorded their utterance, we only analyzed their final production.

2.3. Perception

Directly following the production task, we tested listeners’ perception of the emerging contrast. Recall that the production task contained a large proportion of fillers to targets in order to prevent participants from realizing which segments we were interested in. As a perception task will necessarily draw attention to the contrast of interest, we preferred to test participants’ perception only after they had completed the production task. We compared participants’ perception of the emerging contrast to their perception of the established [p]~[b] contrast, which differs along the same phonetic dimension (i.e., voicing) using the ABX discrimination task. In the ABX task, participants hear two different stimuli A, and B, and then a third stimulus X, and must determine whether X is a token of the same category as A or as B. This task requires participants to hold stimuli in memory and thus encourages a level of phonological abstraction (particularly since we used different voices for the A, B, and X stimuli, creating a great amount of acoustic variation, see below). Participants are required to rely on their abstractions to perform the task, leading to high levels of categorical perception (Gerrits & Schouten, 2004). This should allow us to see if [g] tends to be perceived as different from [k] (the contrast thus forming two distinctive categories) or if it is perceptually assimilated to /k/. None of the previous studies detailed in the introduction considered listeners’ ability to perceive the emerging sound as different from a native one. Yet, listeners’ perception is integral to our understanding of the evolution of emergent phonological structure. If only a small subset of speakers are able to correctly perceive the emerging contrast, there is little hope of it taking hold at a population level.

2.3.1. Stimuli

We created a set of 20 phonotactically legal trisyllabic ‘CVC \emptyset CVC pseudoword frames (where the bold C represents the consonant being tested, drawn from either the [p]~[b] contrast or the [k]~[g] contrast). The initial consonant was drawn from the set {r, h, m, v, f, j, n, s} and the first vowel from the set {i, o, u, ε, ε̄, œ̄, α, y, ø, ɪ, ɔ}. The contrast being tested was always on the second syllable, which was unstressed and contained a schwa, and crucially was not an environment for allophonic [g]. Following Peperkamp and Bouchon (2011), all items ended with real Dutch suffixes (-aar, -ig, -ing, -lijk, -loos).¹¹ This was done to make the stimuli more word-like.

¹¹ For vowel-initial suffixes, a sonorant consonant from the same set as the initial consonants was added to the beginning of the suffix for each frame; this yielded frames like [RiC \emptyset Ra:R] and [muC \emptyset na:R] for the real Dutch suffix -aar.

These 20 frames yielded a total of 80 items (one for each of the critical test segments: [p], [b], [k], [g]; see Appendix B for the full item list). For example, the frame [ho:Cəna:r] yielded the stimuli [ho:pəna:r], [ho:bəna:r], [ho:kəna:r], and [ho:gəna:r]. Each item was recorded by three native speakers of Dutch, all from the Randstad area of the Netherlands, two women, one man (and none having difficulty producing [g]), to be used as the A, B, and X tokens, respectively. This means that the X token was acoustically quite distinct from A and B. Thus, participants could not rely on acoustic memory alone in order to complete the task; some level of abstraction was necessary. A further five frames of the same form (20 total items) were created and used as training stimuli. All tokens were recorded to a Lambda Lexicon external sound card.

From the stimulus frames, we created trials where A and B mismatched only in the voicing of the second consonant. This therefore led to two types of trials: those that tested the emerging [k]~[g] contrast, and those that tested the established [p]~[b] contrast.

2.3.2. Procedure

Participants were informed orally that all instructions for the task would be given on the computer screen. They were seated in a sound-attenuated booth facing a screen at approximately eye level. They were informed they would hear series of three made-up words, the first and second of which were always different, and the third of which was the same as either the first or the second. They were told to hold a button box in their hands and respond with their thumbs by pressing the left button if the third word matched the first, or the right button if it matched the second.

Each trial consisted of the presentation of three stimuli (A, B, and X) produced by the two female speakers and the male speaker, respectively, with an interstimulus interval of 250 ms. In half of the trials, X matched A and in half of the trials X matched B; trial order was randomized for each participant. Participants were given 3,500 ms to respond, after which the trial ended. The next trial began after a delay of 1,000 ms following the button press (or timeout).

Participants began the task by completing a 20-trial training block (which consisted of four trials for each of five training frames, two per contrast), during which they were told if their responses were correct or incorrect. In the event of an incorrect response, the trial was repeated. The test phase consisted of 160 trials that did not include items from the training phase. Each consonant frame yielded four trials per contrast (ABA, ABB, BAA, BAB), such that each individual stimulus item was heard six times, twice by each speaker. During the test phase, no feedback was given. Participants were told both before the task and reminded after the training phase to pay close attention to the words and to keep the first two words in mind, in order to be able to compare them with the third. Halfway through the experiment, participants were offered the opportunity to take a short, self-timed break. The entire experiment lasted approximately 20 minutes.

At the end of the testing session, participants completed an exit questionnaire concerning demographic information and language experience.

3. Results

3.1. Production

Participants' productions were segmented and coded manually offline by the first author. Each target segment was identified by examination of the waveform and spectrogram in Praat (Boersma & Weenink, 2020). Tokens were coded as stops if they had a visible release burst, otherwise they were coded as fricatives. Stops were coded as voiced if they had a visible voicing bar during closure, and voiceless if they did not. Voiced stops were segmented from onset of voicing to release burst; voiceless stops were segmented from release burst to onset of voicing in the following vowel. VOT durations were extracted from this segmentation using a Python script and measured relative to the release burst, such that prevoiced tokens (those with a visible voicing bar during closure) had negative VOTs and voiceless tokens (those where voicing began after the release burst) had positive VOTs. Examples are shown in **Figure 2**.

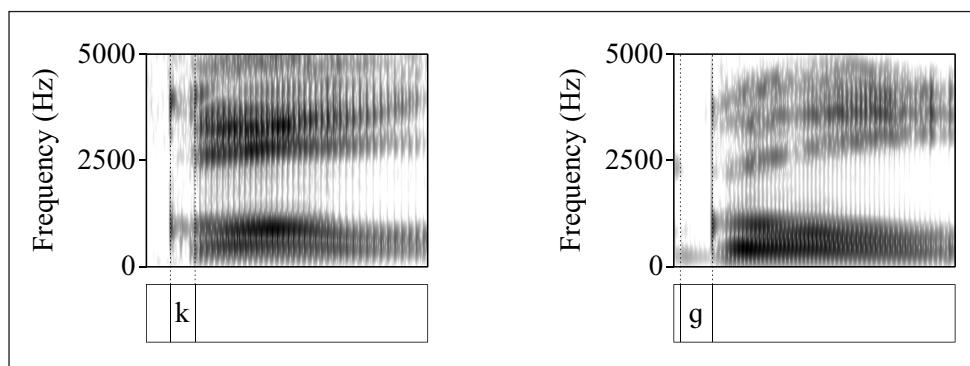


Figure 2: Spectrograms of two tokens of the word *goal* produced in isolation, one produced with [k] (left) and one produced with [g] (right). The token on the left shows that the VOT for the phone [k] was measured from the release burst to the onset of voicing in the following vowel. The token on the right shows that the VOT for the phone [g] was measured from the beginning of visible prevoicing up to the release burst.

Occasionally, target segments were produced in an ambiguous way. Per our coding scheme, segments with no identifiable release burst were categorized as fricatives, though a certain number of these did not impressionistically correspond to any variant of the fricative representing the grapheme <g>. Lenition of underlyingly voiceless obstruents has been documented in spontaneous Dutch speech (Schuppler, Ernestus, Scharenborg, & Boves, 2011), and these tokens could, in principle, correspond to underlying /g/ or /k/. Since it was not possible to objectively determine which category these ambiguous tokens belonged to, they were not included in our analysis; this concerned around 11% of the data points ($N = 148$). A total of 1,264 tokens were thus retained for analysis, 617 from the sentences and 647 from the word list, with a total of 700 [g] tokens, 436 [k] tokens, and 128 [χ] tokens.

We observed a wide range of inter-individual variability in the proportion of [g] produced (see **Figure 3**), with some participants nearly always producing [g], and others rarely producing it. We analyzed these data using logistic mixed-effects models,¹² with fixed effects being deviation coded. We first created a full model that included the fixed effect of Context (word list versus sentence reading) with random intercepts for Participant and Item, each including a random slope for Context (this was the maximal model according to our design). We then created a simpler model that excluded the factor Context and compared the simpler model to the full model using a likelihood ratio test. The full model was found to explain significantly more variance than the reduced model ($\beta = -0.81$, 95% CI = [-1.53 : -0.12]; $\chi^2(1) = 5.13$, $p < 0.05$). This indicates that participants were more likely to produce [g] if the token was produced within a sentence than if it was produced in isolation.

