



Open Library of Humanities

Producing and perceiving socially structured coarticulation: Coarticulatory nasalization in Afrikaans

Andries W. Coetzee, Department of Linguistics, University of Michigan, Ann Arbor, MI, US; North-West University, Potchefstroom, South Africa, ZA; University of Johannesburg, Johannesburg, South Africa, ZA, coetzee@umich.edu

Patrice Speeter Beddor, Department of Linguistics, University of Michigan, Ann Arbor, MI, US, beddor@umich.edu

Will Styler, Department of Linguistics, University of California San Diego, San Diego, CA, US, wstyler@ucsd.edu

Stephen Tobin, Department of Linguistics, University of Michigan, Ann Arbor, MI, US, s.j.tobin.99@cantab.net

Ian Bekker, North-West University, Potchefstroom, South Africa, ZA, Ian.Bekker@nwu.ac.za

Daan Wissing, North-West University, Potchefstroom, South Africa, ZA, Daan.Wissing@nwu.ac.za

Most theories of phonetics assume a tight relation between production and perception, and recent years have also seen increasing evidence for such a relation at the level of the individual. For the most part, however, this evidence comes from socially homogeneous speech communities where the targeted pattern of variation is mostly socially neutral. What implications might socially structured phonetic variation in the speech community have for the perception-production link? If listeners can predict the phonetic patterns of a talker based on the talker's actual or assumed identity, would they adjust their perceptual strategies accordingly, possibly weakening the link between their own production and perception patterns? This study reports the results of a pair of experiments that investigate the production and perception of coarticulatory vowel nasalization in Afrikaans, a language for which variation in coarticulatory nasalization is socially structured. Relying on nasal airflow measures, the production experiment showed that speakers of White Afrikaans produce more extensive coarticulatory nasalization than speakers of Kleurling Afrikaans. The perception experiment used an eye-tracking paradigm to assess listeners' perceptual reliance on coarticulatory nasalization, and found (i) that Afrikaans speakers' use of coarticulatory nasalization in production predicts their perceptual reliance on this information, (ii) that they rapidly adjust to the coarticulatory timing patterns in the speech of other speakers, but also (iii) that they do not adjust their perceptual reliance on coarticulation in response to the assumed identity of the speaker. The link between perception and production therefore persists, even in this situation of socially structured variation in coarticulatory timing.



1. Introduction

This study investigates the perception and production of coarticulated speech in an Afrikaans speech community in which a selected pattern of coarticulation, anticipatory vowel nasalization, is socially structured. The study is situated at the intersection of, on the one hand, the long-standing tradition of phonetic research on the relationship between speech perception and production and, on the other hand, newer lines of research on listeners' adjustment of their perceptual strategies in response to the social identity of their interlocutors. A broad hypothesis underlying this work is that how a language user produces speech is complexly related to how that individual perceives speech. Our main goals are to determine, for coarticulated speech, whether perception is guided by knowledge of social structuring and thus whether the complex production-perception relation is socially mediated.

1.1. The relation between the perception and production of coarticulation

Although speech production and perception cannot be assumed to be isomorphic (e.g., Pardo, 2012) there is nonetheless strong theoretical motivation for assuming a tight relation between speaking and listening. For some theoretical approaches, this tight relation is formalized in the nature of produced and perceived phonetic units. Gesturalist theories of speech perception, for example, postulate that the forms of speaking—vocal tract actions—are, correspondingly, the forms of perception (Fowler, 1986; Liberman & Mattingly, 1985; Liberman & Whalen, 2000). Alternatively, the similarity between produced and perceived forms of speech might reside in the acoustic-auditory domain, as in the assumption of the DIVA (Directions into Velocity of Articulators) production model that the targets of production are auditory/perceptual (Guenther, Hampson, & Johnson 1998). Some theoretical perspectives further postulate that such parity holds not only for the *nature* of phonetic forms but also for the *specific* forms produced and perceived by individual speakers. This assumption emerges in exemplar-based models, which are sometimes agnostic concerning the nature of stored experiences but typically involve a perception-production loop in which productions are drawn from the larger perceptual space (Pierrehumbert, 2001). From a different perspective but along similar lines, many theoretical approaches to sound change rest on the assumption that a listener's innovative percepts (i.e., percepts that differ from the community norm) are manifested in the subsequent production patterns of that individual (Beddor, 2009; Harrington, Kleber, & Reubold, 2008; Lindblom, Guion, Hura, Moon, & Willerman, 1995; Ohala, 1981; Yu, 2013).

For the production and perception of coarticulated speech, this theoretically postulated tight link is well supported by empirical findings for various communities of speaker-listeners. For example, gestural overlap patterns that are language-specific (e.g., Beddor, Harnsberger, & Lindemann, 2002; Beddor & Krakow, 1999) or age group-specific (Harrington et al., 2008; Kleber, Harrington, & Reubold, 2012) have been shown to correspond to language- or age-specific

perception in that the more extensive a speech community's production of coarticulatory overlap (e.g., coarticulatory vowel fronting or nasalization), the greater those language users' perceptual adjustments for the acoustic effects of that overlap. Other studies have shown language variety- and age-specific production and perception of coarticulation to be linked in that the group for which one type of information is especially informative perceptually is also the group that produces that information to a greater extent (Coetzee, Beddor, Shedden, Styler, & Wissing, 2018; Kuang & Cui, 2018; Schertz, Kang, & Han, 2019).

Similar to speech communities, individuals also differ systematically from each other in their production and perception of coarticulated speech (see Beddor, Coetzee, Boland, McGowan, & Styler, 2018 and Yu & Zellou, 2019 for reviews), yet findings are mixed regarding whether an individual's produced coarticulation predicts their perception. Some studies of individuals' perceptual adjustments for, or attention to, the acoustic effects of coarticulation report a positive correlation with produced coarticulation (Beddor et al., 2018; Yu, 2019; Zellou, 2017), while others have failed to establish a link (Grosvald, 2009; Kataoka, 2011). Individuals' perceptual weightings of coarticulatory effects relative to the source of those effects are also not predicted by their production patterns in some studies (Schertz, Cho, Lotto, & Warner, 2015; Shultz, Francis, & Llanos, 2012), and findings that they are (e.g., Coetzee et al., 2018) may be driven in part by group-based differences (see Schertz & Clare, 2020 for a review). Thus, the strength of the production-perception link for individual speaker-listeners appears to be variable, and the factors that mediate the strength of the link are not yet well understood.

One scenario in which this link may be weak or even absent is in speech communities where production patterns are socially structured. In these situations, listeners may interact regularly with speakers whose production patterns are predictably different from their own. Beddor et al. (2018, p. 935), for instance, speculated that a weaker perception-production link may be observed in a situation of an ongoing sound change, where younger and older community members may differ in their adoption of new phonetic norms. Harrington et al.'s review of the relevant literature (Harrington, Kleber, Reubold, Schiel, & Stevens, 2019) suggests that perception can lead production in ongoing changes, and two recent studies of sound changes in progress—Kuang and Cui's study of the tense/lax register contrast in Southern Yi (Kuang & Cui, 2018) and Pinget et al.'s study of obstruent devoicing in Dutch (Pinget, Kager, & van de Velde, 2020)—document this pattern at the level of individual speaker-listeners of the relevant speech varieties. In addition, though, Pinget et al. (2020) find that, when nearing completion, the change has progressed further in production than perception.

We propose that a similar misalignment or relaxation of the perception-production link may be observed in speech communities where differences in production patterns are linked not to an ongoing sound change, but to the social structure of the speech community. If different subgroups of the speech community have different characteristic production patterns, then listeners will

encounter these different patterns and successful communication might be aided by listeners being especially flexible in their perceptual strategies.

In this study, we use an approach, and investigate a scenario, that to some extent is similar to that studied by Beddor et al. (2018). They investigated variability in the production and perception of coarticulatory nasalization in Midwestern American English, finding considerable variation in the extent of produced coarticulatory nasalization. They also found that speakers who produce particularly heavy anticipatory nasalization attend particularly closely to that information in perception—for example, they are faster to identify [sɛnt] as *scent* rather than *set*, indicating that they rely more on the coarticulatory information during the vowel for disambiguation. In that speech community, however, there is no evidence of social structuring of the extent of nasalization, and differences in nasalization do not clearly index a speaker in any meaningful manner. It is hence not possible for a listener to predict, based on the identity of a specific speaker, whether the speaker will nasalize more or less. In the current study, we investigate the same phenomenon in an Afrikaans speech community where a similar degree of variation in the amount and extent of produced coarticulatory nasalization is observed. However, the variation in the extent of coarticulatory nasalization is clearly socially structured in this speech community—producing more or less nasalization marks an individual as being a speaker of a specific socio-ethnic variety of the language. We investigate whether this difference between Midwestern American English and Afrikaans in terms of the social structure of the variation impacts the nature of the perception-production link.

1.2. Talker-sensitive perceptual strategies

In general, listeners use the lawfully structured acoustic variation afforded by coarticulation to facilitate perceptual processing (e.g., Whalen, 1984, among many others). However, as discussed above, individual listeners differ from each other in the extent to which they attend to and use that information in making lexical decisions, and they appear to do so in ways that depend in part on their own production of coarticulation. A question that naturally arises, then, is whether listeners will nonetheless adapt their perceptual strategies to the idiosyncratic coarticulatory patterns of their interlocutors—idiosyncratic patterns that are well-documented in the literature, including for coarticulatory nasalization (Beddor et al., 2018, among others). That is, when attending to the speech of a talker¹ with a coarticulatory pattern different from their own, will listeners adjust their perceptual strategy for the processing of coarticulatory information accordingly?

Based on findings that listeners adapt their perceptual strategies to talker-specific idiosyncrasies, we hypothesize that the same will hold for talker-specific coarticulatory nasalization. Trude

¹ In order to differentiate between individuals whose speech is used as stimuli in speech perception experiments (as in Section 3 below) and general members of speech community, we will use “talker” to refer to the former and “speaker” to the latter throughout this paper.

and Brown-Schmidt (2012; Trude, Duff, & Brown-Schmidt, 2014), for instance, exposed listeners to two talkers who differed in whether they produced a raised diphthong in words ending in /æɡ/ compared to /æk/—that is, although both talkers realized words like *back* as [bæk], one realized *bag* as [bæg] and the other as [berg]. Using a visual world paradigm, they conducted an identification experiment in which participants saw images for minimal pair words like *bag* and *back* while hearing auditory [bæk]. Listeners fixated more quickly on the *back* image for the talker who realized *bag* as [berg], showing that listeners relied on different perceptual strategies depending on talker-specific production patterns. These and other similar findings (e.g., Dahan, Drucker, & Scarborough, 2008; Kraljic, Brennan, & Samuel, 2008; see Samuel & Kraljic, 2009, for a review) show that very limited exposure to a novel pattern of a talker is sufficient for listeners to perceptually adapt to that pattern.

Because varieties of the same language can differ systematically in terms of their timing of coarticulatory nasalization, this study also asks whether listeners bring *existing* knowledge about timing patterns in different language varieties to the perceptual task, and consequently rely differentially on coarticulatory information based on their (possibly unconscious) knowledge about these patterned differences. For coarticulatory nasalization, there is clear evidence, despite variation at the level of individual speakers, of broader community-level patterns. Studies have documented systematic variation in anticipatory vowel nasalization for different regional varieties (e.g., Bongiovanni, 2018, for Caribbean and non-Caribbean Spanish; Delvaux, Huet, Piccaluga, & Harmegnies, 2012, for European French; Stroop, 1994, for Belgian French; Tamminga & Zellou, 2015, for American English) and for age groups within a regional or ethnic variety (Wissing, 2018, for so-called White Afrikaans; Zellou & Tamminga, 2014, for Philadelphia English). This study documents socio-ethnically based nasalization patterns in Afrikaans.

There is a growing body of research showing that, when listeners are led to believe that a talker has a particular (typically regional) identity, they actively adjust their perceptual strategies based on their prior knowledge of, or stereotypes about, that speech variety. Niedzielski (1999), for instance, showed that listeners were more likely to accurately identify a word like *house* as having a raised version of the diphthong /aʊ/ when they were led to believe that the talker was Canadian—that is, a talker of an English variety associated with this form of raising—rather than American. Hay et al. (2006a) replicated this finding for a difference between New Zealand and Australian English, showing that listeners are more likely to perceive a word like *fit* as being produced with a raised vowel when they were led to believe that the talker was from Australia, in agreement with the more raised realization of the high front lax vowel in that dialect. Staum-Casasanto (2009a; 2009b) found that American English-speaking listeners were more likely to identify [mæs] in a phrase like *The [mæs] probably lasted ...* as the word /mæst/ when they were led to believe that the talker was Black rather than White, showing that listeners use their knowledge that word-final deletion of /t/ is more common in the speech of Black than White speakers of American English.

Especially relevant to our question of whether listeners bring existing knowledge about variety-specific timing patterns to the perceptual task is Schertz et al.'s study of vocalic f_0 in relation to preceding stops' voice onset time for speakers of two dialects of Chinese Korean (Schertz et al., 2019). Speakers of these dialects differ from each other in the contributions of f_0 and VOT to the differentiation of lenis and aspirated stops, but both dialects differ from Seoul Korean, in which the f_0 information is primary. Schertz et al. found that younger Chinese Korean listeners weighted f_0 more heavily in their lenis-aspirated judgments when they were led to expect that the talker was from Seoul than when they thought the talker was from their own city.

In the current study's investigation of whether Afrikaans-speaking listeners adjust to the coarticulatory pattern of a talker, the socio-ethnic varieties of the talkers differ in prestige. Several studies have documented differences in processing advantages for 'prestige/standardized' versus 'non-prestige/non-standardized' varieties. In an early study, Weener (1969), for instance, found that children from a predominantly White, middle class Detroit neighborhood recalled more words produced by a speaker from their own neighborhood than by a speaker from a predominantly Black, lower class neighborhood. On the other hand, children from the predominantly Black, lower class neighborhood showed no recall difference between the two different speakers. That is, children who spoke the 'prestige' variety of English showed a processing disadvantage for the other variety, while children who spoke the 'non-prestige' variety of English did not show a comparable processing disadvantage for the prestige variety. Sumner and Kataoka (2013) similarly showed that the prestige of an accent appears to influence listeners' responses to talker-specific variation. They found that American English listeners are faster at identifying a word like *thin* after being primed with an auditory presentation of a semantically related word such as *slender*—but only under certain conditions. Specifically, *slender* primed *thin* identification when realized with a final /ɪ/ (the typical American rhotic pronunciation) or with a non-rhotic British English pronunciation, but not with a non-rhotic, 'non-standard' New York City pronunciation. Under the reasonable assumption that the average American listener would have only limited exposure to (non-rhotic) British English, they argued that listener sensitivity to talker identity does not require extensive exposure to the specific speech variety and hypothesized that the higher prestige associated with British compared to New York City English may result in more robust encoding of British exemplars (see also Sumner, Kim, King, & McGowan, 2014).

1.3. Coarticulatory nasalization in two socio-ethnic varieties of Afrikaans

In this study, we focus on differences in coarticulatory nasalization between two socio-ethnic varieties of Afrikaans. Although there are regional differences observed in Afrikaans, the main dialect groups of the language are differentiated along socio-ethnic rather than regional lines

(Stell, 2011, pp. 57–64).² So-called ‘White Afrikaans’ is spoken predominantly by speakers of European descent, and is also the variety of the language that is more likely to be encountered in the media and taught as either first or second language in school settings. So-called ‘Kleurling Afrikaans’ is spoken predominantly by members of the Kleurling community, comprised of descendants of 17th century Dutch settlers, various communities indigenous to South Africa (including both Khoisan and Bantu speakers), and Malaysian and Indonesian slave laborers brought by the Dutch to South Africa in the late 17th and early 18th centuries. Although there are today more speakers of Kleurling than White Afrikaans as a first language (Stell, 2011, p. 57), Kleurling Afrikaans is often considered the non-standard variety of the language.³ There is a long tradition of impressionistic phonetic descriptions of Afrikaans, including on the dialectal distribution of coarticulatory nasalization. The general observation is that White Afrikaans is characterized by more extensive nasalization, while nasalization is claimed to be limited or even absent in Kleurling Afrikaans (Coetzee, 1981; Coetzee, 1989, pp. 233–234; Coetzee & van Reenen, 1995; Coetzee, 1985; van Rensburg, 1989, p. 440). Although none of these earlier studies relied on acoustic or aerodynamic measures of nasalization, there is no reason to doubt the accuracy of these descriptions. Even so, one of the goals of the current study is to confirm this claimed difference based on nasal airflow measures collected from speakers of the two varieties.

In addition to noting the difference in the prevalence and extent of nasalization between the two varieties of Afrikaans, earlier research also commented on the association between nasalization

² There is also a geographic component to the dialect distribution, with most speakers of White Afrikaans concentrated in the eastern and northern provinces of South Africa, and most speakers of Kleurling Afrikaans in the western provinces (Stell, 2011, pp. 57–59). However, primarily due to segregation enforced on South African society by the apartheid system (1948–1994), even Kleurling communities in the eastern and northern regions of the country speak a variety of Afrikaans that is most closely affiliated with Kleurling Afrikaans, and similarly White communities in the western regions speak predominantly White Afrikaans (modulo smaller regional differences within each of these two socio-ethnic varieties).

³ We acknowledge the problematic nature of the terms ‘White Afrikaans’ and ‘Kleurling Afrikaans.’ The socio-ethnic groupings indicated by the terms ‘White’ and ‘Kleurling’ are problematic constructs that oversimplify the lived realities of Afrikaans-speaking individuals, so that not all speakers will associate with one of these two terms. Similarly, not everyone who may self-identify as belonging to one of these two socio-ethnic groups necessarily speaks (only) the variety of Afrikaans traditionally associated with that particular group. The terms are used here as convenient labels only to refer to two parts along what is more likely a dialect (and perhaps also style) continuum, rather than two distinct varieties of the language. Participants in the study completed a survey at the end of their participation in which they were asked to self-identify in terms of their affiliation with different parts of the Afrikaans speech community. Participants considered as speakers of White Afrikaans for the purposes of this study all self-identified as ‘White,’ while those considered as speakers of Kleurling Afrikaans typically self-identified as ‘Kleurling,’ ‘Coloured,’ or ‘Brown’ (terms that are used mostly interchangeably in South Africa). In the South-African context, the term ‘Coloured’ does not carry the same negative connotations as the term ‘Colored’ in the United States, and it is in fact often the term preferred by members of the community itself, sometimes spelled phonetically as ‘Kallit.’

and socioeconomic or educational factors. For White Afrikaans as spoken in Johannesburg, for instance, A.E. Coetzee (1989) reports more extensive nasalization for individuals in the upper than lower middle class. She also notes that, although there does not appear to be an age- or gender-related difference in the upper middle class, in the lower middle class, younger speakers and women show more extensive nasalization than older speakers or men. This indicates a possible association of nasalization with prestige and upward socioeconomic mobility. I.A. Coetzee (1985) reports similar results for the Kleurling Afrikaans community of Eersterust, not far from Johannesburg. Although he reports an overall low prevalence of nasalization (in accordance with other descriptions of Kleurling Afrikaans), he notes that nasalization rates are higher for individuals from the more affluent neighborhoods of Eersterust than those residing in the poorer neighborhoods (1985, p. 76). This difference again hints at an association of nasalization with prestige, and may also reflect the fact that individuals from the more affluent neighborhoods of Eersterust have more contact with White Afrikaans in both educational and professional settings.

The historical origin of the differences in nasalization patterns between the two varieties of Afrikaans is difficult to determine, especially given that the social valuation of nasalization in modern Dutch is opposite to that in Afrikaans (Coetzee & van Reenen, 1995; van Reenen & Coetzee, 1996). Unlike in Afrikaans, extensive nasalization is associated with non-standard and stigmatized varieties of Dutch, while lesser degrees of nasalization are found in the standard variety of the language. Coetzee and Van Reenen (1995, pp. 63–64) provide a possible explanation for the opposite valuation of nasalization in these two speech communities in terms of the historical settlement patterns of Dutch speakers who provided the input for the development of these varieties of Afrikaans. They note that the Dutch settlers who came to South Africa in the late 17th century originated from regions in the Netherlands where extensive nasalization is common today—the border regions between North and South Holland (excluding Amsterdam) and the southwestern parts of South Holland. These settlers provided the primary input for the variety that later developed into White Afrikaans. The descendants of these early Dutch settlers moved away from the Cape Town region, first to the east in the late 17th century, and eventually also north into the interior of modern South Africa after the British takeover of Cape Town in the early 19th century. Their Afrikaans therefore reflects the nasalization patterns typical of the earliest Dutch settlers. On the other hand, the non-White inhabitants of the Cape Town region for the most part did not migrate away from Cape Town. Once the Dutch settlement was firmly established by the early 18th century, Dutch settlers of higher socioeconomic status (hence coming from regions in the Netherlands where non-nasalization was the norm) came to South Africa and settled in the Cape Town region. Their variety of Dutch therefore had more influence on the development of what later became Kleurling Afrikaans.⁴

⁴ The literature on the early origins of Afrikaans is not extensive, and most of what is available is written in Afrikaans. Interested readers can refer to Den Besten (1989) and Roberge (1994) for two authoritative discussions in English.

1.4. Hypotheses

In this study, nasal airflow and eye-tracking methods are used to assess the production and perception of coarticulatory nasalization by speakers of Kleurling and White Afrikaans. We hypothesize, based on the existing impressionistic descriptions of Afrikaans, that speakers of White Afrikaans will produce more extensive coarticulatory nasalization than speakers of Kleurling Afrikaans. Our perceptual hypotheses are more nuanced and depend not only on the listeners' own socioethnic identity but also on whether they are listening to a Kleurling or White Afrikaans talker.

First, we expect that listeners will rely on coarticulatory information during perception and that their reliance will be sensitive to the time-varying patterns of that information. Given that the stimuli for our study were created in such a way that nasalization onset occurs earlier in the tokens produced by the White than by the Kleurling Afrikaans talker (see Section 3.1.2), we expect listeners to differentiate between word pairs like *bons-bos* ([bɔns]-[bɔs], 'bounce'-'forest') more quickly when listening to the White than Kleurling Afrikaans talker.

Second, given results such as those reported by Beddor et al. (2018) for English, showing a link between the extent of coarticulatory nasalization produced by an individual and that individual's perceptual reliance on nasalization, we hypothesize that a similar pattern will be found for speakers of Afrikaans. However, given the social structuring of coarticulatory variation in the Afrikaans speech community and the evidence that listeners can adjust their perceptual strategies for socially structured variation (see Section 1.2), it is also possible that the link found by Beddor et al. for English may not in fact be observed in the Afrikaans speech community under investigation.

Third, we hypothesize that listeners might adjust their perceptual strategies based on the identity of the two talkers in the perceptual task (Kleurling versus White Afrikaans). Our perceptual design (see Section 3.1.3) tests two possibilities in this regard. Listeners may bring to the perceptual task knowledge about the coarticulatory differences between the two varieties of Afrikaans, and may consequently use different perceptual strategies immediately upon identifying the specific variety. Alternatively, listeners may instead adapt their perceptual strategies over the course of the experiment, based on exposure to the coarticulatory patterns of the specific talkers in the experiment (Dahan et al., 2008; Trude & Brown-Schmidt, 2012; Trude et al., 2014; etc.).

Fourth, based on the social structure of the Afrikaans speech community, we expect potentially different perceptual results for White and Kleurling Afrikaans-speaking participants. Given the sociolinguistic situation in South Africa, speakers of Kleurling Afrikaans typically have extensive exposure to White Afrikaans. Not only is White Afrikaans the variety encountered most often in the media, it is also the variety used most often in professional and academic settings. It can thus

be assumed that speakers of Kleurling Afrikaans will have a relatively high level of exposure to White Afrikaans, which is also the variety with higher social prestige. The average speaker of White Afrikaans, by contrast, would have less extensive exposure to Kleurling Afrikaans. The situation leads to the expectation that, relative to White Afrikaans-speaking listeners, Kleurling Afrikaans-speaking listeners might have stronger prior coarticulatory expectations or be able to more rapidly adjust their perceptual strategies when listening to stimuli from the other variety. That outcome would be in keeping with the finding of Sumner and Kataoka (2013) that American English listeners do not adjust their perceptual expectations to non-prestige New York City English stimuli. On the other hand, given South Africa's racio-political history, and the consequent prominence of race and ethnicity in South African society generally, it is possible that speakers of both varieties of Afrikaans may be attentive to speech patterns related to ethnic identity and hence that speakers of both varieties will adjust their perceptual strategies to a similar extent.

2. Production experiment

Data collection for the production and perception experiments was done over two sessions, typically scheduled one week apart. Both sessions included a perception component, while production data were collected only at the end of the second session. Although the production data were collected last, we present those results first since we investigate whether nasal airflow patterns of individual speakers may predict their reliance on nasalization during perception. Based on the impressionistic descriptions of the patterns of nasal coarticulation in Kleurling and White Afrikaans, we expect both earlier onset and a higher overall volume of nasal airflow for speakers of White than Kleurling Afrikaans in the production of words that contain a nasal coda.

2.1. Methods

2.1.1. Participants

Participants were 81 native speakers of Afrikaans, between the ages of 18 and 30 years, recruited from among the student body at the North-West University, Potchefstroom, South Africa. Of the participants, 37 self-identified as 'Kleurling' (22 female, 15 male; see footnote 3) and 44 as 'White' (24 female, 20 male). All participants reported normal or corrected-to-normal vision, as well as no known speech or hearing deficits. Participants received 500 South African Rand for their participation. Twenty-six additional participants were disqualified for a variety of reasons: seven for failure to complete the full experiment, 11 for problems with accurate airflow measurement, two for poor eye-tracking accuracy, and six for poor performance in the perception task (defined as achieving less than 0.75 proportion target fixations during the time window of interest in any one of the conditions in the perception experiment).

2.1.2. Stimuli

Stimuli consisted of 10 pairs of Afrikaans words, given in **Table 1**, with the structure CVC-CVN(C), where V was either the low vowel [a] or one of the mid vowels [ɛ ɔ], C was an oral consonant (in coda position, either [t] or [s]), and N was the nasal consonant [n].

Oral CVC stimuli			Nasal CVN(C) stimuli		
<i>lat</i>	[lat]	‘whip’	<i>land</i>	[lant]	‘field’
<i>las</i>	[las]	‘joint’	<i>lans</i>	[lans]	‘spear’
<i>kat</i>	[kat]	‘cat’	<i>kant</i>	[kant]	‘lace’
<i>kas</i>	[kas]	‘cupboard’	<i>kan</i>	[kan]	‘tin can’
<i>bot</i>	[bɔt]	‘bud’	<i>bont</i>	[bɔnt]	‘multi-colored’
<i>bos</i>	[bɔs]	‘forest’	<i>bons</i>	[bɔns]	‘bounce’
<i>pot</i>	[pɔt]	‘pot’	<i>pond</i>	[pɔnt]	‘pound’
<i>pos</i>	[pɔs]	‘mail’	<i>pons</i>	[pɔns]	‘punch’
<i>pet</i>	[pet]	‘baseball cap’	<i>pen</i>	[pɛn]	‘pen’
<i>pes</i>	[pɛs]	‘pest’	<i>pens</i>	[pɛns]	‘belly’

Table 1: Stimuli used in the production study.

2.1.3. Procedure

During airflow collection, participants positioned a hand-held pliable silicone mask against their faces, with instructions to create a secure but comfortable seal. For participants with smaller faces, a large metal clip was used to pinch the bottom edge of the mask in order to ensure a tight seal. Nasal airflow was captured via the Glottal Enterprises Oral-Nasal Airflow system using a split oral-nasal silicone mask with mesh port covers and two PT-2E airflow capture transducers. Prior to each block of airflow data collection, each transducer was calibrated by pushing 140 ml of air through a calibration box attached to the transducer; air escaped through a vented-mesh port identical to those in the mask. This produced a known volume pressure signal, which was then used to calculate a conversion factor to transform the electrical pressure response of the transducer into the volume of air (in ml) passing through the mask.

Stimulus presentation and data collection were conducted using SR Research Experiment Builder software. Responses were elicited by presenting a professionally drawn black-and-white line sketch on the computer monitor. Participants were familiar with these images, since the same images were also used during the preceding perception experiment sessions. Even so, to ensure that participants produce the appropriate word, the images were accompanied by an orthographic representation of the relevant word beneath the image. Upon presentation of a stimulus, participants produced the relevant word in the frame sentence *X is die woord* ('X is the word'). Once an image had been presented, participants had two seconds to respond. Trials with incorrect productions or disfluencies were manually flagged by the experimenter for later repetition. Stimuli were presented in random order and repeated 10 times, resulting in 200 airflow samples per participant. After every 50 trials, participants were given a break and allowed to remove the mask from their faces for normal breathing.

2.1.4. Data analysis

Nasal airflow during the vowel portion of each signal was measured at 25 points across the duration of the vowel. Vowel and nasal consonant durations were also measured. Vowel and nasal boundaries were delimited using TextGrid annotations in Praat (Boersma & Weenink, 2013). As illustrated in **Figure 1** for a token of *bons* [bɔns] 'bounce,' segmentation was based on the nasal and oral waveforms, and on spectrograms that were created from the residual acoustic data captured by the airflow transducers. Signals were low-pass filtered below 5,000 Hz (to remove extraneous acoustic information) and high-pass filtered above 40 Hz (to remove the non-acoustic airflow signal). Boundaries for vowel onset and offset were placed at the first and last visually identifiable pitch pulses of the vowel and were based primarily on the oral waveform. Nasal consonant onset was identical to vowel offset, while the offset of the nasal consonant was determined largely on the basis of cessation of the periodic signal in the nasal waveform.

Despite precautions taken during recordings to minimize production errors, specific tokens were excluded from analysis due to speaker error (e.g., incorrect or disfluent production of a target word, or non-production of the carrier sentence), or an unanalyzable nasal waveform (due to mask slippage). Furthermore, to ensure that the data were not unduly affected by outliers, we applied a functional outlier detection method, from R's rainbow package, on a by-participant basis (Hyndman & Ullah, 2007; Shang & Hyndman, 2019). The outlier detection method calculates for each trial the integrated squared deviation from the mean airflow over time. Those trials that fall outside of the smallest area that captures 99% of the data were removed from further analysis. We excluded 107 trials from 34 participants (16 speakers of Kleurling Afrikaans, 18 speakers of White Afrikaans) on this basis (1.4% of the total trials).