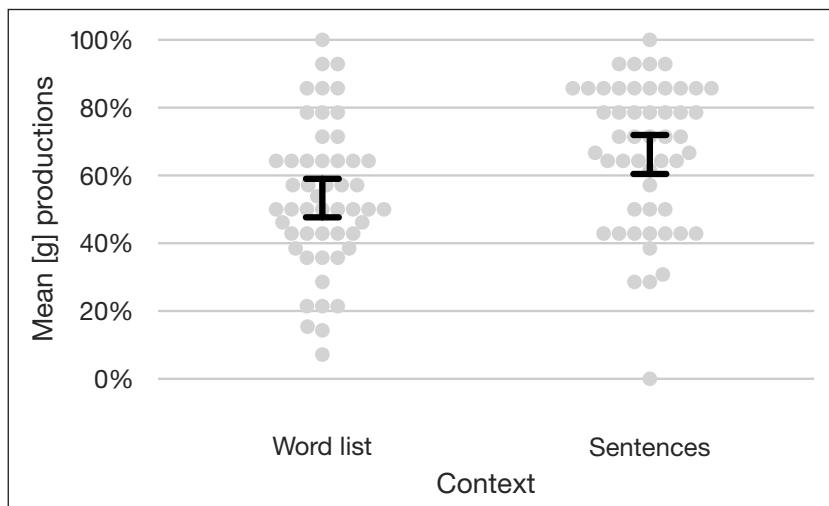


Figure 3: Average percentage [g] production by context. Each dot represents a participant and error bars show a 95% confidence interval calculated over participant means.

In addition to whether or not each token was produced in isolation or embedded in a sentence, some target segments appeared in an intervocalic position (potentially easing the articulation of a voiced compared to a voiceless stop). Some of these always appeared in intervocalic position (e.g., when they were word internal like in the word *spaghetti*), while others were in intervocalic position only in the sentence reading (e.g., when they were preceded by a vowel-final word like in the phrase *redelijke goulashsoep*). A summary of the frequency of each token type by phonological context is shown in **Table 1**. To test the influence of phonological context on [g] versus other production, we designed a logistic mixed-effects model with the fixed factor Intervocalic (deviation

¹² All analyses were implemented in R, and mixed models were designed using the lme4 package (Bates, 2014).

coded) and random intercepts for Participant and Item (including a random slope for Intervocalic under Participant, the maximal structure), and compared it to a simpler model excluding the factor Intervocalic using a likelihood ratio test. We combined the sentence and word list data in order to increase statistical power¹³ and found that the full model was a significantly better fit than the reduced model ($\beta = 0.89$, 95% CI = [0.39 : 1.40], $\chi^2(1) = 12.12$, $p < 0.001$), indicating that intervocalic contexts did appear to facilitate the production of [g].

	intervocalic	non-intervocalic
g	321	379
k	140	296
χ	102	26

Table 1: Frequency of token types by phonological context.

Next, we considered the VOTs of stop tokens. VOT was measured as the beginning of a visible voicing bar relative to the stop release burst. VOTs for all stop tokens (i.e., both [g] and [k] tokens) are visualized in **Figure 4**. Tokens with a positive VOT correspond to those tokens coded as [k], with a peak around the average VOT for native Dutch /k/ (25 ms, per Lisker & Abramson, 1964). Our [k] versus [g] coding thus seems to properly reflect the bimodal distribution of the stop tokens' VOTs in our dataset, and further follows the pattern reported in Lisker and Abramson (1964) with the voiced category showing a wider range than the voiceless category.

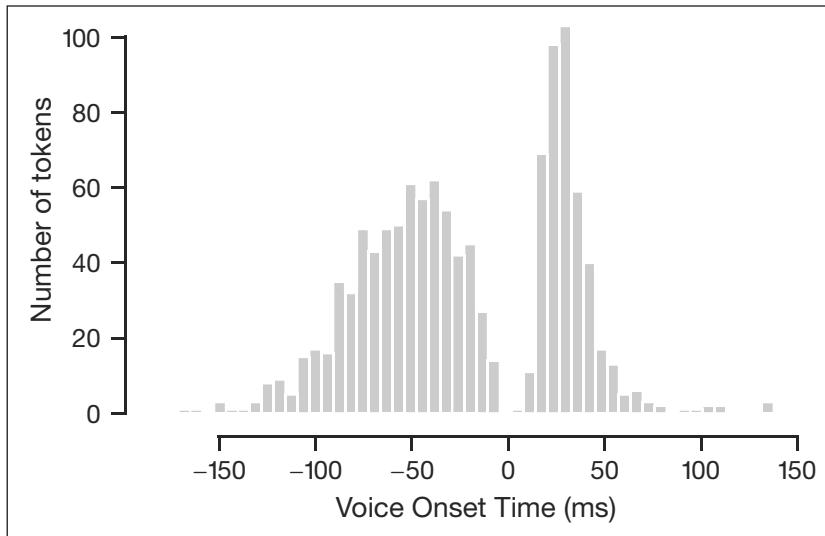


Figure 4: Distribution of VOT for all velar stop tokens in our production task.

¹³ A model including Context as an additional factor failed to converge.

To test whether there is a relation between the amount of [g] usage (a representational question) and the phonetic realization of the sound (an implementational question), we measured VOTs of the [g] stop tokens in individual participants and compared this with their average [g] usage. A total of 700 tokens (again, combining the sentence and word list data) were coded as containing [g] and were therefore eligible for VOT analysis. We examined whether speakers who produce longer prevoicing on [g] tokens (thus a heightened contrast between native [k] and emerging [g]) are more likely to use [g] in general. We did this by performing a regression on the proportion of average [g] production per participant as a function of their average [g] VOT. This regression is shown in **Figure 5**. We indeed found that participants with longer average prevoicing on [g] tokens also produced *more* [g] tokens (adjusted $R^2 = 0.12$, $t = -2.82$, $p < 0.01$).

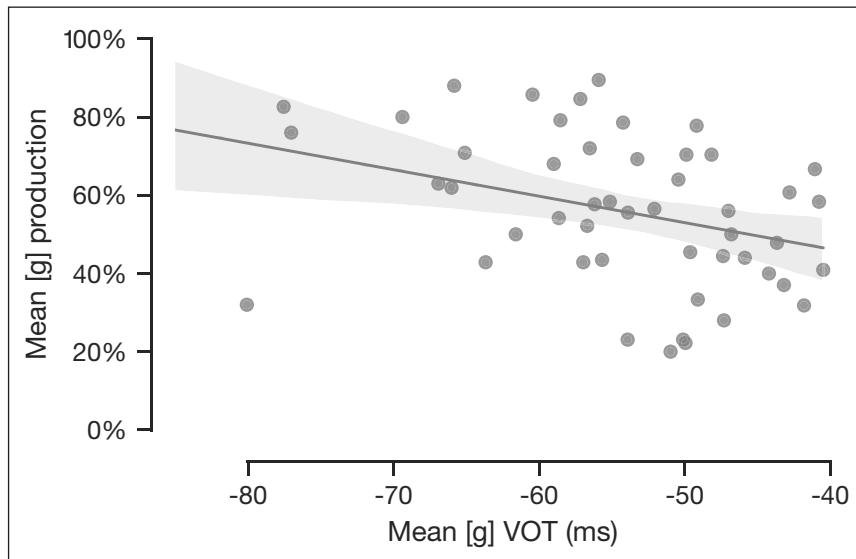


Figure 5: Participants' average [g] production as a function of their average VOT for [g] tokens with a fitted regression line and the 95% CI around that line.

3.2. Perception

Having considered our results in production where we observed varying rates of [g] usage by participant, we turn to the results of our perception task to see if we observe similar variability. Despite the seemingly difficult design of the experiment (three different voices pronouncing long stimuli, thus a wide range of acoustics; crucial contrast in an unstressed syllable containing schwa), performance was overall very high and variability relatively low. We set an *a priori* inclusion criterion such that participants who performed below 80% accuracy on the established contrast were excluded from data analysis ($N = 5$). This concerned three participants who responded with only one of the buttons, one participant with low overall accuracy (70.1%), and one participant whose accuracy indicated they may have inverted the response buttons (27.2%).

Average accuracy for the remaining 46 participants on the established contrast was 94.5% ($SD = 22.8\%$), and 91.3% ($SD = 28.2\%$) on the emerging contrast. Participants' performance in the two conditions is plotted in **Figure 6**.