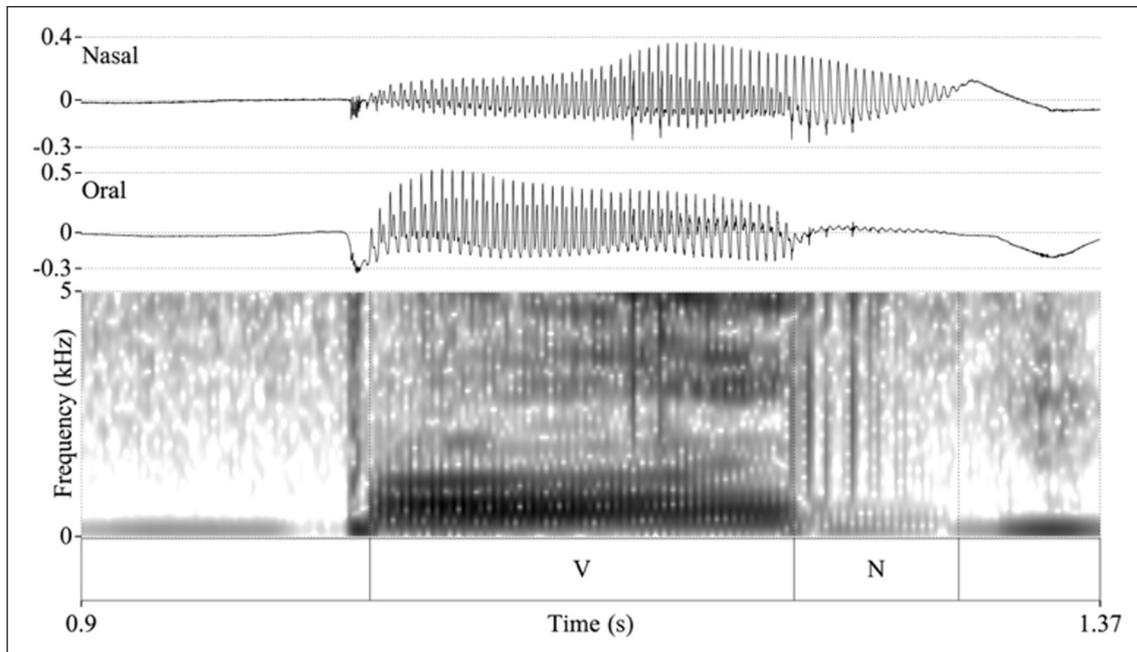


Figure 1: Nasal (top) and oral (middle) waveforms and spectrogram (bottom) for a token of *bons* [bɔns] ‘bounce.’ See text for explanation of placement of V and N boundaries.

2.2. Results

We start by briefly describing the observed nasal airflow patterns in order to confirm that the expected differences between speakers of White and Kleurling Afrikaans were obtained. We then present an analysis of the nasal airflow patterns relying on Generalized Additive Mixed Modeling (GAMM) to capture the dynamic changes in nasal airflow over time. Finally, we conduct a functional principal component analysis of the airflow data to capture differences between speakers. The first principal component from this analysis will be used later (see Section 3) to make speaker-level predictions about perceptual reliance on nasal coarticulation.

The left panel in **Figure 2** presents the raw average nasal airflow across normalized time for CVN(C) tokens, separately for speakers of Kleurling and White Afrikaans. As this image shows, the onset of nasal airflow is earlier and the overall volume is greater for the speakers of White Afrikaans.

Before statistical modeling, airflow measures were normalized on a by-trial basis by dividing each of the 25 raw nasal airflow measures in the vowel by the maximum nasal airflow value attested within the following nasal consonant. The normalized nasal airflow values therefore constitute the ratio of nasal airflow during the vowel to maximum nasal airflow during the nasal

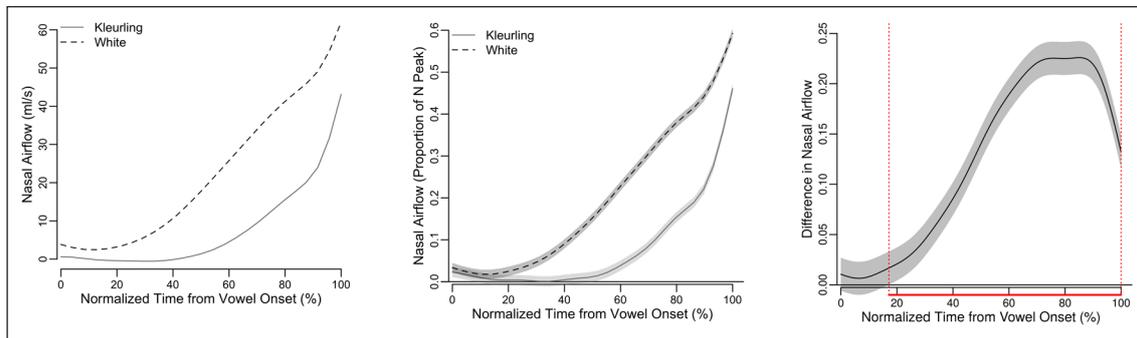


Figure 2: Left: Mean nasal airflow across normalized time for speakers of White (black dashed line) and Kleurling (solid grey line) Afrikaans. Middle: Model predicted nasal airflow across normalized time for speakers of White and Kleurling Afrikaans, with 95% confidence bands for each curve. Right: Model predicted differences in nasal airflow between White and Kleurling Afrikaans, with 95% confidence band (shaded region). Differences were calculated by subtracting predicted values for Kleurling Afrikaans from that for White Afrikaans; positive values indicate more airflow for White than Kleurling Afrikaans. The region of significant difference (where the confidence interval is above zero) is marked in red and bounded by dotted lines.

consonant. These normalized values adjust for both across-speaker differences and between-trial differences within a speaker. To give an indication of the structure of and variation in the non-normalized data, we report measures of vowel length and nasal airflow in **Table 2**. As seen in this table, vowel length was marginally longer in White than Kleurling Afrikaans, and the peak nasal flow was higher in White Afrikaans in both the nasal coda consonant and the pre-nasal vowel.

		Kleurling Speakers		White Speakers	
		Mean	SD	Mean	SD
Vowel duration (ms)	CVN(C)	146.6	40.5	155.5	41.1
	CVC	125.2	30.7	137.5	33.7
Peak nasal flow (CVN(C) only) (ml/s)	Vowel	45.3	29.3	64.8	37.2
	Nasal	95.6	48.8	103.0	45.5

Table 2: Mean vowel durations (in ms) of vowels in CVC and CVN(C) tokens, and peak nasal airflow (in ml/s) in nasal codas and pre-nasal vowels in CVN(C) tokens, as produced by speakers of White and Kleurling Afrikaans.

The normalized airflow measures were subjected to Generalized Additive Mixed Modeling (GAMM) analyses, using the *mgcv* (Wood, 2019) and *itsadug* (van Rij, Wieling, Baayen, & van

Rijn., 2020) packages in R (R Core Team, 2020). GAMMs make no assumptions about the shape of continuous data, fitting the data to a sum of splines (smooth functions). Fixed effects in the model included participant Ethnicity (White versus Kleurling), Normalized Time (as a smooth), and their interaction. Participant-wise smooths for Word were entered as a random effect. We corrected for autocorrelation observed among the predicted nasal flow values (typical in time-series data) by running a further autoregressive AR(1) model, to account for dependence among these values.

The model predictions are shown in the middle panel in **Figure 2**. Comparison with the raw flow patterns in the left panel confirms that the model accurately captures the overall shape of the curves and the nasalization differences between the productions of speakers of Kleurling and White Afrikaans. While significant differences between the curves in the middle panel are evident where their confidence intervals do not overlap, the panel on the right plots the model-predicted difference between Kleurling and White Afrikaans with 95% confidence bands revealing significant differences in nasal airflow from as early as 20% of the duration of the vowel. This provides the first empirical support for earlier impressionistic descriptions of Afrikaans that have claimed more extensive nasal coarticulation for White than for Kleurling Afrikaans (see Section 1.3).

Finally, in order to reduce the number of dimensions required to describe airflow differences across different speakers, we follow Beddor et al. (2018) in submitting participants' normalized nasal flow patterns to a functional principal components analysis (fPCA) using functions from the *fda* package in R (Ramsay, Wickham, Graves, & Hooker, 2020). fPCA represents data points sampled over time as smooth functions (splines), and extracts independent (orthogonal) modes of variation (harmonics) among the functions. The first principal component (henceforth PC1) accounted for 92% of the variance in the data.

The left panel in **Figure 3** shows the mean predicted nasal airflow (solid line), with predicted airflow for speakers with PC1 values one standard deviation above and below the mean (broken lines). Higher PC1 values correspond both to earlier onset of nasal airflow, and higher overall volume of flow. As further confirmation of this relation between PC1 and the time course of nasal airflow, the right panel plots the average normalized nasal airflow for the five participants in our study with the highest and lowest PC1 values, respectively. As seen in this figure, speakers with low PC1 values have a very late onset of nasal airflow, typically near the vowel offset, indicating that these speakers have very little if any nasal coarticulation. On the other hand, speakers with high PC1 values have nasal airflow through all or most of the vowel, indicating substantial nasal coarticulation.⁵

⁵ Ingressive nasal airflow in the first part of the vowel (here especially for speakers with low PC1 values) has been documented in other aerodynamic studies of coarticulatory nasalization (e.g., Delvaux, Demolin, Harmegnies, & Soquet, 2008). We confirmed that these measures do not reflect measurement error by inspecting the nasal airflow

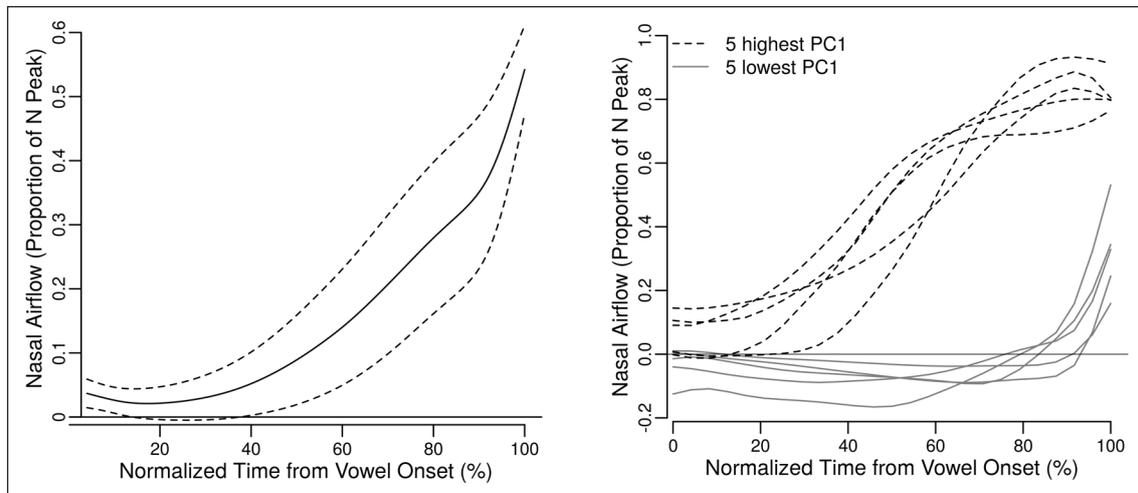


Figure 3: Left: Mean normalized predicted nasal airflow for speakers with a PC1 value of 1 standard deviation above and below the mean (upper and lower dashed lines). The solid line in the middle represents the average normalized nasal airflow across all speakers. Airflow was normalized as explained in the text by taking the ratio of nasal airflow in the vowel to the maximum nasal airflow in the following nasal consonant. Right: Average normalized nasal airflow for the five speakers with the lowest (solid grey lines) and highest (broken black lines) PC1 values.

To investigate the observed difference between White and Kleurling Afrikaans further, we order the participants in **Figure 4** by their PC1 values. Speakers of White Afrikaans (black bars) tend to cluster towards the higher endpoint of PC1 values, while speakers of Kleurling Afrikaans (grey bars) tend to cluster toward the lower endpoint. This figure also shows that there is more variation overall among speakers of Kleurling than White Afrikaans—speakers of Kleurling Afrikaans are found across nearly the full range of PC1 values, while no speakers of White Afrikaans are found in the lower 20–25% of the PC1 range.⁶

for these same speakers in oral CVC words, and noting that there is no evidence of ingressive nasal airflow during their productions of CVC words. We hypothesize that ingressive nasal airflow may be the result of the lowering of the velum before the velic seal is broken, slightly increasing the volume of the nasal cavity, and hence resulting in weak ingressive nasal airflow. See Hayes and Stivers (2000) for evidence that such ‘pumping action’ of the velum can result in measurable ingressive airflow.

⁶ Speakers (and especially speakers of non-standardized/stigmatized language varieties) ‘style-shift’ based on the specific social communicative setting in which their language use occurs—see Scanlon and Wassink (2010) and Britt and Weldon (2015) about such style-shifting in African American English, for example. The PC1 values used here should therefore be interpreted as reflecting the nasalization patterns typical of these speakers in the specific social communicative setting in which the data were collected—that is, a fairly formal setting on a university campus where White Afrikaans is the majority language variety and the assumed standard. It is therefore possible that those Kleurling speakers with PC1 values typical of White Afrikaans may be style-shifting to accommodate to the specific social communicative setting of the experiment and that they may nasalize less in settings where White Afrikaans is not the social normative variety of the language (Coetzee, 2018, p. 188).

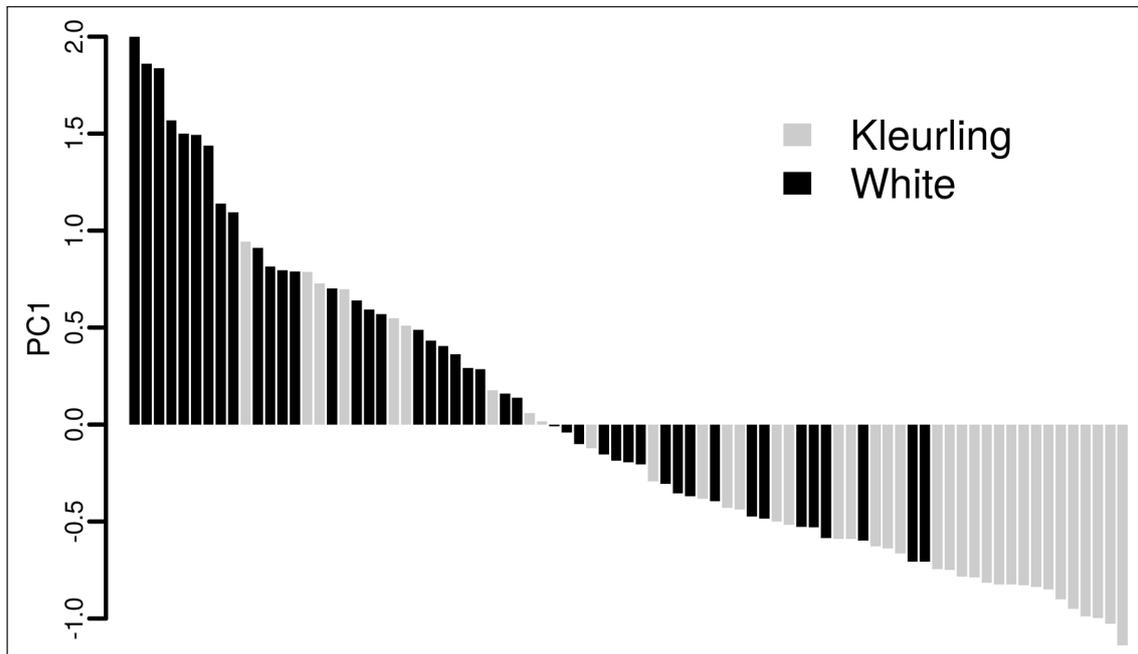


Figure 4: Kleurling (grey bars) and White (black bars) Afrikaans-speaking participants ordered by PC1 values.