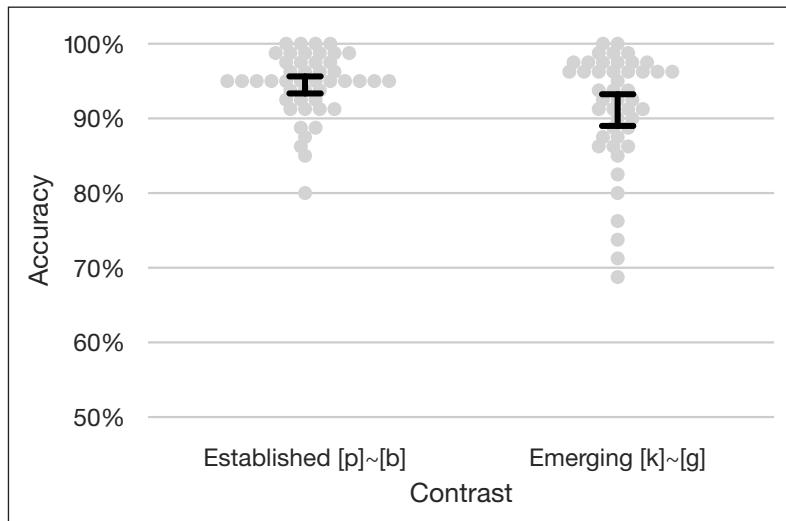


Figure 6: Participants' average accuracy in the ABX task by contrast, plotted from chance level. Each dot represents a participant mean and error bars represent 95% confidence intervals calculated on those means.

We analyzed these data using logistic mixed-effects models, with fixed effects being deviation coded. We first created a full model that included the fixed effect of Contrast (established versus emerging) with random intercepts for Participant and Frame, each including a random slope for Contrast (this was the maximal model according to our design). We then created a simpler model that excluded the factor Contrast and compared it to the full model using a likelihood ratio test. This model was found to predict significantly less variance than the full model that included the factor Contrast ($\beta = -0.41$, 95% CI = $[-0.81 : -0.01]$, $\chi^2(1) = 3.99$, $p < 0.05$), indicating that the established contrast was perceived better than the emerging contrast.

We additionally conducted a follow-up analysis based on Best, McRoberts, and Goodell's (2001) consideration of recency.¹⁴ The ABX task includes two types of trials: 1) those where X matches the initial stimulus (i.e., ABA trials) and 2) those where X matches the most *recent* stimulus (i.e., ABB trials). Previous work has suggested that participants in the ABX task tend to have a small bias to answer X = B (Schouten, Gerrits, & van Hessen, 2003). One potential explanation for this bias is that if participants have difficulty discriminating the critical sounds, then all three stimuli sound similar. This might lead participants to think that the last two stimuli

¹⁴ We would like to thank an anonymous reviewer for this recommendation.

in particular sound the same, if they have lost the auditory trace of the A stimulus. This would cause them to answer X = B more often than X = A, leading to ceiling performance on ABB trials, and lower accuracy on ABA trials. We therefore conducted a post hoc analysis splitting our data by trial type. This split is visualized in **Figure 7**, showing that the difference between the established and emerging contrast is greater on ABA trials than it is on ABB trials, though performance is overall still very high and variability low.

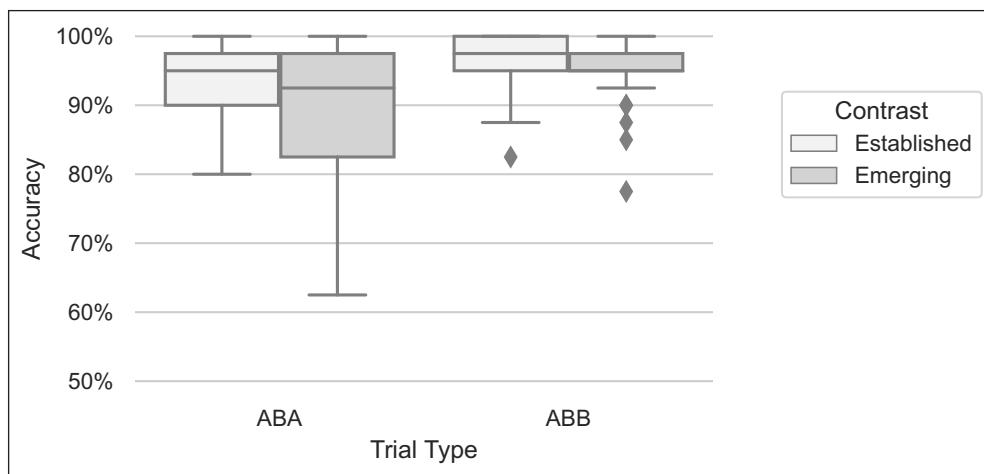


Figure 7: Participants' average accuracy in the ABX task by contrast, separated by trial type, plotted from chance level.

To explore the possible difference between ABA and ABB trials, we re-analyzed the data including an additional deviation coded factor in our logistic regression model. Thus, we first created a full model that included the fixed effect of Contrast (established versus emerging) and a fixed effect of Trial Type (ABA versus ABB) and the interaction between these factors, along with random intercepts for Participant and Frame, and including a random slope for Contrast within Participant. This was the maximal random effects structure that converged. We then created simpler models that excluded one of the factors or the interaction and compared them to the full model using likelihood ratio tests. A model excluding Contrast was again found to predict significantly less variance than the full model ($\beta = 0.40$, 95% CI = $[-0.65 : -0.16]$, $\chi^2(1) = 9.88$, $p < 0.01$), indicating that the established contrast was perceived better than the emerging contrast. A model excluding Trial Type was also found to predict significantly less variance than the full model ($\beta = 0.68$, 95% CI = $[0.52 : 0.84]$, $\chi^2(1) = 75.38$, $p < 0.001$), indicating that performance was better on ABB trials than ABA trials, in line with a response bias for B. We also found that a model excluding the interaction predicted significantly less variance than the full model ($\beta = 0.33$, 95% CI = $[0.02 : 0.64]$, $\chi^2(1) = 4.25$, $p < 0.05$), indicating that the difference between the established and emerging contrasts was greater on ABA trials than on ABB trials.

3.3. Production-perception link

In order to investigate the association between participants' production and their perception, we first established a by-participant score for each of the tasks. For the production task, we took the average proportion of [g] use (unadapted compared to adapted variants), collapsing across production contexts (word list and sentence reading).

For the ABX task, we require a measure of how well each participant can discriminate the emerging contrast. Taking the raw accuracy on the emerging contrast fails to take into account task-related issues (e.g., attention). It is therefore necessary to consider performance on the emerging contrast *relative* to performance on the established contrast, which represents baseline performance in the task. A relative measure also makes comparison amongst participants more meaningful, since some listeners will be better at contrast perception generally. This is often done using a difference score, where participants' average performances in one condition are subtracted from their average performance in the other. However, recent work has demonstrated that a method using the residuals extracted from a linear regression between the two conditions is a more robust measure (DeGutis, Wilmer, Mercado, & Cohan, 2013). We followed the method described in that study by performing a linear regression on the ABX accuracy data, using performance on the established contrast as a predictor for performance on the emerging contrast. From this regression, we extracted the residuals (a given point's deviation from the fitted regression line), with lower values representing worse discrimination of the emerging [k]~[g] contrast relative to the established [p]~[b] contrast. We then used these residuals in a regression predicting the average proportion of [g] production by participant (that is, we used the by-participant perceptual score to predict the proportion of by-participant [g] production); see **Figure 8**.

The residuals from the ABX task regression were found to be a significant predictor of the proportion of [g] production (adjusted $R^2 = 0.12$, $t = 2.72$, $p < 0.01$), indicating that participants who perceive the emerging contrast less robustly are less likely to produce [g] in loanwords, and instead tend to adapt it as one of the native sounds [k] and [χ], while participants who perceive the emerging contrast more robustly are, on average, more likely to produce [g]. We thus observed some evidence of a link between perception and production. Note though that the observed effect appears to be driven by a couple of participants, with most participants showing no discernible link between their production and perception scores, despite the relatively small amount of variability we observed in our perceptual results compared to our production results.

Given the results from the follow-up analysis of our ABX task, we additionally performed an analysis that used residuals from a regression that looked only at performance on ABA trials (the more difficult trials, where we observed a significantly stronger difference between the

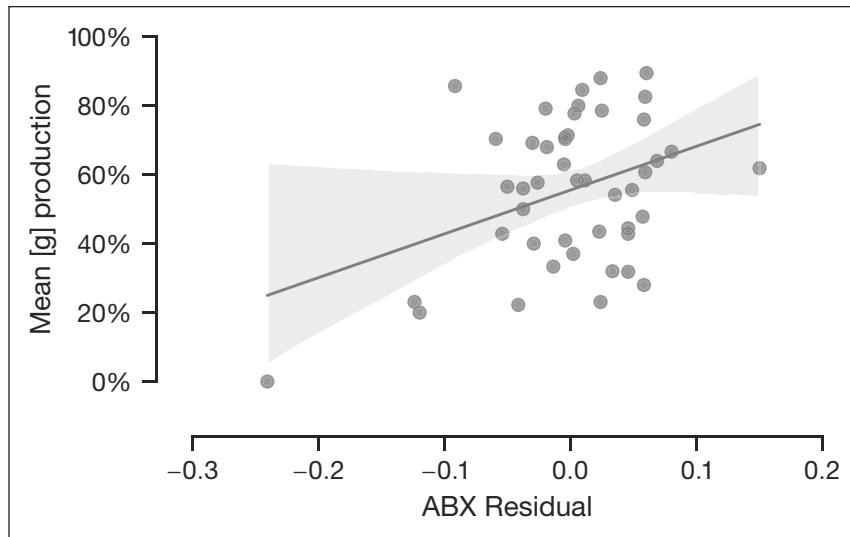


Figure 8: Average [g] production by participant (both contexts combined) as a function of the residuals extracted from the ABX task regression with a fitted regression line and the 95% CI around that line.

established and emerging contrasts than on ABB trials). The relation between ABA residuals and average [g] production is visualized in **Figure 9**. We similarly found that the ABA residuals were a significant predictor of [g] production (adjusted $R^2 = 0.12$, $t = 2.66$, $p < 0.05$) though again, this appears to be driven mainly by the same couple of participants.

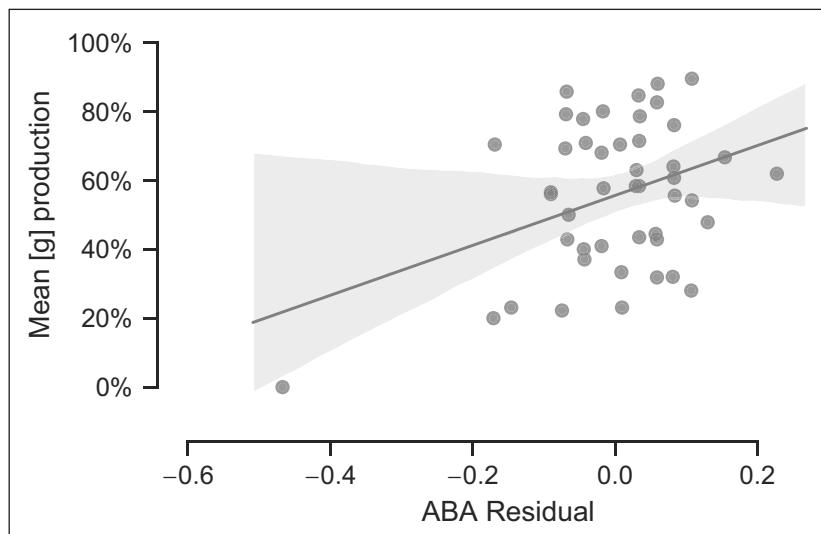


Figure 9: Average [g] production by participant (both contexts combined) as a function of the residuals extracted of the ABX task regression for ABA trials only, with a fitted regression line and the 95% CI around that line.

4. Discussion

Recent work has highlighted the myriad types of phonological relations that are ill-captured by a categorical distinction between full contrast and allophony (e.g., K. C. Hall, 2013), and indeed, understanding the various aspects of *marginal contrasts* (i.e., going beyond lexical distribution) is key to our understanding of phonology. Here, we have focused on a case of marginal contrast in Dutch: the emergence of contrast between /k/ and /g/ through extensive borrowing from languages like English, French, and Italian. In a production task requiring participants to read out sentences and individual loanwords extracted from those sentences, we found a wide range of variability in the amount of [g] productions relative to the native equivalent [k], or an adaptation in line with the orthography [χ]. Additionally, we observed a positive correlation between the amount of unadapted [g] use compared to use of an adapted phone and how strong of a phonetic contrast participants made between [g] and native [k].¹⁵ The same participants took part in an ABX discrimination task that assessed their ability to perceive the emerging [k]~[g] contrast relative to the established [p]~[b] contrast. Overall, participants were able to distinguish the contrasts very well, though group performance was better for the established contrast than for the emerging one. We then considered the relationship between individuals' own productions and their perception. We found some potential evidence for a link between the two, with individuals who produce more [g] (i.e., less adaptation) also perceiving the emerging contrast slightly better than those who produce less [g], though the effect appeared to be driven by only a couple of participants. Below, we discuss in turn our production results—including lexical and social factors—and our perception results—including factors that favour perception of the emerging contrast. We conclude with a discussion of the implications of marginal contrasts emerging through borrowing on phonological representation.

4.1. Production

Let us first consider our production data. Participants produced three types of tokens in loanwords which contain /g/ in their source language. Some tokens were produced with the fricative [χ], which corresponds to the orthography: the Dutch letter <g> is pronounced as a fricative. Others were produced with stops, either the established voiceless [k] or the emerging voiced [g]. We posited [g] productions to represent non-adaptation while [k] tokens were posited to be adapted to the native phonology. But, before we settle down on this notion, we have to consider an alternative interpretation. Could these [k] tokens simply be faithful, English-like productions, at least for the tokens occurring in English borrowings? After all, the <g> in English *goal* is produced without

¹⁵ Of course, we are unable on the basis of these data alone to determine whether it is the fact that these participants produce more [g] than they are producing ‘better’ tokens (*practice makes perfect*), or if their ability to produce longer prevoicing modulates their willingness to produce [g] at all (*confidence is key*).

prevoicing by native English speakers.¹⁶ However, this alternative interpretation is unlikely to fully account for our production data. While the speakers we tested were all proficient in English, considering [k] as a faithful production (thus not as an adaptation) requires speakers to have not only in-depth knowledge of English phonetics, but also in-depth knowledge of the phonetics of other source languages (e.g., how Italian native speakers produce *spaghetti*). Speakers would have to know to produce [k] for loanwords from English or German and [g] for loanwords from French or Italian. Furthermore, if the target words were being produced as phonetically-faithful code-switches, we should expect no phonological integration at all (Poplack, Robillard, Dion, & Paolillo, 2020); clear and consistent final devoicing on items like *Go Ahead Eagles* in our data discounts this possibility. Finally, if such loanwords were to be produced entirely faithfully, then we should expect (some) Dutch speakers to produce words like *computer* with aspiration on the initial /k/, in line with the English pronunciation. But lack of aspiration on initial stops in stressed syllables is a well known feature of Dutch L2 speakers of English (Collins & Mees, 2003) and we are unaware of such productions on loanwords in Dutch. For these reasons, we considered [k] tokens to be productions adapted to the native phonology.

It seems that the phonetic realization of the crucial sound in the source language has a limited effect compared to its position in the phonological system of that language, and specifically how it contrasts with other sounds. This is in line with accounts of structural borrowing in cases of intensive language contact (like borrowing of a phoneme) that propose that such borrowing is initially restricted to a small group of highly bilingual individuals before spreading through the community (e.g., Thomason & Kaufman, 1988; Weinreich, 1953). The abundance of [g] tokens we observed suggests that speakers know that /g/ contrasts with their native /k/ in a specific way (i.e., voicing), or at the very least that this was true of the bilingual speakers who initially borrowed these words. The prevoiced [g] is different from the way /g/ is usually produced in English (though is closer to the way it is produced in other languages Dutch has borrowed from, such as French or Italian), and the contrast that it forms with [k] for our participants is thus specifically relevant to Dutch. Since Dutch speakers use a native phonetic contrast (i.e., prevoicing versus short lag VOT, in line with other stop voicing contrasts in their language), they rather implement the source language phonological contrast in a way that is meaningful within the Dutch phonetic space, even though it might not strictly correspond to the phonetic realization in the source language. We may consider how this Dutch phonetic contrast came about through borrowing from languages like English (where the contrast is usually realized as a difference in the presence versus absence of aspiration). One possibility is that the speakers who brought in the loanwords originally were aware of the abstract contrast between /g/ and /k/ (in English or other source languages) and implemented the contrast in a way that is meaningful in

¹⁶ Recall that Dutch uses prevoicing to distinguish voiced from voiceless stops rather than the presence or absence of aspiration.