3. Perception Experiment

The perception experiment assessed listeners' perceptual reliance on the presence versus absence of coarticulatory nasalization using an eye-tracking design similar to that used by Beddor et al. (2018; see also Beddor, McGowan, Boland, Coetzee, & Brasher, 2013). In this experiment, listeners were presented with an auditory CVC (*kat* [kat] 'cat,' *pet* [pet] 'baseball cap') or CVN(C) (*kant* [kant] 'lace,' *pen* [pen] 'pen') stimulus, and two images corresponding to the presented auditory stimulus and its minimal pair competitor (*kat-kant*, *pet-pen*). Participants' task was to look at the image corresponding to the auditory stimulus. Auditory stimuli were produced either by a White or Kleurling Afrikaans talker, with relatively minor manipulation so that all tokens would show coarticulatory patterns typical of these two varieties of the language (see Section 3.1.2 for more on these manipulations).

For each talker condition (Kleurling or White Afrikaans), stimuli were presented according to a blocked design in which participants first heard only CVC auditory stimuli followed by CVC and CVN(C) stimuli intermixed (with both blocks also containing fillers). This design allowed us to test the two versions of our third perception hypothesis (Section 1.4) concerning adjustment of perceptual strategies based on talker identity. If listeners use prior knowledge about coarticulatory nasalization in the two varieties when hearing, say, *kat* and deciding between *kat* and *kant*, they should look more quickly to the *kat* image when hearing the token produced by the White

Afrikaans talker than the Kleurling Afrikaans talker, even prior to hearing *kant*. This is because, on average, orality disambiguates *kat* and *kant* early in the vowel for White but not for Kleurling Afrikaans. Alternatively, if listeners only adapt their perceptual strategies over the course of the experiment, differences in responses to the stimuli produced by the two talkers should not emerge until after listeners hear CVN(C) stimuli.

3.1. Methods

3.1.1. Participants

The participants were the same individuals as those who participated in the production experiment.

3.1.2. Stimuli

Stimuli were the same 10 CVC-CVN(C) minimal pairs that were used in the production experiment. Auditory stimuli were modified versions of the words as produced by two adult female talkers, one Kleurling Afrikaans talker and one White Afrikaans talker. In order to select talkers who could easily and reliably be identified as speaking the relevant variety of Afrikaans, we first conducted a talker norming experiment using the voices of 11 Kleurling Afrikaans (5 female, 6 male) and 13 White Afrikaans (8 female, 5 male) individuals, each reading the instruction sentences used during the eye-tracking experiment (see Section 3.1.3). These recordings were presented, through an online interface and in random order, to 19 Afrikaans listeners who were tasked with identifying the variety of Afrikaans spoken by each talker. The talkers' variety was generally identified accurately, with an average of 94% correct and a range of 86 to 100%. From the 19 talkers, we selected one female talker of each variety for whom their variety was correctly identified by 100% of the participants.

In addition to the instruction sentences, these two talkers also produced the CVC and CVN(C) target words (in addition to some filler words). To ensure that the stimuli consistently had the coarticulatory nasalization patterns typical of Kleurling and White Afrikaans, the original stimuli were waveform edited in Praat. For each minimal CVC-CVN(C) word pair (*kat-kant*), the initial C and onset of V were taken from a token of the CVC word. To create the CVC stimulus (*kat*), this initial portion (ka_{onset}) was then spliced onto the $V_{\text{offset}}C$ of a different token of the relevant CVC word ($a_{\text{offset}}t$). The corresponding CVN(C) stimulus (*kant*) was created by using the same initial portion (ka_{onset}), and splicing that onto the $V_{\text{offset}}N(C)$ portion of the relevant CVN(C) token ($\tilde{a}_{\text{offset}}nt$ from *kant*). Splicing was done such that approximately the last 75% of the vowel was realized with nasalization in the White Afrikaans tokens, and approximately 20% for the Kleurling Afrikaans tokens. This editing (typically involving only a few pitch pulses per vowel) resulted in tokens with coarticulatory patterns characteristic of the two relevant varieties of Afrikaans. For all nasal vowel portions, nasalization was clearly audible, with acoustic correlates of the nasalization being a decrease in waveform amplitude and a flattening and broadening of

the F1 region of FFT spectra, relative to the oral portion. **Table 3** contains average durations of the oral and nasal portions of the vowel, and the nasal consonant in CNV(C) tokens in each of the two varieties. Filler stimuli were 10 minimal pairs differing in oral codas (e.g., *tas* /tas/ ‘suitcase,’ *tak* /tak/ ‘branch’).

This splicing results in stimuli in which the temporal onset of nasalization (the main difference of interest between White and Kleurling Afrikaans in this study) is carefully controlled. An alternative approach could have been used in which naturally produced tokens with the requisite timing patterns were selected as stimuli. We opted to use the splicing methodology in order to have more exact control over both the pre-nasalization portion of the stimuli (i.e., so that there would be no other information on which listeners might base their target looks) and the temporal onset of nasalization in the stimuli.

	Oral portion of vowel	Nasal portion of vowel	Nasal consonant
White Afrikaans	41	123	101
Kleurling Afrikaans	101	31	128

Table 3: Average durations (in ms) of relevant portions of the vowel in CVC and CVN(C) tokens, and of the nasal consonants in CVN(C) tokens.

Visual stimuli were black and white line drawings corresponding to each of the 40 words (20 target stimuli and 20 fillers), which were used as prompts in both the production and perception studies.

3.1.3. Procedure

Data collection for the perception experiment was done over two sessions, usually scheduled a week apart. Talker identity was blocked by session such that each session consisted of only tokens produced by the Kleurling or White Afrikaans talker. The order of sessions was counterbalanced across participants. In the first session, prior to testing, participants learned the labels for each of the target images used for the eye-tracking study (and also the production study reported in Section 2). Participants first saw the randomly ordered images one at a time, with the corresponding word written below the image. To aid memorization, they read each label aloud to the experimenter and explained how the image related to the label. Participants were then shown, in a self-paced procedure, each of the images in random order, and had to produce the word corresponding to the image aloud. Each image had to be identified correctly twice before moving on to the main task. An incorrect answer resulted in the correct label being shown on the screen and the word being reentered into the randomization. The testing part of this familiarization procedure was repeated at the start of the second data collection session.

Eye movements were captured with a remote monocular eye-tracker (EyeLink 1000 Plus, SR Research), using a 25 mm lens and a sampling rate of 500 Hz. Participants were seated so that their eyes were between 550 and 650 mm from the camera and about 800 mm from the monitor. During testing, auditory and visual stimuli were presented using SR Research Experiment Builder software; auditory stimuli were heard over AKG 271 Mk2 headphones. After familiarization but prior to testing, the experimenter performed a calibration.

In each test trial, participants were presented with two visual stimuli, arranged as in **Figure 5**. Participants then heard the instruction sentence (*Kyk na die sketse* ‘Look at the drawings’), as produced by the White or Kleurling Afrikaans talker. After 2.5 seconds, a fixation cross appeared along with the instruction sentence to *Staar na die kruis* ‘Stare at the cross.’ One second later, the cross disappeared as the participant heard *Fokus nou op ...* ‘Focus now on ...,’ followed half a second later by the target auditory stimulus produced by the same talker. The trial ended two seconds later. Before presentation of the test trials, participants responded to 10 practice trials, consisting of fillers only.



Figure 5: Screenshot of a trial with visual stimuli for *bos* [bɔs] ‘forest’ and *bons* [bɔns] ‘bounce.’

Each of the two perception sessions consisted of 160 trials of three different types. In oral auditory trials, the target (auditory) stimulus was an oral word (e.g., *bos* [bɔs] ‘forest’), and the visual stimulus consisted of the corresponding visual image of the oral word, paired with a distractor image of the corresponding nasal word (e.g., *bons* [bɔns] ‘bounce’). In nasal auditory trials, the target visual stimulus consisted of an image of the relevant nasal word and the distractor image was of the corresponding oral word. Each of the 10 oral and 10 nasal tokens was presented five times, for 50 oral and 50 nasal auditory trials each. In filler trials, an oral filler word (e.g., *gaar* [xɑ:r] ‘cooked’) was presented auditorily with images corresponding to the auditory token and a minimal pair oral competitor (e.g., *gaas* [xɑ:s] ‘screen’). The 10 filler tokens were each presented six times, for 60 filler trials.

Stimuli were organized into two blocks: The initial block was an ‘oral only’ block, and contained 40 oral target stimuli and 35 filler auditory stimuli, followed by a ‘mixed’ block containing 10 oral, 50 nasal, and 25 filler trials. Two stimulus randomizations were created (without mixing stimuli from the mixed and oral only blocks), and were alternated between consecutive participants. Participants assigned to the first of the two orderings for the first perception experiment session were then assigned to the other ordering for the second session (i.e., a participant always heard different orderings for the White and Kleurling Afrikaans sessions, respectively). Participants were given a short break after every 50 eye-tracking trials.

Participants’ eye movements were monitored during each trial, starting from the onset of the auditory stimulus and for a duration of 1000 ms. The computed measure was the proportion of fixations on the target image over time, beginning at 200 ms after stimulus onset, and for forty 20 ms temporal bins. The 200 ms delay is based on the standard assumption of the time required for the planning and execution of a saccade (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; for a review of the cognitive bases for this delay, see Hutton, 2008). A fixation was counted as a target fixation if it fell within the target image’s ‘square’ (as in **Figure 5**). Thus, a proportion of 0.50 for, say, the temporal bin 400–420 ms for auditory *bons* in *bos-bons* trials means that 50% of those trials included a fixation on visual *bons* at some point during that 20 ms interval.

3.2. Results

Results from the eye-tracking experiment were modeled with generalized additive mixed models (GAMMs) implemented in *R* (R Core Team, 2020) using the *mgcv* (Wood, 2019) and *itsadug* (van Rij et al., 2020) packages. Since fixations are binomial (a participant either does or does not fixate on the target), a logit link function was used in the regression. To investigate our hypothesis of a link between production and perception patterns at the level of the individual, we ran a model that included as fixed factors the PC1 values from the production experiment (to represent the extent of coarticulatory nasalization produced by an individual speaker), Speaker Ethnicity (Kleurling, White), Participant Ethnicity (Kleurling, White), Auditory Target Nasality (Nasal/CVN(C), Oral/CVC), and Block (Oral Only, Mixed). These four factors were coded as an interaction variable for ease of incorporation into the model. The model also contained two fixed smooths for Time and PC1. All fixed factors were fully interacted. Random Word-specific smooths for Participant over Time were also included. To avoid artefactual overfitting of the data, the non-linearity penalty, gamma, was increased to double its default value (see Baayen, Vasishth, Kliegl, & Bates, 2017; Wood, 2011). The remainder of our hypotheses are independent of the assumed link between perception and production, and were hence evaluated with a model that was identical to that described above, except that PC1 and its interactions with other factors were not included in the model. The full model structures and results are available in the supplementary materials.

3.2.1. Do listeners use coarticulatory nasalization?

Our first hypothesis is that listeners will rely on coarticulatory vowel nasalization to differentiate CVN(C) and CVC words (e.g., *kant* versus *kat*) and will not wait for the disambiguating post-vocalic consonantal information (*-nt* versus *-t*). Since coarticulatory nasalization starts earlier in the CVN(C) words produced by the White than Kleurling Afrikaans talker, we expect that participants will fixate on the target CVN(C) image earlier in the White Afrikaans condition than in the Kleurling Afrikaans condition. The panels in the top row of **Figure 6** show the average observed fixations over time, starting 200 ms after vowel onset; the left and right panels show the results for Kleurling and White Afrikaans-speaking listeners, respectively. That both groups of listeners fixated earlier on the target CVN(C) image when listening to the White Afrikaans talker than the Kleurling Afrikaans talker can be seen in the top panels and is confirmed by the GAMM. Panels in the middle row of **Figure 6** show model-predicted patterns with 95% confidence bands, and those in the bottom row show the difference between the model-predicted patterns in log odds (with the same confidence bands) for the two talker conditions.⁷ Analysis of the binomial data requires use of the logit link function, and as such the model predicts log odds of fixation on the target. Correspondingly, comparison of model predictions for pairs of conditions is on this scale. A difference of zero between conditions on the log odds scale indicates that target fixations were equally likely in the conditions being compared. A positive difference indicates more fixations on the target image for stimuli produced by the White than by the Kleurling Afrikaans talker. Regions of significant difference are delineated in red in the middle and bottom panels.⁸ Starting between 250 and 300 ms after vowel onset, both Kleurling and White Afrikaans listeners fixate more on the CVN(C) targets produced by the White Afrikaans than by the Kleurling Afrikaans talker. This difference persists up to the end of the 1000 ms time period for Kleurling Afrikaans listeners and up to 800 ms for White Afrikaans listeners. For White Afrikaans listeners, the final 100 ms show the opposite pattern (more looks for the Kleurling than for the White Afrikaans talker)—an inversion of look patterns that arguably reflects the possibility that some listeners

⁷ In order to isolate the predicted effects of the fixed factors in the model, these and all similar plots later in the paper show model-predicted fixation proportions for conditions excluding effects attributable to participant- and word-specific variation, as specified in the model random effect structure. In the relevant plots, this is indicated by the words “fitted values, excl. random” in the righthand margin of the plots.

⁸ Results in the difference plots (bottom panels) of this and following figures are given in log odds, since responses are binary (a participant either looks at an image or not), so that we have to rely on regression models with a logit link function to model participant looks. Log odds should be interpreted carefully given that the relationship between log odds and proportions is not linear. The same size change in log odds (a change of one unit from 0 to 1, and from 1 to 2) can correspond to very different size changes in proportion (here a difference of 0.23 from 0.5 to 0.73, and a difference of 0.15 from 0.73 to 0.88). The R packages used to model the data in this paper do not have the functionality to back-transform modeled differences in log odds to proportions, and we therefore opt to represent the difference plots in log odds rather than proportions. Readers interested in transforming the log odds to proportions can use the following formula: $(\exp(\text{logodds})) / (1 + \exp(\text{logodds}))$.