Dutch (in analogy with other stop voicing contrasts). For example, a Dutch speaker of English might know that the English word *coast* – /koust/ – [k^houst] contrasts with the English word *ghost* – /goust/ – [koust] and thus that there is a contrast between /k/ and /g/. If they also know that words like *post* and *boast* are similarly contrasted by the presence or absence of aspiration, and that the aspirated categories map onto their native voiceless categories in Dutch while the unaspirated categories map onto their voiced categories, the jump to a voiced velar phone is reasonably straightforward. Additionally, the presence of [k]~[g] allophony in Dutch might allow other listeners to more accurately perceive this new contrast when the borrowing speaker uses it (we return to this point below when discussing our perception results). Naturally, this whole process may be aided by orthography, since Dutch speakers of English are invariably literate in both languages, and attention is drawn to this contrast in English language teaching in the Netherlands.

One aspect of lexical borrowing that we have yet to address is lexical specificities that might affect the way individual words are adapted (e.g., some words may be systematically produced with the unadapted variant, while other words may display inter- or even intra-speaker variability). For example, Lev-Ari and Peperkamp (2014) mention the Arabic sound [w] in loanwords borrowed into Hebrew (with Hebrew lacking [w] natively); some words are systematically borrowed with an adapted sound [v], while others maintain the Arabic pronunciation. Any number of word-specific properties could drive such differences. Crawford (2007) suggests that the type frequency of emerging phonological structures may affect their (non-)adaptation, taking as an example the previously illicit and now often unadapted coronal + [i] sequences in Japanese (e.g., in English *team*). He argues that if a structure occurs in many words, listeners are more likely to encounter it, and in turn be better at perceiving it without perceptual adaptation to a native structure. Nagy (2010), however, points out that certain previously foreign sounds which recombine features already present in the borrowing language may be reliably borrowed even if they appear in only a small amount of (infrequent) words. For example, he notes the emergence of contrast between long [y:] (borrowed) and short [y] (native) in Dutch (which recombines vowel length with a native category), despite the long phone only occurring in a handful of borrowed words (e.g., *centrifuge*). In the statistical analysis we performed on our experimental results, we simply considered word-specific factors as a type of noise by including word as a random effect in our models. Indeed, with only 12 individual items in our production task it is difficult to disentangle the various lexical attributes of individual words (e.g., frequency, age of introduction, source language). Additionally, many (though not all) of the words have domain-specific uses; *goal* for example is only used in reference to scoring points during a sports match and does not replace native *doel* globally (cf. examples like *lunch* which has supplanted *middageten* almost entirely for most speakers). Given the restricted semantic domains in which many of these words appear and their relative rarity, it may prove rather difficult to systematically tease apart the word-level effects that might influence

adaptation strategy. Future work might thus seek to include a wider range of target words in order to better ascertain which lexical characteristics influence (non-)adaptation.

One might also ponder the social aspects that could affect the emergence of the [k]~[g] contrast in Dutch. Previous explorations of Dutch /g/ pointed to regional variability in the use of the phone [g], with speakers from more peripheral regions (those located further away from the *Randstad*, the main population centre of the Netherlands) generally producing fewer [g] tokens and using rather more adapted tokens (van Bezooijen & Gerritsen, 1994; Van de Velde & van Hout, 2002). Specifically, van Bezooijen and Gerritsen (1994) found that the participants who produced an adapted sound such as the fricative [χ]—which corresponds to the orthographic pronunciation of the word—were nearly always from the far south of the Dutch language area. Van de Velde and van Hout (2002) similarly found higher proportions of an adapted variant in the south and higher proportions of [g] in the north. This could potentially reflect the urban environment of the Randstad compared to the relatively rural environment in much (though not all) of the peripheral areas. While we did not set out to test this specific hypothesis, we explored this in our data post hoc, by log-transforming the population density of the home region¹⁷ of our speakers and using it as a predictor of [g] production in a mixed effects logistic regression, including a random intercept for Item (which itself included a random slope for Population Density).¹⁸ We found that this model was a significantly better fit than a simpler model without the predictor Population Density, suggesting that speakers from areas with a higher population density tended to produce more [g] ($\beta = 0.32$, 95% CI = [0.08 : 0.60], $\chi^2(1) = 10.86$ $p < 0.001$). We may speculate what could drive such differences in production. Language attitudes, and specifically views of foreign languages like English being a threat to Dutch could differ between more urban versus more rural areas. However, a more straightforward interpretation might be that, in more urban areas, there is simply more contact with non-native speakers of Dutch (and potentially use of English as *lingua franca*, especially amongst the more highly educated in university towns), leading urban speakers to show different production patterns from rural speakers. This is a clear avenue of interest for future research.

4.2. Perception

Next, we turn to our consideration of perception alongside production. Following the variability observed in production, we might have expected to observe a high range of variability in perception. Specifically, individuals who often produce [g] in loanwords could be expected to

¹⁷ We asked participants to report the town(s) they grew up in. Population densities were extracted using the NUTS3 level from the 2004 census data made available as part of the Dutch *Centraal Bureau voor de Statistiek*'s Open Data initiative and from the 2011 census data made available by Belgium's statistical office StatBel.

¹⁸ We did not include a random intercept for Participant since nearly all participants were from different hometowns, and the Population Density was thus in a near 1:1 relation with Participant.

also perform well in a perceptual task requiring them to distinguish between emerging [g] and native [k], with individuals who produce less [g] in turn being less good at distinguishing the new sound from the native one. While we found some evidence for this, the observed effect is not overwhelming. We did find that participants discriminated the emerging contrast less well than an established contrast in Dutch, but overall accuracy was very high. This means that variability amongst participants was low, and this limited our ability to observe a strong link between production and perception. Any conclusions concerning the link between the two are therefore tentative. We will start by presenting a limitation of our perception results and then discuss two plausible lines of explanation for the lack of variability we observed, one concerning the specifics of our sample and the other concerning the specifics of [g].

First, we should point out a caveat to the interpretation of our perception results. We compared perception of the emerging contrast relative to an established baseline, but we did not compare perception of the emerging contrast relative to a pair of sounds that are completely absent from the Dutch inventory. Our perceptual results indicate that the contrast between [k] and [g] is perceived slightly less well than the contrast between [p] and [b], and while we take the position that this is due to the emerging nature of the phoneme /g/, it is also compatible with a view that says that /g/ does not form a contrast with /k/ in the Dutch inventory. Such an account would predict that the ABX results are reflective of our participants' ability to perceive any similar 'non-contrast.' Inclusion of a contrast like [c]~[J] (also a stop voicing contrast) in our experiment would help disentangle these views: If our listeners have a harder time discriminating the non-contrast than the emerging [k]~[g] contrast, that would be strong evidence for an abstract representation of /g/ (and in line with a gradient view of contrast since [k]~[g] was still perceived as less distinct than [p]~[b]). In fact, Peperkamp and Bouchon (2011), who used an ABX discrimination task with an identical procedure, including a 20-trial training phase with feedback, found that accuracy on a non-native, non-emerging contrast only reached about 70%. This suggests that performance in these kinds of tasks can in fact be much lower than what we observed for our emerging contrast. This also suggests that the feedback during training was unlikely to have provided much training on our target contrasts (beyond training on the task itself).

Additionally, certain aspects of the present study make the gradient contrastiveness account more plausible. First, our ABX task required participants to abstractly represent the different stimuli; a simple acoustic comparison (which might allow participants to distinguish even a contrast they were not familiar with) was not available due to the nature of the task. This was likely bolstered by the fact that we used two different female voices for the A and B stimuli and a male voice for the X stimulus, with the crucial contrast appearing in a non-prominent position within three-syllable non-words. Second, we observed an individual-level link between phonetic implementation and frequency of use of the new sound [g] (participants who produced longer prevoicing also produced more prevoiced tokens). This is exactly the kind

of relation predicted by a gradient view of contrast. Thus, although worse discrimination of a contrast like [c]~[ʃ] compared to the emerging contrast would be strong evidence for gradient representation, evidence from similar studies in addition to features of our own design already speak to this issue.

One aspect that may have limited our ability to observe large differences in perception has to do with our experimental sample. Our testing site, Utrecht, is located in the aforementioned urban *Randstad* area, where [g] usage is reportedly high. This means that all of our participants, who were all young university students, have been exposed to productions of [g] in loanwords and may therefore be well trained in the perceptual realm, without necessarily producing many [g] tokens themselves. Therefore, we may have observed less variability in the perception of the contrast than we saw in production due to the social specifics of the sample we tested. Previous work has shown that perceptual adaptation to contrasts not present in one's native dialect does not completely follow shifts in production, though an individual level link may still be present (Evans & Iverson, 2007). This suggests that phonological perception and production may evolve separately (or at least semi-independently), and is a clear avenue for future research around social aspects affecting loanword (non-)adaptation.