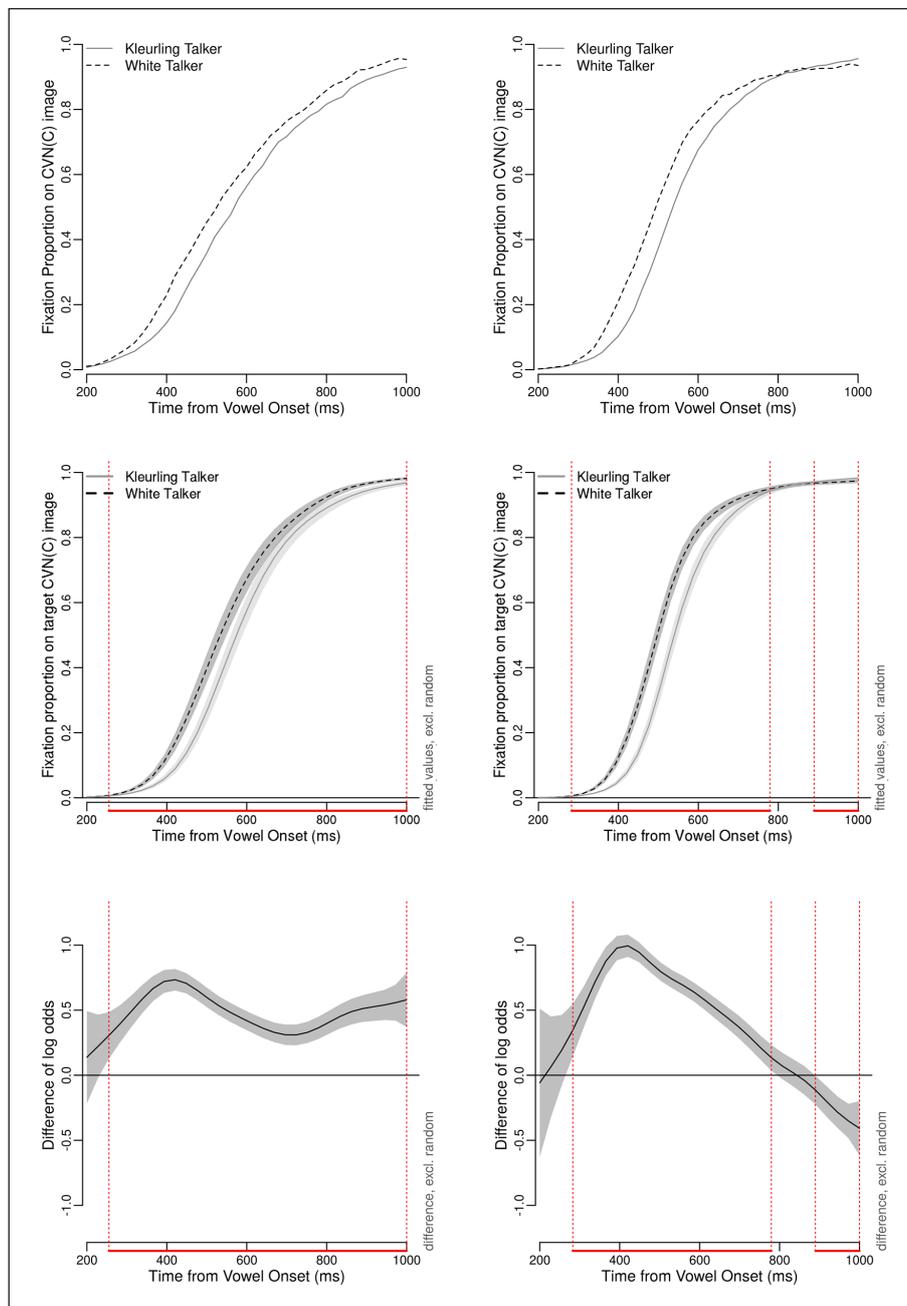


Figure 6: Fixations over time on the target CVN(C) image for Kleurling Afrikaans (left) and White Afrikaans (right) listeners. Top row: observed average fixations over time, starting 200 ms after vowel onset. Middle: model-predicted looks with 95% confidence bands. Bottom: model-predicted fixation differences for the Kleurling and White Afrikaans stimuli (with positive differences indicating more fixations for the stimuli produced by the White Afrikaans talker). Middle and bottom rows: Temporal regions marked in red are regions of significant difference between fixations for White and Kleurling Afrikaans stimuli.

have completed the task for the White Afrikaans talker (where nasalization onset is earlier) and are hence starting to look away from the target image for this talker towards the end of the trial (see Beddor et al., 2018, pp. 954–955, for a similar ‘look away’ pattern).

3.2.2. Do individuals’ patterns of produced coarticulatory nasalization predict their perceptual reliance on coarticulatory nasalization?

Having established that both Kleurling and White Afrikaans-speaking listeners rely perceptually on acoustic information for coarticulatory nasalization, we turn to the question of whether there is a relation between an individual’s production of coarticulatory nasalization and that same individual’s perceptual use of this information. To investigate this question, we rely on a GAMM that includes PC1 (and its interactions) as a fixed factor. As shown in Section 2.2, higher PC1 values correspond to speakers with both earlier onset and higher overall volume of nasal airflow in CVN(C) words. If an individual’s produced coarticulatory nasalization predicts their perception, we would expect participants with higher PC1 values to fixate earlier on nasal CVN(C) target images. Based on the general assumption of a production-perception link, this effect is expected to hold irrespective of whether listeners are attending to the speech of the Kleurling or White Afrikaans talker.

The results given in **Figure 7** and **Figure 8** assess the PC1 effect when participants respond to the Kleurling and White Afrikaans talker, respectively. These figures show observed and predicted fixation patterns for individuals whose PC1 values fall around the 75th and 25th percentile of observed PC1 values, separately for the Kleurling (left panels) and the White (right panels) Afrikaans-speaking participants. We use the 75th and 25th percentile of PC1 values to represent high and low PC1 values since more extreme PC1 values (in both directions) are more sparsely distributed across the PC1 range, such that model estimates may be less accurate for these more extreme PC1 values. Panels in the top row show the average observed fixations for participants whose PC1 value falls within the 15% range of PC1 values centered around the 75th and 25th percentiles for each listener group. The middle row shows the model predicted fixations on the target CVN(C) image, including 95% confidence bands, and the panels in the bottom row provide the model-predicted difference in fixations for a participant with a PC1 value at the 75th and 25th percentile, calculated such that a positive difference indicates more looks at the target-image at that time point for a participant at the 75th percentile. Regions of significant difference are marked as before.

Inspection of **Figure 7** shows that, when listening to the White Afrikaans talker, participants with higher PC1 values (represented by the 75th percentile) fixated on the CVN(C) target image earlier than those with lower PC1 values (the 25th percentile), and that this effect is observed for Kleurling (on the left) and White (on the right) Afrikaans listeners. The inversion in the proportion of fixations observed later in the 1000 ms time period, especially for the Kleurling Afrikaans listeners, can most likely be attributed to participants with higher PC1 values having completed the task and hence looking away from the target image.

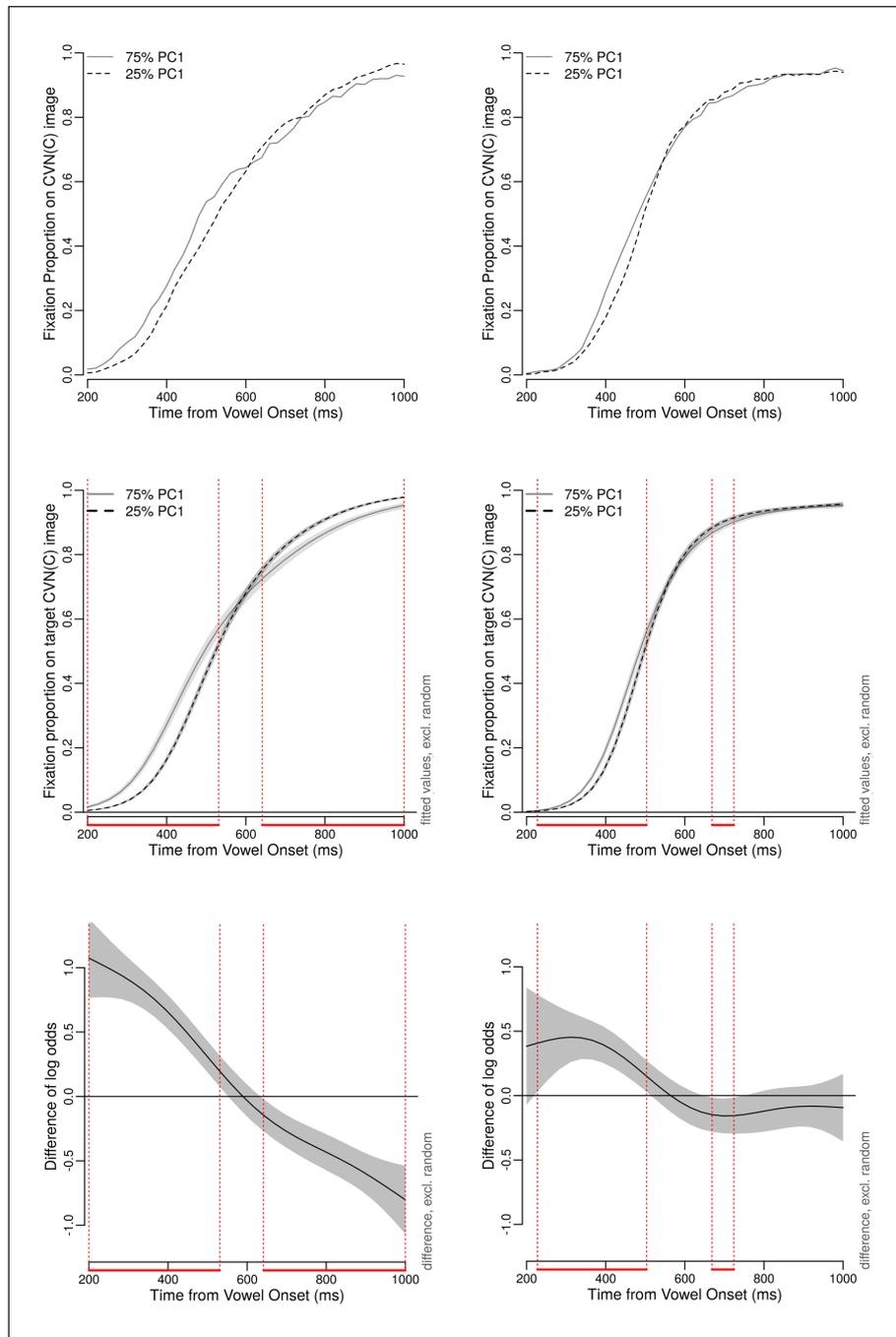


Figure 7: Kleurling (left) and White (right) Afrikaans-speaking listeners' fixations over time on the target CVN(C) image for tokens produced by the White Afrikaans talker. Top panels: average observed fixation proportions for participants whose PC1 values fall within the 15% range of PC1 values centered around the 25th and 75th percentile of observed PC1 values. Middle panels: model-predicted fixations for listeners whose PC1 value falls at the 25th and 75th percentile of observed PC1 values within each listener group. Bottom panels: model-predicted differences between listeners at the 25th and 75th percentile of observed PC1 values (positive difference: more fixations on the target CVN(C) image for a listener at the 75th percentile). Middle and bottom panels include 95% confidence bands; red lines indicate regions of significance.

Figure 8 shows the results for perceptual responses to the Kleurling Afrikaans talker, and the same patterns arise here as for the White Afrikaans talker in **Figure 7**. These results therefore provide support for the hypothesis that individuals who produce more extensive coarticulatory nasalization in CVN(C) tokens also rely more on coarticulatory nasalization perceptually, and hence add to the evidence for a link between individuals' production and perception patterns.

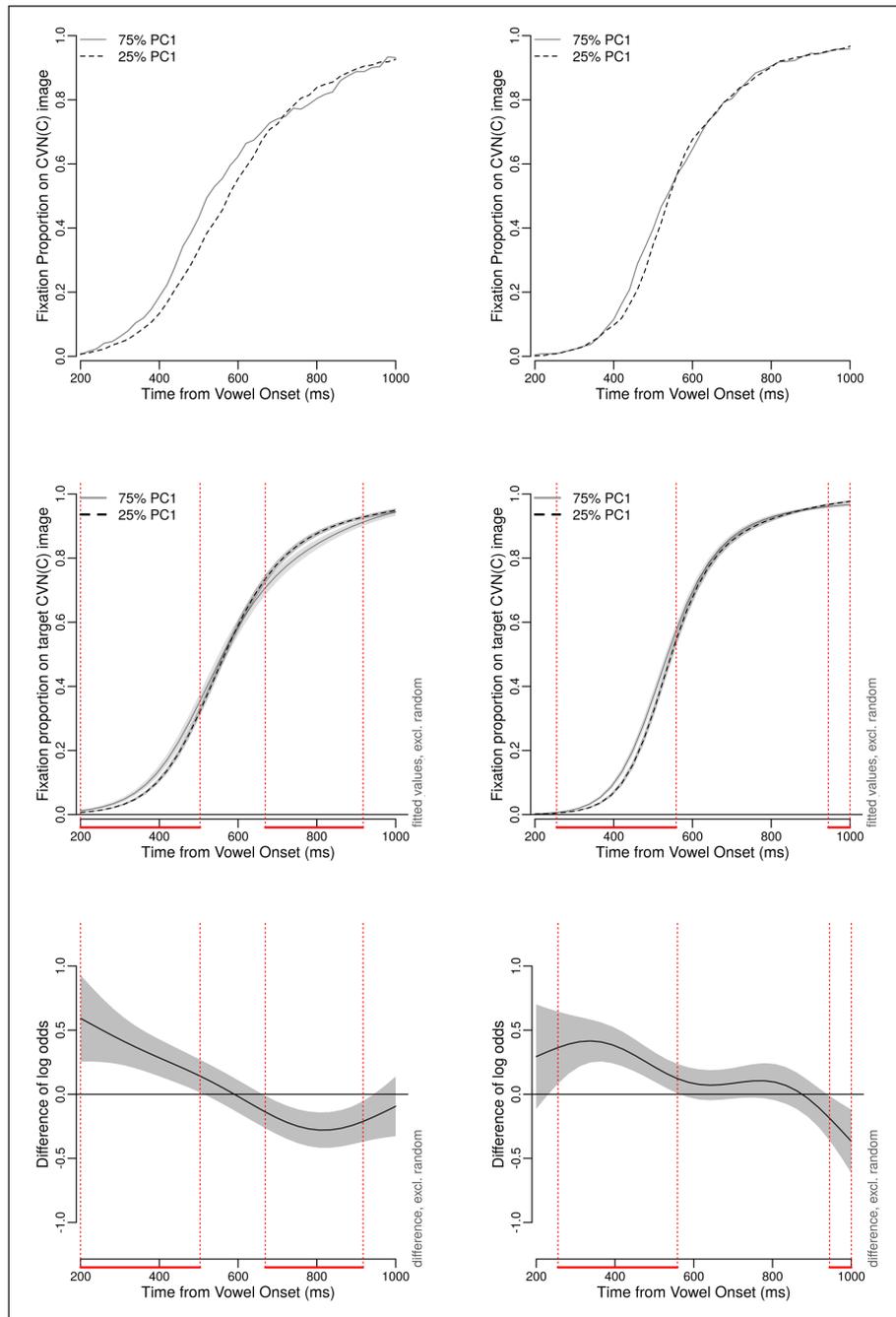


Figure 8: Kleurling (left) and White (right) Afrikaans-speaking listeners' fixations over time on the target CVN(C) image for tokens produced by the Kleurling Afrikaans talker. The structure of the panels in the top, middle, and bottom rows is the same as in Figure 7.