A second aspect that could explain some of the lack of variability we observed in perception has to do with what makes the contrast between [k] and [g] an ideal candidate to emerge in the first place: the place of [g] in the native Dutch phonological system. The phone [g] is not entirely absent from native Dutch phonology, as it exists as a contextual allophone of /k/ (specifically, in regressive voicing assimilation). As noted in the introduction, previous work has found that allophone perception is considerably better when the target contrast is presented outside of the native licensing environment (Peperkamp & Dupoux, 2003; Whalen et al., 1997). In our case, target [g] in our ABX discrimination task was not presented in a context that licenses regressive voicing assimilation. Thus, perhaps participants' performance in our ABX task was particularly strong since the target contrast is a well established case of allophony in Dutch. This high perceptability could in fact bolster the chances that a foreign contrast will come to distinguish words in the borrowing language. Indeed, if all speakers are capable of perceiving the contrast relatively well, it has a higher likelihood of taking hold at the population level.

Allophonic status has been pointed to in explaining some cases of phonological contrast emergence through borrowing. For example, Blevins (2009) notes that while Mekeo lacks contrastive coronals in its native lexicon, coronals do surface as allophones of velars in some phonological contexts. This, then, could explain why certain recent borrowings from English may be pronounced with coronals, like *tauli* for “towel.” This in turn could lead to the emergence of a contrast between velars and coronals in the language. Kennard and Lahiri (2020) go so far as to propose that phonological emergence through lexical borrowing is *necessarily* dependent on the new contrast already existing as an allophonic alternation in the borrowing language. Our

Dutch case would then be a prototypical example of this process: Dutch speakers can distinguish between [k] and [g] if they occur in onset position (i.e., outside of the context where they natively alternate, namely preceding obstruents), making a non-adaptation more likely to take hold in a word like *buggy*.

Furthermore, it has been claimed that foreign sounds which recombine features or gestures already present in the borrowing language may be more likely to be borrowed without being adapted (Cohen, 2019; Ussishkin & Wedel, 2003). This could also favour the emergence of a phonological contrast between /k/ and /g/ in Dutch, since the feature [voice] is already used to make contrasts in stop consonants. As noted above, ‘filling the gap’ in the native inventory adheres to the principle of *feature economy* (Clements, 2003), since [voice] becomes maximally contrastive within the stop series, in turn making the system on the whole more economical. While phonological inventories are by no means universally maximally economical, they do seem to be more economical than might be expected by chance (Coupé, Marsico, & Philippson, 2011; Dunbar & Dupoux, 2016; Mackie & Mielke, 2011). This typological evidence suggests that there may be a cognitive bias constraining learning that would lead to expectations of economical systems, and indeed, higher-order organizational pressures, such as economy, are often suggested to play a role during language acquisition (e.g., Thompson & de Boer, 2017). For our case, this could mean that Dutch-learning children might expect [k] to contrast with [g], since other pairs of stops are contrasted along the same dimension. Relatively little input might therefore be required for the contrast to be acquired. It might suffice for a single word in a young child’s lexicon to contain /g/ (e.g., *buggy* pronounced [bygi], a quite common word in most Dutch-speaking children’s lives) for them to assume a voicing contrast for velar stops.

There is some evidence from Dutch-speaking children that this account is reasonable. Stoehr, Benders, van Hell, and Fikkert (2022) found that children between 3 and 6 years old produced varying rates of [g] after being exposed to novel names produced by an experimenter with prevoicing (e.g., *Gabi* produced as [gabi]).¹⁹ Overall, 27% of target segments were produced with prevoicing (other tokens were produced like [kabi]), but at least one child produced consistent prevoicing on all tokens. This inter-speaker variability is in line with the results of our production task and with previous investigations of Dutch /g/. The authors of that study argue that their results are evidence of generalization of the feature [voice] from labial and coronal contrasts (contrasts that are ubiquitous in the children’s input) to the velar place of articulation (an emerging contrast that some of the children tested may not yet have heard). This is precisely the type of generalization predicted by a feature economy account.

¹⁹ Their study also included one real loanword target: *goal*. Some of the parents of the child participants reported that their children already knew this word, but the authors state that they found similar results when they excluded *goal* tokens from their analyses.

4.3. Putting it all together

Overall, our participants may all have been capable of perceiving the emerging contrast well, but their use of the contrast in production might be subject to further constraints (which could explain the relatively larger variability we observed in our production results). Two explanations seem likely to have a role to play. Firstly, prevoicing is more effortful to produce during velar constrictions than during coronal or labial constrictions simply due to the fact that the space behind the constriction is smaller (and pressure thus builds up more quickly). Indeed, this can explain why cross-linguistically, missing /g/ from a stop inventory is more likely than missing /d/ or /b/, and also why prevoicing at more posterior places of articulation tends to be shorter (e.g., Ohala, 1983). This argument is bolstered by the fact that we observed higher [g] rates in our production data in intervocalic contexts (which favour voicing) than in non-intervocalic contexts (see **Table 1**). Secondly, as mentioned above, we did not explore the social associations that speakers have with the different variants. There may be specific (potentially stigmatized) social meanings associated with any of the variants, and this might differ by speaker. For example, some speakers might find it pompous to use the unadapted variant in casual speech, while others might find it uneducated not to. Future research might specifically explore these potential differences.

Let us turn finally to what this case of loanword phonology can tell us about the representation of phonological contrasts. Taking the distribution of contrasts as a starting point, we can imagine that a contrast like /t/~/p/, which is ubiquitous in Dutch, and is not subject to neutralization, should be perceived on the far end of a continuum of distinctiveness. A contrast like /p/~/b/, which is also ubiquitous, but is subject to neutralization (i.e., final devoicing means these sounds are not contrastive at the ends of words) is likely to be perceived as slightly less distinctive than /t/~/p/, and should therefore be placed less far along a continuum of distinctiveness. This is in line with previous work showing that listeners' perceptions of different contrasts reflect properties of those contrasts' lexical distribution (e.g., K. C. Hall & Hume, 2013; K. C. Hall et al., 2015; Martin & Peperkamp, 2017). Indeed, this continuum seems to closely track K. C. Hall's (2013) predictability of distribution continuum, but evidence from perception shows that the story is even more complex. Martin and Peperkamp (2017), for example, show that French listeners are indeed biased by their knowledge of contrast distribution in the French lexicon during word recognition, but also that those same listeners are more attentive to contrasts with a greater acoustic difference. The latter type of knowledge is posited to be part of the grammar in substance-based theories of phonology (i.a., Archangeli & Pulleyblank, 1994; Bybee, 2001; Hayes & Steriade, 2004), and can be used to explain why certain phonological rules are more common cross-linguistically than others (namely, that certain alternations are perceptually more distinct than others, see, e.g, Steriade, 2001). While proposing how language-specific and language-independent knowledge integrate to influence contrast perception is beyond the scope

of this paper, the data presented here show that fine-grained knowledge should be part of our understanding of contrast. It does not do justice to the status of /g/ to simply say that /g/ only appears in a handful of words in Dutch (and therefore plays only a marginal role, if any, in the phonology). The contrast between /k/ and /g/ is already used by many speakers, and some of the minimal pairs distinguished by this contrast may be present in Dutch children's input (e.g., Stoehr et al., 2022, mention that some of the children in their study knew the word *goal* which forms a minimal pair with the native word *kool*, "cabbage"). Clearly we must consider contrasts to be dynamic, evolving, and intertwined with perception. The predictability of the distribution of a contrast is an important starting point, but data like we have presented here can begin to shed light on the representation of a contrast on the one hand, and also on the intricacies of a contrast's implementation in production (including complex patterns of variability) on the other.

5. Conclusion

To conclude, we have shown the active role that the phoneme /g/ plays in the phonology of Dutch, going beyond allophony to form a new contrast with /k/. We further showed that this contrast is subject to inter-speaker variability in both production and perception. Marginal contrasts, such as the emerging /k/~/g/, can shed light on the multi-faceted nature of contrast and should be considered full-fledged parts of phonology. While the distribution of production patterns tells us a great deal about how distinctive a contrast is, considering listeners' individual-level perception and production can inform a more holistic understanding of contrast.