3.2.3. Do listeners adapt to the coarticulatory nasalization patterns of the talker?

As reviewed in Section 1.2, under certain conditions, listeners can rapidly adjust their perceptual strategies to the acoustic patterns present in the speech of a specific talker. In our study, stimuli were blocked such that participants were presented with only oral CVC auditory targets in the initial part of the experiment (Oral Only block), and with both oral CVC and nasal CVN(C) targets in the second part (Mixed block). It is therefore only in the Mixed block that participants get information about differences in the timing of coarticulatory nasalization for the two talkers (earlier onset for the White Afrikaans talker). Once participants reach the Mixed block they could hence, based on the timing patterns of nasalization in the White Afrikaans tokens, identify a token as CVC rather than CVN(C) with confidence relatively early in the vowel of the CVC token (if the vowel is still fully oral about 25% into the vowel, it can only be the oral CVC token). Consequently, we might expect earlier fixations on the CVC target image in the Mixed than Oral Only blocks for the White Afrikaans talker. Conversely, for the Kleurling Afrikaans talker, participants will get information in the Mixed block that CVC and CVN(C) tokens are ambiguous up to the very end of the vowel (due to late onset of nasalization in CVN(C) tokens). Minimally, no change in the speed of looks to the CVC target images would be expected in Kleurling Afrikaans condition. It is also possible that confirmation of the late disambiguation between CVC and CVN(C) tokens in the Kleurling Afrikaans condition can result in additional uncertainty on the part of the listeners, which could lead to a slow-down in fixations on the CVC target images in the Mixed versus Oral Only Block in the Kleurling Afrikaans conditions.

As with earlier hypotheses, this hypothesis is assessed using a GAMM. **Figure 9** shows the fixations in response to the Kleurling talker's CVC stimuli, with patterns for Kleurling Afrikaans listeners in the left panels and for White Afrikaans listeners in the right panels. Panels in the top row show the observed average target fixations over time in response to auditory CVC tokens in the Oral Only and Mixed blocks. The middle panels show model-predicted fixations, and the bottom panels show model-predicted differences calculated such that a positive difference indicates more target fixations in the Oral Only block (i.e., later looks in the Mixed than Oral Only block). (Regions of significant difference are marked in the middle and bottom rows as before.) As inspection of this figure shows, the predicted slow-down in target fixations is observed for both the Kleurling and White Afrikaans listeners.

Figure 10 shows the corresponding patterns in response to the stimuli of the White Afrikaans talker, where we expect to find earlier fixations on the oral CVC image in the Mixed compared to the Oral Only block. The pattern of results here is less clear. For the Kleurling Afrikaans listeners, we find a momentary difference in the predicted direction at around 250 ms, then a difference in the opposite-to-predicted direction between about 400 and 500 ms, and then again a difference in the expected direction after 800 ms. For the White Afrikaans listeners, the only significant difference is in the opposite-to-expected direction, late in the trial after 800 ms. The opposite-to-expected patterns, and in particular the flip in patterns observed for the Kleurling Afrikaans listeners, are difficult to explain. The observed patterns in this condition, however, do not in general provide support for the hypothesized speed-up in target fixations in response to CVC auditory tokens.

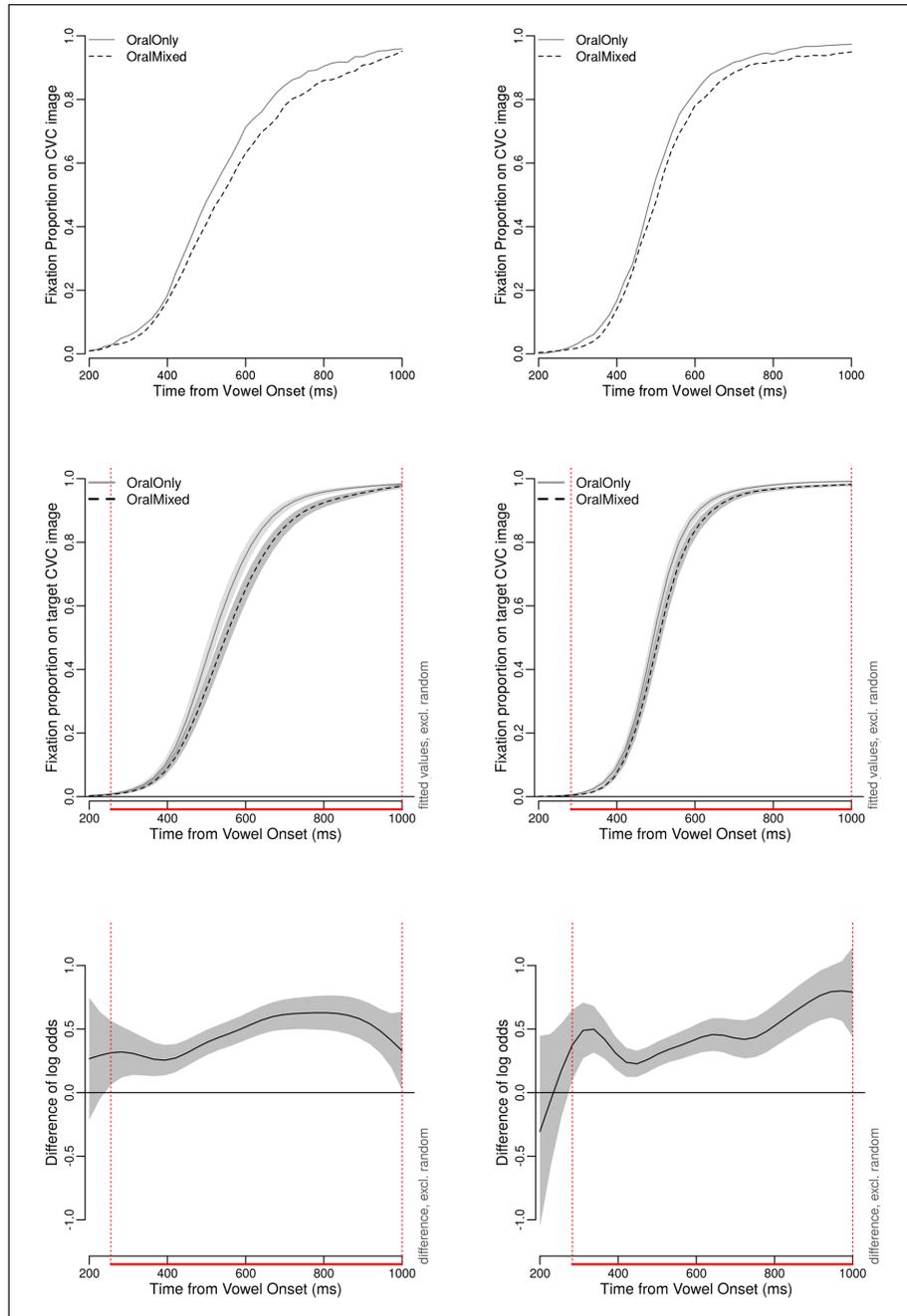


Figure 9: Kleurling (left) and White (right) Afrikaans-speaking listeners’ fixations over time on the target CVC image in the Oral Only and Mixed blocks for stimuli produced by the Kleurling Afrikaans talker. Top: observed average fixations. Middle: model-predicted fixations. Bottom: model-predicted differences calculated such that a positive difference indicates more target fixations in the Oral Only block (i.e., a slow-down). Middle and bottom panels include 95% confidence bands; temporal regions marked in red are regions of significant difference.

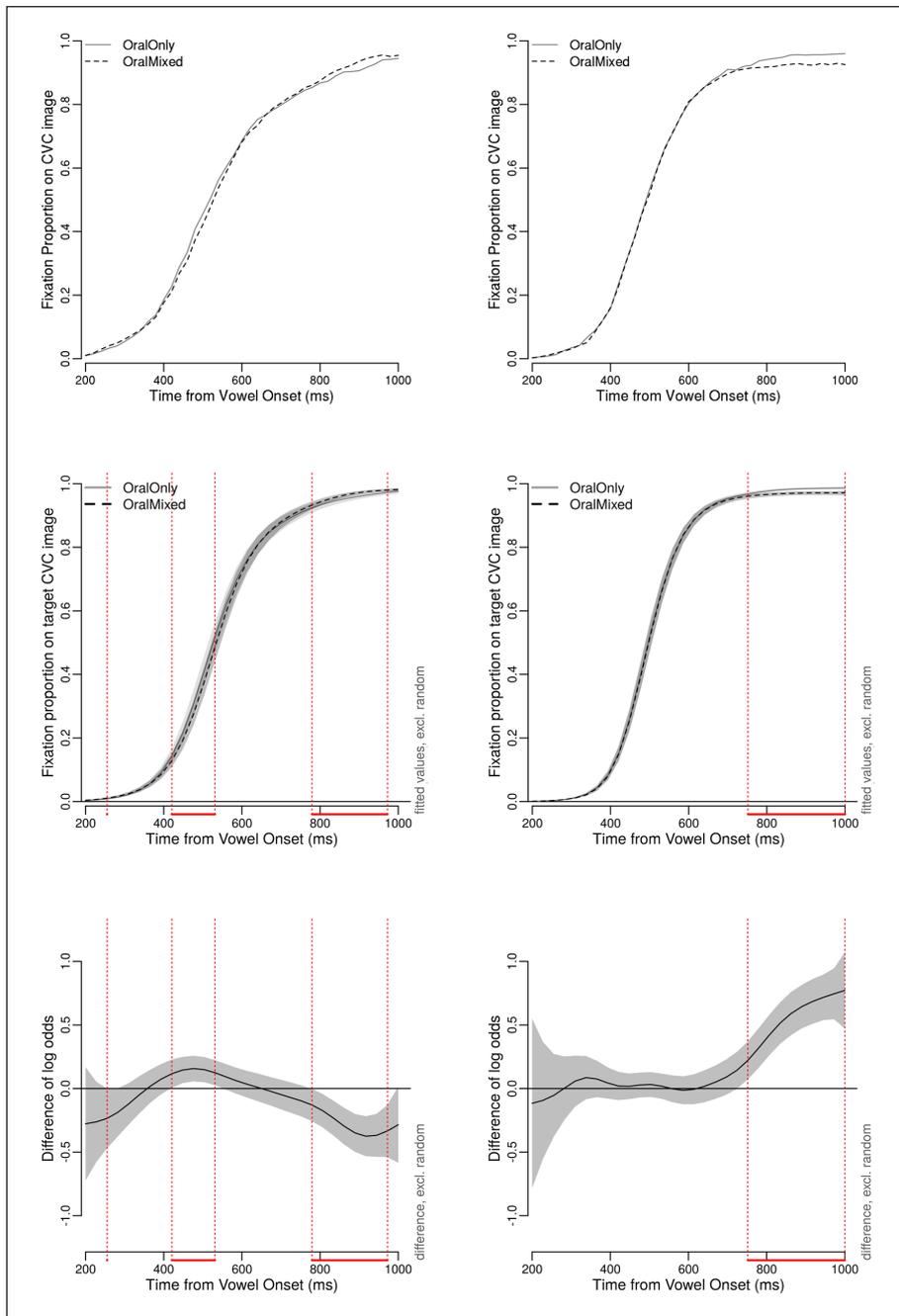


Figure 10: Kleurling (left) and White (right) Afrikaans-speaking listeners’ fixations over time on the target CVC image in the Oral Only and Mixed blocks for tokens produced by the White Afrikaans talker. The structure of the panels in the top, middle, and bottom rows is as in Figure 9.

Although both White and Kleurling Afrikaans-speaking listeners showed the expected slow-down in fixations on the CVC images in the Mixed relative to the Oral Only block for the Kleurling

Afrikaans talker, the corresponding speed-up for the talker of White Afrikaans was not found. We therefore find partial support for the hypothesis that listeners rapidly adjust their perceptual strategies based on speaker-specific acoustic timing patterns.

3.2.4. Do listeners anticipate differences between the Kleurling and White Afrikaans talkers?

We have found partial support in Section 3.2.3 for the hypothesis that listeners rapidly adjust their perceptual strategies, at least for the Kleurling Afrikaans talker, once they receive information about the specific timing patterns of this talker's coarticulatory nasalization in CVN(C) tokens. We turn now to whether listeners anticipate these timing patterns. The timing of coarticulatory nasalization in the perception stimuli reflect typical patterns for Kleurling and White Afrikaans (cf. Section 2.2). Additionally, the two talkers who produced these stimuli are easily and unambiguously identified as speakers of these varieties based on the instruction sentences that introduce each stimulus (cf. Section 3.1.2). If the participants have pre-existing knowledge about the typical timing patterns of coarticulatory nasalization for these two varieties of Afrikaans, they may rely on this knowledge even during the initial Oral Only block of the experiment, before receiving information about coarticulatory nasalization for these specific talkers. In this case, listeners would fixate on the CVC target image in the initial Oral Only block earlier for the White Afrikaans talker because for the White, but not the Kleurling, Afrikaans talker listeners would, by hypothesis, be able to identify the auditory target as a CVC word early during the vowel based on the absence of acoustic evidence for nasalization.

As before, we rely on a GAMM to assess whether this effect is observed in our data. **Figure 11** shows the proportion fixations over time in the Oral Only block for the CVC auditory targets as produced by the White and Kleurling Afrikaans talkers. The left and right panels show the patterns for the Kleurling and White Afrikaans listeners, respectively. Panels in the top row show the average observed fixations, while those in the middle show model-predicted fixations. The bottom row shows model-predicted differences where a positive difference indicates more target fixations in response to the White Afrikaans stimuli. Differences would be expected fairly early during the trial, given that disambiguation between CVC and CVN(C) tokens happen early in the vowel in White Afrikaans. Inspection of the figure shows, however, that the observed differences happen comparatively late (between 400 and 600 ms after vowel onset) and, for both Kleurling and White Afrikaans listeners, the difference is in the opposite-to-expected direction (i.e., more looks to the CVC tokens for the Kleurling than White Afrikaans talker). The opposite-to-expected pattern is difficult to explain. However, given the lateness of this effect, as well as its direction, the current experiment does not provide support for the hypothesis that listeners adjust their perceptual strategies based on presumed pre-existing knowledge about the variety of Afrikaans being spoken.