Additional files

The additional files for this article can be found as follows:

- **Appendix A.** Target sentences. DOI: <https://doi.org/10.16995/labphon.6454.s1>
- **Appendix B.** ABX stimuli. DOI: <https://doi.org/10.16995/labphon.6454.s2>

Acknowledgements

This work was supported by ANR-13-APPR-0012 LangLearn and ANR-17-EURE-0017 and was carried out while Alexander Martin and Marieke van Heugten were members of the LSCP. We would like to thank the following people for their contributions to this work: Yue Sun for preparing the audio recording script; Nicoline van der Sijs for providing us with an electronic version of the database used in compiling her Dutch etymological dictionary; Marieke de Bree and Stephen Whitmarsh for recording stimuli for our ABX task; Iris Mulders and Chris van Run for participant recruitment and technical assistance in Utrecht. Thanks also go to James Kirby and Lauren Hall-Lew for giving comments on a previous version of this manuscript.

Competing interests

The authors have no competing interests to declare.

References

- Archangeli, D., & Pulleyblank, D. (1994). *Grounded Phonology*. MIT Press.
- Babel, M., & Johnson, K. (2010). Accessing psycho-acoustic perception and language-specific perception with speech sounds. *Laboratory Phonology*, 1(1). DOI: <https://doi.org/10.1515/labphon.2010.009>
- Bates, D. M. (2014). *Lme4: Mixed-effects modeling with R*.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, 109(2), 775–794. DOI: <https://doi.org/10.1121/1.1332378>
- Blevins, J. (2009). Another Universal Bites the Dust: Northwest Mekeo Lacks Coronal Phonemes. *Oceanic Linguistics*, 48(1), 264–273. DOI: <https://doi.org/10.1353/ol.0.0033>
- Boersma, P., & Hamann, S. (2009). Loanword adaptation as first-language phonological perception. In A. Calabrese & L. Wetzel (Eds.), *Loanword Phonology* (pp. 11–58). John Benjamins. DOI: <https://doi.org/10.1075/cilt.307.02boe>
- Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer*. Retrieved from <http://www.praat.org>
- Booij, G. (1999). *The Phonology of Dutch* (Vol. 5). Oxford University Press.

- Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In *Contrast in Phonology: Theory, Perception, Acquisition* (Vol. 13, p. 143–172).
- Boretzky, N. (1991). Contact-Induced Sound Change. *Diachronica*, 8(1), 1–15. DOI: <https://doi.org/10.1075/dia.8.1.02bor>
- Brown, G., Anderson, A., Shillcock, R., & Yule, G. (1985). *Teaching talk: Strategies for production and assessment*. Cambridge University Press.
- Bybee, J. L. (2001). *Phonology and language use*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511612886>
- Clements, G. N. (2003). Feature Economy in Sound Systems. *Phonology*, 20(3), 287–333. DOI: <https://doi.org/10.1017/S095267570400003X>
- Cohen, E.-G. (2019). Loanword phonology in Modern Hebrew. *Brill's Journal of Afroasiatic Languages and Linguistics*, 11(1), 182–200. DOI: <https://doi.org/10.1163/18776930-01101012>
- Collins, B., & Mees, I. M. (Eds.) (2003). *The Phonetics of English and Dutch* (5th Revised Edition ed.). Brill.
- Coupé, C., Marsico, E., & Philippson, G. (2011). How economical are phonological inventories? In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 524–527).
- Crawford, C. (2007). The role of loanword diffusion in changing adaptation patterns: A study of coronal stops in Japanese borrowings. *Working Papers of the Cornell Phonetics Laboratory*, 16, 32–56.
- Daland, R., Oh, M., & Kim, S. (2015). When in doubt, read the instructions: Orthographic effects in loanword adaptation. *Lingua*, 159, 70–92. DOI: <https://doi.org/10.1016/j.lingua.2015.03.002>
- Davidson, L. (2007). The Relationship between the Perception of Non-Native Phonotactics and Loanword Adaptation. *Phonology*, 24(2), 261–286. DOI: <https://doi.org/10.1017/S0952675707001200>
- DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, 126(1), 87–100. DOI: <https://doi.org/10.1016/j.cognition.2012.09.004>
- de Jong, K., & Cho, M.-H. (2012). Loanword phonology and perceptual mapping: Comparing two corpora of Korean contact with English. *Language*, 88(2), 341–368. DOI: <https://doi.org/10.1353/lan.2012.0035>
- Donaldson, B. (1983). *Dutch: A linguistic history of Holland and Belgium*. Uitgeverij Martinus Nijhoff.
- Dunbar, E., & Dupoux, E. (2016). Geometric Constraints on Human Speech Sound Inventories. *Frontiers in Psychology*, 7. DOI: <https://doi.org/10.3389/fpsyg.2016.01061>
- Eckman, F. R., & Iverson, G. K. (2015). *Second Language Acquisition and Phonological Change*. In P. Honeybone & J. Salmons (Eds.). Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199232819.013.005>

- Ernestus, M. (2011). Gradience and Categoricality in Phonological Theory. In M. van Oostendorp, C. J. Ewen, E. Hume & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 1–22). John Wiley & Sons, Ltd. DOI: <https://doi.org/10.1002/9781444335262.wbctp0089>
- Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America*, 121(6), 3814. DOI: <https://doi.org/10.1121/1.2722209>
- Fagyal, Z., Hassa, S., & Ngom, F. (2002). L'opposition [e]-[ɛ] en syllabes ouvertes de fin de mot en français parisien : étude acoustique préliminaire. In *Actes des journées d'études sur la parole* (pp. 165–168).
- Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, 66(3), 363–376. DOI: <https://doi.org/10.3758/BF03194885>
- Goossens, J. (1974). *Historische Phonologie des Niederländischen*. Niemeyer. DOI: <https://doi.org/10.1515/9783111411576>
- Hall, D. C., & Hall, K. C. (2016). Marginal contrasts and the Contrastivist Hypothesis. *Glossa: A journal of general linguistics*, 1(1), 50. DOI: <https://doi.org/10.5334/gjgl.245>
- Hall, K. C. (2009). *A Probabilistic Model of Phonological Relationships from Contrast to Allophony* (Doctoral dissertation). The Ohio State University.
- Hall, K. C. (2013). A typology of intermediate phonological relationships. *The Linguistic Review*, 30(2), 215–275. DOI: <https://doi.org/10.1515/tlr-2013-0008>
- Hall, K. C., & Hume, E. V. (2013). Perceptual confusability of French vowels. In (pp. 060113–060113). DOI: <https://doi.org/10.1121/1.4800615>
- Hall, K. C., Letawsky, V., Turner, A., Allen, C., & McMullin, K. (2015). Effects of predictability of distribution on within-language perception. In S. Vinerte (Ed.), *Proceedings of the 2015 Annual Conference of the Canadian Linguistics Association* (pp. 1–14). Canadian Linguistics Association.
- Hamann, S., & de Jonge, A. (2015). Eliciting the Dutch loan phoneme /g/ with the Menu Task. In *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Haugen, E. (1950). The Analysis of Linguistic Borrowing. *Language*, 26(2), 210–231. DOI: <https://doi.org/10.2307/410058>
- Hayes, B., & Steriade, D. (2004). Introduction: The phonetic bases of phonological Markedness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 1–33). Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511486401.001>
- Hume, E., & Johnson, K. (2001). A Model of the Interplay of Speech Perception and Phonology. In *The Role of Speech Perception in Phonology*. DOI: <https://doi.org/10.1163/9789004454095>
- Itô, J., & Mester, A. (1999). The phonological lexicon. In N. Tsujimura (Ed.), *The Handbook of Japanese Linguistics* (pp. 62–100). Blackwell. DOI: <https://doi.org/10.1002/9781405166225.ch3>
- Itô, J., & Mester, A. (2001). Covert generalizations in Optimality Theory: The role of stratal faithfulness constraints. In *Studies in Phonetics, Phonology and Morphology*, 7.