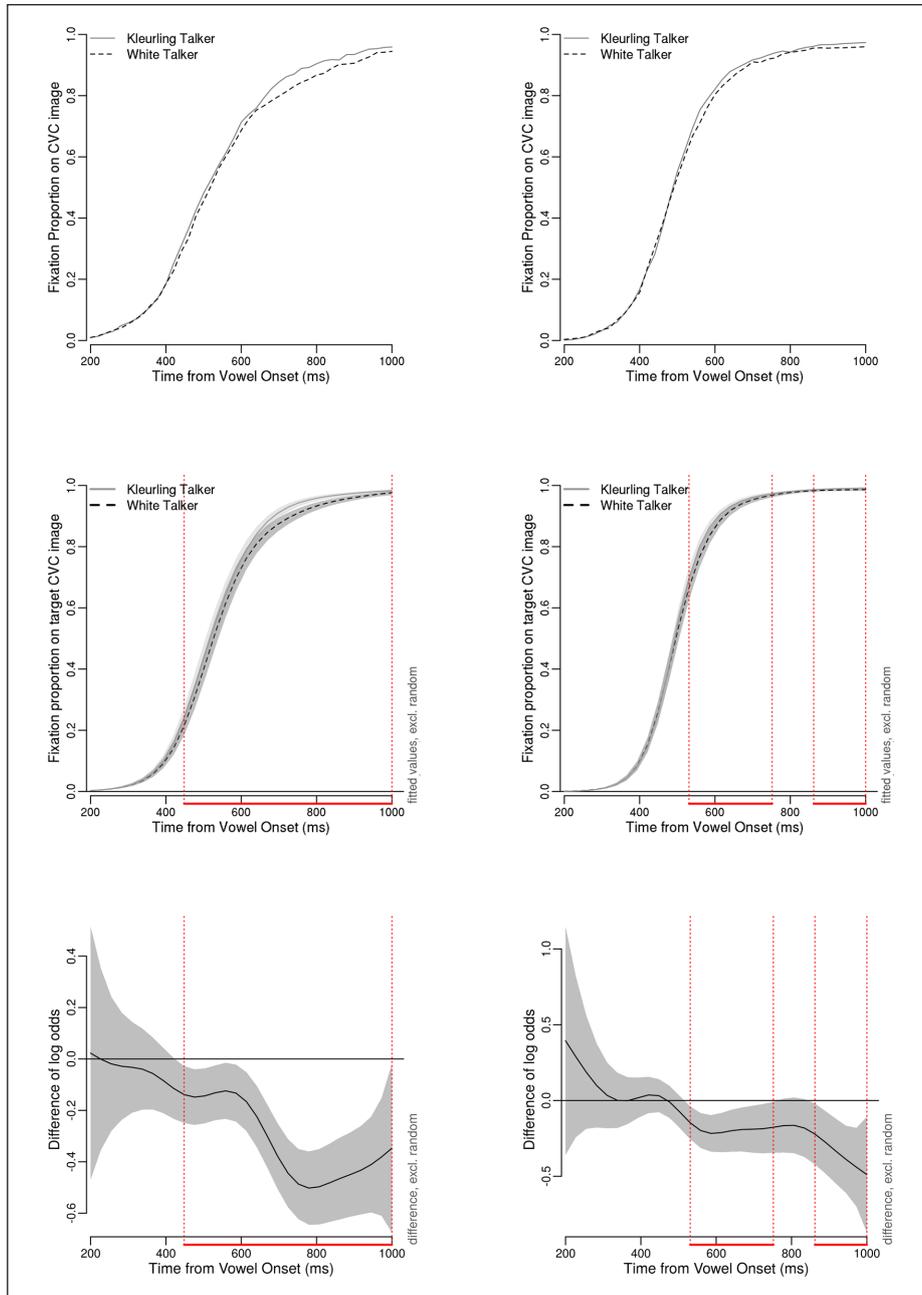


Figure 11: Fixations over time on the target CVC image in the Oral Only block in response to stimuli from the Kleurling (solid line) and White (dashed line) Afrikaans talkers. Fixation patterns for Kleurling Afrikaans listeners are in the left panels, and those for White Afrikaans listeners in the right panels. Top: average observed fixations. Middle: model-predicted fixations. Bottom: model-predicted differences, calculated such that a positive difference corresponds to more target fixations for the White than Kleurling Afrikaans stimuli. Middle and bottom panels include 95% confidence bands; temporal regions marked in red are regions of significant difference.

4. Discussion

4.1. Summary

This study investigated patterns of produced nasal coarticulation and the perceptual reliance on this information by members of an Afrikaans speech community in which variation in nasalization is socially structured. Consistent with earlier impressionistic descriptions, we confirmed more extensive coarticulatory nasalization in White than Kleurling Afrikaans by showing that, in the production of CVN(C) words, nasal airflow both starts earlier and reaches higher overall volumes for speakers of White than for speakers of Kleurling Afrikaans (**Figure 2**). In addition, we documented variation in the amount of produced coarticulatory nasalization within each of the two varieties of Afrikaans through submitting the nasal airflow patterns to an fPCA. In this analysis, the first principal component (PC1) accounted for over 90% of observed variation, with higher PC1 values corresponding to earlier onset and higher overall volume of nasal airflow (**Figure 3**). Although variation was observed within both speaker groups, PC1 values for speakers of White Afrikaans were higher overall than those for speakers of Kleurling Afrikaans (**Figure 4**).

On the perception side, we first demonstrated that both White and Kleurling Afrikaans listeners rely on nasal coarticulation in differentiating CVN(C) and CVC words. Specifically, given that nasalization started early during the vowel for the White Afrikaans stimuli and late for the Kleurling Afrikaans stimuli, systematically earlier fixations on the CVN(C) target image for the White Afrikaans stimuli would be evidence that listeners rely perceptually on coarticulatory nasalization. We found this pattern both for White and Kleurling Afrikaans listeners (**Figure 6**), consistent with our hypothesis that listeners' attention to coarticulation will be sensitive to the time-varying patterns of that information.

Replicating the finding of Beddor et al. (2018) for American English, and consistent with theoretical frameworks that assume a link between individual speakers' perception and production repertoires, we found that speakers who produce more extensive coarticulatory nasalization also rely more on this information as listeners. This was confirmed by showing that participants who, as speakers, produce more extensive nasalization (higher PC1 values) fixate earlier on CVN(C) target images than participants who produce less nasalization. This pattern was observed for both the Kleurling and White Afrikaans stimuli, and for both Kleurling and White Afrikaans listeners (**Figure 7** and **Figure 8**).

We also tested the hypothesis that listeners would adjust their perceptual strategies based on coarticulatory timing differences in the speech of the Kleurling and White Afrikaans talkers by investigating listeners' responses to CVC auditory stimuli. We expected that listeners' exposure (in the Mixed perception block) to early onset of vowel nasalization for the White Afrikaans talker and late onset for the Kleurling Afrikaans talker would lead to faster and slower fixations, respectively, on the CVC target images (relative to fixations on the CVC target images in the Oral Only perception block). This hypothesis, tested by comparing the Oral Only and Mixed perception

blocks, was partially supported. As expected, both White and Kleurling Afrikaans listeners were slower to fixate on oral CVC target images in the (later occurring) Mixed perception block for the Kleurling Afrikaans talker (**Figure 9**), that is, for the talker for whom disambiguation between CVC and CVN(C) auditory stimuli only happens towards the end of the vowel. However, contrary to expectations, for the White Afrikaans talker, for whom CVC-CVN(C) disambiguation happens early in the vowel, we did not find evidence of a speed-up in fixations for either the White or Kleurling Afrikaans listeners (**Figure 10**).

Lastly, we investigated whether listeners have knowledge of the difference in the typical timing patterns of coarticulatory nasalization in White and Kleurling Afrikaans; pre-existing knowledge that they could potentially bring to the perceptual task. In this case, that CVC and CVN(C) can, in general, be disambiguated earlier for White than for Kleurling Afrikaans voices, might lead to faster identification of a word as CVC (rather than CVN(C)) for the White than the Kleurling Afrikaans talker in the experiment—even prior to hearing any CVN(C) produced by that talker. However, as tested within the context of the Oral Only perception block (**Figure 11**), we did not find evidence for this pattern for either the White or Kleurling Afrikaans listeners.

4.2. Is the perception-production link socially mediated?

Although this study is situated within a theoretical framework that postulates a close link between perception and production, we asked whether the production-perception link might nonetheless be relatively weak within a speech community in which coarticulation is socially structured. In asking this question, we have in mind that, similar to perception leading production or production leading perception for some community members in conditions of an ongoing sound change (e.g., Coetzee et al., 2018; Kuang & Cui, 2018; Pinget et al., 2020), perhaps socially structured coarticulatory variation might lead speaker-listeners to attend perceptually to the information encoded in that variation more than might be expected on the basis of their own production patterns.

As just summarized, though, for participants in this study, those individuals who produced more extensive coarticulatory nasalization also relied more on this information perceptually. Clearly, then, any possible social mediation of the production-perception link was not sufficiently strong to override the basic finding in this study that the extent of produced coarticulatory nasalization predicts perceptual reliance on this information, for both White and Kleurling Afrikaans-speaking individuals. Thus, our findings suggest that listeners are applying a perceptual strategy determined at least partially by their own production patterns when listening to either White or Kleurling Afrikaans talkers.

On the other hand, if perception were very closely tied to production, we would expect that Kleurling Afrikaans listeners, who overall produce limited coarticulatory nasalization, would exhibit relatively limited perceptual reliance on nasalization and would hence exhibit

comparably weak influences of speaker ethnicity on their perceptual judgments. Instead, as shown in **Figure 6** (left panels), these listeners fixated more on CVN(C) images when listening to the White Afrikaans compared to the Kleurling Afrikaans talker nearly across the entire duration of the relevant eye-tracking trials. This effect is not simply driven by those Kleurling Afrikaans listeners who produce heavier vowel nasalization since even the Kleurling Afrikaans participants with lower PC1 values fixated reliably earlier on the CVN(C) image when listening to the White Afrikaans talker than to the Kleurling Afrikaans talker. (For example, the dashed curves in the middle left panels of **Figures 7** and **8** show that low-PC1 participants were estimated to fixate on the target 50% of the time at 525 ms after vowel onset for the White talker's stimuli but not until about 575 ms after onset for the Kleurling talker's stimuli.) What we cannot determine from these data, though, is whether the perceptual attention to nasalization by these Kleurling Afrikaans listeners is socially mediated. To determine this, we would need to have stimuli similar to those in this study but presented to participants with at most minimal exposure to speakers of the other variety. If the perception-production link is weakened by the social structure of the speech community in which our study was conducted, we should find an even stronger relation between production and perception for participants with no or limited exposure to the other variety.

4.3. On the nature of talker-specific perceptual adjustment

Both Kleurling and White Afrikaans listeners attend to the talker-specific patterns of coarticulatory vowel nasalization. This outcome might emerge from perceptual learning of a talker's coarticulatory timing, but it could also follow more generally from listeners' close attention to the coarticulatory information as it becomes available in the unfolding acoustic signal—information that is available earlier for the CVN(C) words produced by the White than by the Kleurling Afrikaans talker. However, if listeners are adjusting to the talker's timing patterns for nasality, this perceptual adjustment should also emerge in their responses to the CVC words produced by these talkers. As reviewed in Section 1.2, listeners are adept at rapidly adjusting their perceptual strategies based on the specific acoustic properties of an interlocutor's speech. In this study, similar evidence of perceptual adjustment was expected to emerge in listeners' slower fixation on target images across the course of the experiment for the Kleurling Afrikaans talker's CVC words versus faster fixations for the White Afrikaans talker's CVC words.

That only the first pattern was found (**Figures 9** and **10**) is unexpected. Dahan et al. (2008) and Trude and Brown-Schmidt (2012), for instance, both show that listeners are faster to respond once they get evidence for early disambiguation of stimuli (in their case, faster to respond to *back* after learning that the speaker produced *bag* with a raised diphthong). This is exactly the pattern that we did not find—faster fixations for the White Afrikaans talker. Instead, we found evidence for a slow-down once listeners receive information for late disambiguation for the Kleurling

Afrikaans talker. The reason for this difference is difficult to explain. We note, however, that the phenomenon that is the focus of the studies by Dahan et al. and Trude and Brown-Schmidt is above the level of consciousness—that is, American English listeners will most likely consciously notice the difference between a non-raised and raised production of a word like *bag* ([bæg] versus [beɪg]) because both vowels have phonemic status in English. The difference between early and late onset of coarticulatory nasalization in Afrikaans, however, is below the level of consciousness—coarticulatory nasalization is not phonemic in Afrikaans. These differences in the status of the phenomena may hence be relevant in the different patterns of perceptual adjustment seen in the Dahan et al. and Trude and Brown-Schmidt studies versus our study.

4.4. Differential perceptual strategies based on the identity of the talker

Contrary to results reported by Niedzielski (1999), Hay et al. (2006a, 2006b), Staum Casasanto (2008, 2009a, 2009b), Schertz et al. (2019), and others, which showed that listeners adjust their perceptual strategies based on their prior knowledge of a targeted speech variety, we did not find evidence that listeners adjust their perceptual strategies based on the assumed identity of the talker. Specifically, we did not find for either the White or the Kleurling Afrikaans listeners that they were faster to fixate on the CVC target images in the Oral Only block for the White than for the Kleurling Afrikaans talker (**Figure 11**). That is, we did not find evidence that these listeners anticipated, based on pre-existing knowledge about nasalization patterns in different dialects of Afrikaans, talker-specific coarticulatory patterns.

The absence of this effect may be at least in part methodological: Previous studies showing an influence of anticipated talker variety on listeners' judgments have used visual priming (e.g., orthographic label or purported picture of the talker). In spite of the fact that we found high accuracy in identifying the variety of Afrikaans spoken by the two talkers who provided stimuli for the perception study (see Section 3.1.2), it may be that the auditory instructions for each trial produced by the Kleurling and White Afrikaans talkers served as less explicit information about talker identity than the explicit visual or orthographic cues used in other studies. (In this regard, we note that Munson, Ryherd, & Kemper, 2017, found that explicit priming of talker sex with a picture of a woman or a man had stronger influences on linguistic judgments than implicit priming based on female- versus male-associated sentence content.)

Another methodological difference between our study and some studies that found evidence of an influence of anticipated talker identity is that all auditory stimuli in our study were congruent with speaker variety—that is, all White Afrikaans CVN(C) stimuli had early onset nasalization and all corresponding Kleurling Afrikaans stimuli had late onset nasalization. In comparison, participants in some other studies were also presented with trials in which the acoustic properties of the stimuli were incongruent with the assumed identity of the speaker. In their study of the perception of New Zealand and Australian English vowels, Hay and Drager (2010), for instance,

presented all stimuli (both those typical of New Zealand and Australian English) in both the New Zealand and Australian conditions of their study. The mismatch between the acoustic stimuli and the patterns expected based on the assumed identity of the talker could cause participants to attend more closely to these expected patterns.

Alternatively, or in addition, for the White Afrikaans listeners, absence of evidence of a difference in anticipatory response to White as opposed to Kleurling Afrikaans might be ascribed to the fact that, due to the structure of South African society, they have less exposure to Kleurling Afrikaans and so may not have sufficient knowledge of the differences between the two varieties of the language to adjust their perceptual strategies relative to the identity of the talker. Moreover, along the lines of the claim by Sumner, Kim, King, and McGowan (2014) that socially stigmatized varieties of a language receive less robust exemplar encoding, even if White Afrikaans listeners have sufficient exposure to Kleurling Afrikaans they may not use this information to inform differential perceptual strategies in response to a speaker of Kleurling versus White Afrikaans. However, these explanations are not available for the lack of evidence of differential perceptual strategies for the Kleurling Afrikaans listeners. These listeners would have ample exposure not only to Kleurling Afrikaans (in the home and family context) but also to White Afrikaans (through the media and as students at a majority White Afrikaans university), which is also the prestige variety of the language, especially in the academic context where data collection for this study took place.

Yet another possible explanation for the absence of differential perceptual anticipation strategies may again (see Section 4.3 for perceptual adjustment strategies) be that the phenomenon of interest here (coarticulatory nasalization) is below the level of consciousness and is non-phonemic in Afrikaans. This differentiates this phenomenon from at least some of the phenomena for which such differential perceptual strategies have been documented.

5. Conclusion

Many phonetic theories assume a close relation between speech production and perception including, for some approaches, between the production and perception repertoires of individual language users. At the same time, successful communication depends on listeners being able to accurately perceive speech produced by speakers whose production patterns may be quite different from their own, implying a need for flexibility in the perception-production link. Understanding the factors that mediate this link at the level of the individual and the speech community is therefore central to phonetic theory. In this study, we investigated how the production-perception link may be mediated by socially structured variation in the extent of produced coarticulatory nasalization in an Afrikaans speech community. For this community, we found evidence for a production-perception link at the level of the individual, such that individuals who produce

more coarticulatory nasalization also rely more on this information in perception—and they do so regardless of the talker’s (predictably structured) pattern of nasalization. The persistence of the production-perception link, even in a context of socially structured variation, provides evidence for the robustness of this link. At the same time, although the relative perceptual usefulness of coarticulatory information is informed by listeners’ own productions, our results also show that even language users who themselves produce little to no anticipatory nasalization are nonetheless adept at using that information in perception. The evidence provided in this study further shows, though, that listeners’ perceptual adjustments for speaker-specific, real-time information occur only under certain circumstances. No clear evidence was found for the social mediation of the link between production and perception based on pre-existing knowledge of different coarticulatory patterns in different socio-ethnic varieties of Afrikaans. The continuing challenge for phonetic theory is to determine how individual language users balance, from moment to moment, their reliance on the acoustic patterns in the speech of their interlocutors, and their reliance on their own production patterns.

Supplementary Materials

The supplementary materials (data and statistical code) are too large to be uploaded to the journal website. We are instead making it available at this link: http://bit.ly/Afr_Nas_Supplementary

Start by reading Readme.txt.

Acknowledgements

The research reported here was supported in part by NSF Grant BCS 1348150 to Patrice Beddor (PI) and Andries Coetzee (co-PI). We thank Ian Calloway for his input on this project, and also Claire Laing, Skye Huerta, Karen Tze Hui Tan, and Deon du Plessis, who assisted in various ways in the collection and coding of the data. We also extend our gratitude to two anonymous reviewers, and to the members of the *Laboratory Phonology* editorial team, including Abby Walker and Lisa Davidson, for their input on this paper. Earlier versions of this work were presented at ICPhS 2019 and to the University of Michigan's Phonetics/Phonology Discussion Group, and we thank these audiences for valuable discussions.

Competing Interests

The authors have no competing interests to declare.

References

- Baayen, R. H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, *94*, 206–234. DOI: <https://doi.org/10.1016/j.jml.2016.11.006>
- Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*, 785–821. DOI: <https://doi.org/10.1353/lan.0.0165>
- Beddor, P. S., Coetzee, A. W., Boland, J. E., McGowan, K. B., & Styler, W. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language*, *94*, 931–968. DOI: <https://doi.org/10.1353/lan.2018.0051>
- Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, *30*, 591–627. DOI: <https://doi.org/10.1006/jpho.2002.0177>
- Beddor, P. S., & Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *Journal of the Acoustical Society of America*, *106*, 2868–2887. DOI: <https://doi.org/10.1121/1.428111>
- Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *Journal of the Acoustical Society of America*, *133*, 2350–2366. DOI: <https://doi.org/10.1121/1.4794366>

- Boersma, P., & Weenink, D. (2013). *Praat: Doing Phonetics by Computer (Version 5.3.56)*. [Computer Program.] Retrieved May 23, 2012, from <http://www.praat.org>.
- Bongiovanni, S. (2018). *Production of Anticipatory Vowel Nasalization and Word-final Nasal Consonants in Two Dialects of Spanish* (Doctoral Dissertation). Indiana University.
- Britt, E., & Weldon, T. L. (2015). African American English in the middle class. In J. Bloomquist, L. J. Green & S. L. Lanehart (Eds.), *The Oxford Handbook of African American Language* (pp. 800–816.) Oxford: Oxford University Press.
- Coetzee, A. E. (1981). Variasies by nasalering in Afrikaans. [Variations in nasalization in Afrikaans.] In A. J. L. Sinclair (Ed.), *LVSA-Kongresreferate*. (pp. 128–146).
- Coetzee, A. E. (1989). *Uitspraakvariasie in die Afrikaans van die Johannesburgse Bruin gemeenskappe: 'n Vergelykende Studie*. [Pronunciation Variation in the Afrikaans of the Johannesburg Brown Communities: A Comparative Study] (Doctoral dissertation). Randse Afrikaanse Universiteit.
- Coetzee, A. E., & van Reenen, P. Th. (1995). Die Afrikaanse nasalering en nienasalering se verband met 17de-eeuse Nederlands. [The relation of Afrikaans nasalization and non-nasalization with 17th century Dutch.] *South African Journal of Linguistics*, 13, 62–73. DOI: <https://doi.org/10.1080/10118063.1995.9723978>
- Coetzee, A. W. (2018). Individual and community-level variation in phonetics and phonology. In D. Bradley & R. Mesthrie (Eds.), *The Dynamics of Language* (pp. 176–192.) Cape Town: University of Cape Town Press.
- Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics*, 66, 185–216. DOI: <https://doi.org/10.1016/j.wocn.2017.09.009>
- Coetzee, I. A. (1985). Nasalering in die Afrikaans van die Bruin gemeenskap in Eersterust, Pretoria. In N. J. Grieshaber & J. L. Venter (Eds.), *LVSA-Kongresreferate* (pp. 64–82.)
- Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, 108, 710–718. DOI: <https://doi.org/10.1016/j.cognition.2008.06.003>
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16, 507–534. DOI: <https://doi.org/10.1080/01690960143000074>
- Delvaux, V., Demolin, D., Harmegnies, B., & Soquet, A. (2008). The aerodynamics of nasalization in French. *Journal of Phonetics*, 36, 578–606. DOI: <https://doi.org/10.1016/j.wocn.2008.02.002>
- Delvaux, V., Huet, K., Piccaluga, M., & Harmegnies, B. (2012). Inter-gestural timing in French nasal vowels: A comparative study of (Liège, Tournai) Northern French vs. (Marseille, Toulouse) Southern French. *INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association*, 3, 2681–2684. DOI: <https://doi.org/10.21437/Interspeech.2012-666>
- Den Besten, H. (1989). From Khoekhoe foreigner talk via Hottentot Dutch to Afrikaans: The creation of a novel grammar. In M. Pütz & R. Driven (Eds.), *Wheels within Wheels: Papers of the Duisberg Symposium on Pidgin and Creole Languages*. (pp. 207–249.) Frankfurt am Mein: Peter Lang.

- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3–28. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30607-2](https://doi.org/10.1016/S0095-4470(19)30607-2)
- Grosvald, M. (2009). Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics*, 37, 173–188. DOI: <https://doi.org/10.1016/j.wocn.2009.01.002>
- Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105, 611–633. DOI: <https://doi.org/10.1037/0033-295X.105.4.611-633>
- Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America*, 123, 2825–2835. DOI: <https://doi.org/10.1121/1.2897042>
- Harrington, J., Kleber, F., Reubold, U., Schiel, F., & Stevens, M. (2019). The phonetic basis of the origin and spread of sound change. In W. F. Katz & P. F. Assmann (Eds.), *The Routledge Handbook of Phonetics*. (pp. 401–426.) London: Routledge. DOI: <https://doi.org/10.4324/9780429056253-15>
- Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48, 865–892. DOI: <https://doi.org/10.1515/ling.2010.027>
- Hay, J., Nolan, A., & Drager, K. (2006a). From *fish* to *feesh*: Exemplar priming in speech perception. *The Linguistic Review*, 23, 351–379. DOI: <https://doi.org/10.1515/TLR.2006.014>
- Hay, J., Warren, P., & Drager, K. (2006b). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34, 458–484. DOI: <https://doi.org/10.1016/j.wocn.2005.10.001>
- Hayes, B., & Stivers, T. (2000). Postnasal Voicing. Ms., UCLA. [Available online at <http://www.linguistics.ucla.edu/people/hayes/#phonetics>.].
- Hutton, S. B. (2008). Cognitive control of saccadic eye movements. *Brain and Cognition*, 68, 327–340. DOI: <https://doi.org/10.1016/j.bandc.2008.08.021>
- Hyndman, R. J., & Ullah, Md. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, 51, 4942–4956. DOI: <https://doi.org/10.1016/j.csda.2006.07.028>
- Kataoka, R. (2011). *Phonetic and Cognitive Bases of Sound Change* (Doctoral dissertation). University of California, Berkeley. Linguistics.
- Kleber, F., Harrington, J., & Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech*, 55, 383–405. DOI: <https://doi.org/10.1177/0023830911422194>
- Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, 107, 54–81. DOI: <https://doi.org/10.1016/j.cognition.2007.07.013>
- Kuang, J. & Cui, A. (2018). Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics*, 71, 194–214. DOI: <https://doi.org/10.1016/j.wocn.2018.09.002>

- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36. DOI: [https://doi.org/10.1016/0010-0277\(85\)90021-6](https://doi.org/10.1016/0010-0277(85)90021-6)
- Lieberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4, 187–196. DOI: [https://doi.org/10.1016/S1364-6613\(00\)01471-6](https://doi.org/10.1016/S1364-6613(00)01471-6)
- Lindblom, B., Guion, S., Hura, S. L., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica*, 7, 5–37.
- Munson, B., Ryherd, K., & Kemper, S. (2017). Implicit and explicit gender priming in English lingual sibilant fricative perception. *Linguistics*, 55, 1073–1107. DOI: <https://doi.org/10.1515/ling-2017-0021>
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18, 62–85. DOI: <https://doi.org/10.1177/0261927X99018001005>
- Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R.A. Hendrick & M.F. Miller (Eds.), *Chicago Linguistic Society: Papers from the Parasession on Language and Behavior* (pp. 178–203.) Chicago: Chicago Linguistic Society.
- Pardo, J. S. (2012). Reflections on phonetic convergence: Speech perception does not mirror speech production. *Language and Linguistics Compass*, 6, 753–767. DOI: <https://doi.org/10.1002/lnc3.367>
- Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency Effects and the Emergence of Lexical Structure*. (pp. 137–157.) Amsterdam: Benjamins. DOI: <https://doi.org/10.1075/tsl.45.08pie>
- 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, 660–685. DOI: <https://doi.org/10.1177/0023830919880206>
- R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Ramsay, J. O., Wickham, H., Graves, S., & Hooker, G. (2020). fda: Functional Data Analysis, R package version 2.4.8.1. <https://cran.r-project.org/web/packages/fda/index.html>.
- Roberge, P. T. (1994). The formation of Afrikaans. *Stellenbosch Papers in Linguistics*, 27, 1–121. DOI: <https://doi.org/10.5774/27-0-69>
- Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, 71, 1207–1218. DOI: <https://doi.org/10.3758/APP.71.6.1207>
- Scanlon, M., & Wassink, A. B. (2010). African American English in urban Seattle: Accommodation and intraspeaker variation in the Pacific Northwest. *American Speech*, 85(2), 205–224. DOI: <https://doi.org/10.1215/00031283-2010-011>
- Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, 52, 183–204. DOI: <https://doi.org/10.1016/j.wocn.2015.07.003>

- Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *WIREs: Wiley Interdisciplinary Reviews: Cognitive Sciences*, 11, e1521. DOI: <https://doi.org/10.1002/wcs.1521>
- Schertz, J., Kang, Y., & Han, S. (2019). Sources of variability in phonetic perception: The joint influence of listener and talker characteristics on perception of the Korean stop contrast. *Laboratory Phonology*, 10, 13. DOI: <https://doi.org/10.5334/labphon.67>
- Shang, H. L., & Hyndman, R. J. (2019). Rainbow: Bagplots, Boxplots and Rainbow Plots for Functional Data, R package version 3.6. <https://CRAN.R-project.org/package=rainbow>.
- Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *Journal of the Acoustical Society of America*, 132, EL95–EL101. DOI: <https://doi.org/10.1121/1.4736711>
- Staum Casasanto, L. (2008). Does social information influence sentence processing? *Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society*, 30. Retrieved from <https://escholarship.org/uc/item/8dc2t2gf>.
- Staum Casasanto, L. (2009a). *Experimental Investigations of Sociolinguistic Knowledge* (Doctoral dissertation). Stanford University. Linguistics.
- Staum Casasanto, L. (2009b). What do listeners know about sociolinguistic variation? *University of Pennsylvania Working Papers in Linguistics*, 15, 39–49.
- Stell, G. (2011). *Ethnicity and Language Variation: Grammar and Code-switching in the Afrikaans Speech Community*. Frankfurt am Main: Peter Lang.
- Stroop, J. (1994). Afgedwongen nasalering. [Forced nasalization.] *Tijdschrift voor Nederlandse Taal- en Letterkunde*, 110, 55–67.
- Sumner, M., & Kataokoa, R. (2013). Effects of phonetically-cued talker variation on semantic encoding. *Journal of the Acoustical Society of America*, 134, EL485–491. DOI: <https://doi.org/10.1121/1.4826151>
- Sumner, M., Kim, S. K., King, Ed., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psychology*, 4, 1–13. DOI: <https://doi.org/10.3389/fpsyg.2013.01015>
- Tamminga, M., & Zellou, G. (2015). Cross-dialectal differences in nasal coarticulation in American English. In The Scottish Consortium for ICPhS 2015 (Ed.) *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow.
- Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, 27, 979–1001. DOI: <https://doi.org/10.1080/01690965.2011.597153>
- Trude, A. M., Duff, M. C., & Brown-Schmidt, S. (2014). Talker-specific learning in amnesia: Insight into mechanisms of adaptive speech perception. *Cortex*, 54, 117–123. DOI: <https://doi.org/10.1016/j.cortex.2014.01.015>
- van Reenen, P. Th., & Coetzee, A. E. (1996). Afrikaans, a daughter of Dutch. In H. F. Nielsen & L. Schøsler (Eds.), *The Origins and Development of Emigrant Languages: Proceedings from the Second*

- Rasmus Rask Colloquium, *Odense University, November 1994* (pp. 71–102.) Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/nss.17.07ree>
- van Rensburg, M. C. J. (1989). Soorte Afrikaans. [Types of Afrikaans.] In T. J. R. Botha, F. A. Ponelis, J. G. H. Combrink & F. F. Odendal (Eds.), *Inleiding tot die Afrikaanse Taalkunde*. [Introduction to Afrikaans Linguistics.] (pp. 436–467.) Pretoria: Academica.
- van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2020). itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. R package version 2.4. <https://cran.r-project.org/web/packages/itsadug/>.
- Weener, P. D. (1969). Social dialect differences and the recall of verbal messages. *Journal of Educational Psychology*, 60(3), 194–199. DOI: <https://doi.org/10.1037/h0027559>
- Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics*, 35, 49–64. DOI: <https://doi.org/10.3758/BF03205924>
- Wissing, D. (2018). Nasalization. *Taalportaal*. Retrieved from <http://taalportaal.org/taalportaal/topic/pid/topic-15052933831939084>. [Accessed 20 June 2018].
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 3–36. DOI: <https://doi.org/10.1111/j.1467-9868.2010.00749.x>
- Wood, S. N. (2019). mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. R package version 1.8-31. <https://cran.r-project.org/web/packages/mgcv/>.
- Yu, A. C. L. (2013). Individual differences in socio-cognitive processing and the actuation of sound change. In A. C. L. Yu (Ed.) *Origins of Sound Change: Approaches to Phonologization* (pp. 201–227.) Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199573745.003.0010>
- Yu, A. C. L. (2019). On the nature of the perception-production link: Individual variability in English sibilant-vowel coarticulation. *Laboratory Phonology*, 10, 2. DOI: <https://doi.org/10.5334/labphon.97>
- Yu, A. C. L., & Zellou, G. (2019). Individual differences in language processing: Phonology. *Annual Review of Linguistics*, 5, 131–150. DOI: <https://doi.org/10.1146/annurev-linguistics-011516-033815>
- Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, 61, 13–29. DOI: <https://doi.org/10.1016/j.wocn.2016.12.002>
- Zellou, G., & Tamminga, M. (2014). Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics*, 47, 18–35. DOI: <https://doi.org/10.1016/j.wocn.2014.09.002>