- Jansen, W. (2007). Dutch regressive voicing assimilation as a ‘low level phonetic process’: Acoustic evidence. In J. van de Weijer & E. J. van der Torre (Eds.), *Current Issues in Linguistic Theory* (Vol. 286, pp. 125–151). John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/cilt.286.06jan>
- Kang, Y. (2003). Perceptual similarity in loanword adaptation: English postvocalic wordfinal stops in Korean. *Phonology*, 20(2), 219–273. DOI: <https://doi.org/10.1017/S0952675703004524>
- Kang, Y. (2011). Loanword phonology. In *The Blackwell Companion to Phonology* (pp. 1–25). DOI: <https://doi.org/10.1002/9781444335262.wbctp0095>
- Kennard, H. J., & Lahiri, A. (2020). Nonesuch phonemes in loanwords. *Linguistics*, 58(1), 83–108. DOI: <https://doi.org/10.1515/ling-2019-0033>
- Lee, P. (2013). The impact of borrowed sounds and neutralization on Korean contrasts: An entropy-driven analysis. In *IULC Working Papers*.
- Lev-Ari, S., & Peperkamp, S. (2014). An experimental study of the role of social factors in language change: The case of loanword adaptations. *Laboratory Phonology*, 5(3). DOI: <https://doi.org/10.1515/lp-2014-0013>
- Lewis, P. M., Simons, G. F., & Fenning, C. D. (Eds.) (2016). *Ethnologue: Languages of the World* (19th ed. ed.). SIL International.
- Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 527–565. DOI: <https://doi.org/10.1080/00437956.1964.11659830>
- Mackie, S., & Mielke, J. (2011). Feature economy in natural, random, and synthetic inventories. In G. N. Clements & R. Ridouane (Eds.), *Where Do Phonological Features Come From?: Cognitive, Physical and Developmental Bases of Distinctive Speech Categories* (Vol. 6, pp. 43–64). John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/lfab.6.03mac>
- Martin, A., & Peperkamp, S. (2015). Asymmetries in the exploitation of phonetic features for word recognition. *The Journal of the Acoustical Society of America*, 137(4), EL307–EL313. DOI: <https://doi.org/10.1121/1.4916792>
- Martin, A., & Peperkamp, S. (2017). Assessing the distinctiveness of phonological features in word recognition: Prelexical and lexical influences. *Journal of Phonetics*, 62, 1–11. DOI: <https://doi.org/10.1016/j.wocn.2017.01.007>
- Nagy, R. (2010). *Kölcsönszavak fonológiai integrációja a holland nyelvbe* (Doctoral dissertation). Eötvös Loránd Tudományegyetem.
- Ohala, J. J. (1983). The Origin of Sound Patterns in Vocal Tract Constraints. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 189–216). Springer New York. DOI: https://doi.org/10.1007/978-1-4613-8202-7_9
- Peperkamp, S., & Bouchon, C. (2011). The Relation Between Perception and Production in L2 Phonological Processing. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech 2011)* (pp. 161–164). Causal Productions. DOI: <https://doi.org/10.21437/Interspeech.2011-72>
- Peperkamp, S., & Dupoux, E. (2003). Reinterpreting Loanword Adaptations: The Role of Perception. In M.-J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 367–370). Causal Productions.

- Peperkamp, S., Pettinato, M., & Dupoux, E. (2003). Allophonic Variation and the Acquisition of Phoneme Categories. In *Proceedings of the 27th Annual Boston University Conference on Language Development*. Cascadilla Press.
- Peperkamp, S., Vendelin, I., & Nakamura, K. (2008). On the perceptual origin of loanword adaptations: Experimental evidence from Japanese. *Phonology*, 25(1), 129–164. DOI: <https://doi.org/10.1017/S0952675708001425>
- Pinget, A.-F., Kager, R., & Van de Velde, H. (2020). Linking Variation in Perception and Production in Sound Change: Evidence from Dutch Obstruent Devoicing. *Language and Speech*, 63(3), 660–685. DOI: <https://doi.org/10.1177/0023830919880206>
- Poplack, S., Robillard, S., Dion, N., & Paolillo, J. C. (2020). Revisiting phonetic integration in bilingual borrowing. *Language*, 96(1), 126–159. DOI: <https://doi.org/10.1353/lan.2020.0004>
- Renwick, M. E. L. (2014). *The Phonetics and Phonology of Contrast: The Case of the Romanian Vowel System*. De Gruyter. DOI: <https://doi.org/10.1515/9783110362770>
- Renwick, M. E. L., & Ladd, D. R. (2016). Phonetic Distinctiveness vs. Lexical Contrastiveness in Non-Robust Phonemic Contrasts. *Laboratory Phonology*, 7(1), 19. DOI: <https://doi.org/10.5334/labphon.17>
- Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, 41(1), 71–80. DOI: [https://doi.org/10.1016/S0167-6393\(02\)00094-8](https://doi.org/10.1016/S0167-6393(02)00094-8)
- Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, 39(1), 96–109. DOI: <https://doi.org/10.1016/j.wocn.2010.11.006>
- Scobbie, J. M., & Stuart-Smith, J. (2008). Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English. In P. Avery, B. E. Dresher & K. Rice (Eds.), *Contrast in phonology: Theory, perception, acquisition* (pp. 87–113). Mouton de Gruyter.
- Simonović, M. (2015). *Lexicon immigration service: Prolegomena to a theory of loanword integration = De immigratiедienst van het lexicon: Prolegomena tot een theorie over leenwoordintegratie: (met een samenvatting in het Nederlands) = Imigracijska služba leksikona: Prolegomena za teoriju integracije tuđica: (sa sažetkom na srpskohrvatskom)* (Doctoral dissertation). Utrecht University.
- Steriade, D. (2001). The phonology of perceptability effects: The P-map and its consequences for constraint organization. In S. Inkelas & K. Hanson (Eds.), *The nature of the word: Essays in honor of Paul Kiparsky*. MIT Press.
- Stevenson, S., & Zamuner, T. (2017). Gradient phonological relationships: Evidence from vowels in French. *Glossa: A journal of general linguistics*, 2(1), 58. DOI: <https://doi.org/10.5334/gjgl.162>
- Stoehr, A., Benders, T., van Hell, J. G., & Fikkert, P. (2022). Feature generalization in Dutch-German bilingual and monolingual children's speech production. *First Language*, 42(1), 101–123. DOI: <https://doi.org/10.1177/01427237211058937>
- Thompson, B., & de Boer, B. (2017). Structure and abstraction in phonetic computation: Learning to generalise among concurrent acquisition problems. *Journal of Language Evolution*, 2(1), 94–112. DOI: <https://doi.org/10.1093/jole/lzx013>

- Thomason, S. G., & Kaufman, T. (1988). *Language Contact, Creolization, and Genetic Linguistics*. University of California Press. DOI: <https://doi.org/10.1525/9780520912793>
- Uffmann, C. (2015). *Loanword Adaptation*. In P. Honeybone & J. Salmons (Eds.). Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199232819.013.019>
- Ussishkin, A., & Wedel, A. (2003). Gestural Motor Programs and the Nature of Phonotactic Restrictions: Evidence from Loanword Phonology. In M. Tsujimura & G. Garding (Eds.), *WCCFL 22 Proceedings* (pp. 505–518). Cascadilla Press.
- van Bezooijen, R. A. M. G., & Gerritsen, M. (1994). De uitspraak van uitheemse woorden in het Standaard-Nederlands: Een verkennende studie. *De Nieuwe Taalgids*, 87, 145–161.
- van Coetsem, F. (1988). *Loan Phonology and the Two Transfer Types in Language Contact*. Foris. DOI: <https://doi.org/10.1515/9783110884869>
- van der Sijs, N. (2002). *Chronologisch woordenboek: De ouderdom en herkomst van onze woorden en betekenissen*.
- Van de Velde, H., & van Hout, R. (2002). Uitspraakvariatie in leenwoorden. In *NVT-onderwijs en -onderzoek in Franstalig gebied* (pp. 77–95). Vantilt.
- Vendelin, I., & Peperkamp, S. (2006). The influence of orthography on loanword adaptations. *Lingua*, 116(7), 996–1007. DOI: <https://doi.org/10.1016/j.lingua.2005.07.005>
- Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128(2), 179–186. DOI: <https://doi.org/10.1016/j.cognition.2013.03.002>
- Weinreich, U. (1953). *Languages in contact: Findings and problems*. Linguistic Circle of New York.
- Whalen, D., Best, C. T., & Irwin, J. R. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics*, 25(4), 501–528. DOI: <https://doi.org/10.1006/jpho.1997.0058>
- Winford, D. (2005). Contact-induced changes: Classification and processes. *Diachronica*, 22(2), 373–427. DOI: <https://doi.org/10.1075/dia.22.2.05win>
- Winter, B., & Roettger, T. (2011). The Nature of Incomplete Neutralization in German: Implications for Laboratory Phonology. *Grazer Linguistische Studien*, 76, 55–74.
- Zuraw, K., O'Flynn, K. C., & Ward, K. (2019). Non-native contrasts in Tongan loans. *Phonology*, 36(1), 127–170. DOI: <https://doi.org/10.1017/S095267571900006X>