Stop! Hey, what's that sound? the representation and realization of Danish stops

Puggaard-Rode, R.

Part II

Corpus studies

CHAPTER 4

The rarity of intervocalic voicing

4.1 Introduction

Intervocalic voicing of underlyingly voiceless stops is phonetically well-understood but is a phonological conundrum. Voicing is usually difficult to maintain during closure, leading to the common assumption that the feature [voice] is phonologically marked in stops. Intervocalically, however, the vocal folds are initially adducted and tensed, and subglottal pressure is high, providing ideal conditions for closure voicing (Westbury and Keating 1986). Hence, voicing is often found in this position, even in languages where voicing does not play a role in distinguishing stops (Kaplan 2010). In other words, the markedness of voicing depends on position; voicing requires an effort in initial and final position, while voicelessness requires an effort intervocalically. This distribution of markedness is difficult to account for phonologi-

The research reported in this chapter is collaborative work with Camilla Søballe Horslund and Henrik Jørgensen. This chapter is based (largely verbatim) on a published paper (Puggaard-Rode et al. 2022a), and an earlier version of the study was presented at a conference (Puggaard et al. 2020). Audio data are available online in password-protected form (Grønnum 2016); replication data and code are freely available (Puggaard-Rode et al. 2022b).

cally, where voicing is usually associated with a [voice] feature, i.e. a more complex and marked structure.

As discussed in Section 2.3.2, there is no closure voicing in absolute initial position in Modern Standard Danish, and only negligible voicing in final position. Voicing in intervocalic position is less well-understood. There are essentially two different positions in the literature on the status of intervocalic voicing in Modern Standard Danish. One position holds that stops are systematically voiced in medial position (Abrahams 1949; Fischer-Jørgensen 1954, 1980; Spore 1965; Keating et al. 1983; Kingston and Diehl 1994); the other holds that Danish stops are systematically voiceless in all positions (Jessen 1999, 2001; Beckman et al. 2013). The lack of concrete knowledge about intervocalic voicing is a major gap in the phonetics and phonology of Danish stops. On the phonetics side, the glottal activity during stop production is well-described (Frøkjær-Jensen et al. 1971; Fischer-Jørgensen and Hirose 1974; Hutters 1985; see Section 2.3.3), as are accompanying F_0-perturbations in following sonorant sounds (Fischer-Jørgensen 1969; Jeel 1975; Petersen 1983; see Section 2.3.6). On the phonology side, intervocalic voicing has played a pivotal role in discussions of the underlying representation of Danish stops, and both aforementioned positions have been taken as evidence in favor of underlying representations (e.g. Kingston and Diehl 1994; Beckman et al. 2013; see Section 2.4.3.1). The status of intervocalic voicing could be a skeleton key to both 1) contextualizing existing studies of the phonetics of Danish stops, and 2) evaluating proposals regarding the representation of the laryngeal contrast in Danish.

The problem is intriguing beyond just the Danish context. Much of our knowledge about intervocalic voicing comes from aerodynamic models (such as Westbury and Keating 1986); excluding English, there are few quantitative studies of intervocalic voicing in 'aspiration languages', and particularly few where the data has high ecological validity. As laid out in Section 2.3, the breadth of our existing knowledge about Danish stop production is quite impressive; concrete knowledge about intervocalic voicing may help tie together our existing knowledge of voice onset time, glottal activity, and F_0-perturbations. The relationship between the resulting articulatory actions and acoustic cues are central to our understanding of how laryngeal contrast should be phonologically represented.

In this chapter, I present the first empirical study of intervocalic stop voicing in Danish. The study is based on the existing DanPASS corpus (Danish Phonetically Annotated Spontaneous Speech; Grønnum 2009, 2016; see Section 4.5.1 below). In Section 2.3, I summarize a number of observations about the production of Danish stops which make intervocalic voicing patterns difficult to predict. The unaspirated stops /b d g/ are produced with longer closure duration and greater muscular tension than the aspirated stops /p t k/ (Fischer-Jørgensen 1954; Fischer-Jørgensen and Hirose 1974; see Section 2.3.5). From this perspective, /p t k/ should actually be more conducive to voicing. It is sometimes claimed that these differences are too small to be of significance (e.g. Grønnum 2005), and that both sets of stops are phonetically lenis, which suggests that both sets are equally likely to be voiced intervocalically. On the other hand, glottographic and electromyographic (EMG) investigations have shown that both stop types are characterized by a glottal opening gesture during the closure when produced between vowels in careful speech, and that this gesture lasts longer and is of greater magnitude in /p t k/ (Frøkjær-Jensen et al. 1971; Fischer-Jørgensen and Hirose 1974; Hutters 1985; see Section 2.3.3). This suggests that /b d g/ should be most conducive to voicing, but also that voicing is actively blocked in both sets.

The results show that intervocalic voicing is very rare in /p t k/. Although much more frequent in /b d g/, intervocalic voicing occurs in less than half of /b d g/ tokens. This rarity of intervocalic voicing is in essence the opposite conundrum of the one I mentioned at the beginning of the chapter. Voicing is natural in this position, so its rarity can only be accounted for with reference to some mechanism that blocks it. The occurrence of intervocalic voicing is generally correlated with other variables that are associated with phonetic lenition; it occurs more frequently in quick speech, in morphological affixes, before neutral vowels, and in unstressed syllables. This suggests that intervocalic voicing is in itself lenition phenomenon. I suggest that this lenition is best modeled as gesture reduction: Danish has phonologized glottal spreading gestures in all stops, which usually block voicing; however, in environments that are generally prone to lenition, this gesture is lost at a relatively high rate. This happens more frequently in /b d g/, where the gesture has less of a critical function.

The chapter is structured as follows: Section 4.2 provides some background on closure voicing in phonetics and phonology with particular focus on intervocalic position, and Section 4.3 briefly recaps some important facts about the phonetics and phonology of Danish stops from Chapter 2. In Section 4.4, I summarize the research questions and motivate all independent variables of the study. In Section 4.5, I provide an overview of the methods: I introduce the corpus used for the study, and explain how the data were processed. In Section 4.6, I provide an exploratory analysis of the data. In Section 4.7, I describe the selection of a logistic mixed-effects regression model, and give the results of that model. In Section 4.8, I discuss the research questions in light of the results, and in Section 4.9, I briefly summarize the chapter.

4.2 Closure voicing and [voice]

I already alluded to the phonetic pressure against closure voicing in Section 3.5. Here, I will cover the relevant literature in more detail, and discuss the role that this pressure plays in discussions of the underlying representation of laryngeal contrast.

Closure voicing in stops is "unnatural" (Ohala 1983a). A sufficient transglottal pressure differential is required to maintain vocal fold vibration. As air continually flows from the lungs while both the oral and nasal cavities are sealed, supraglottal air pressure rises quickly. This effectively means that it is impossible to maintain closure voicing for a long duration of time. If the size of the supraglottal cavity remained constant during a stop closure, it would only be possibility to maintain a sufficient transglottal pressure differential for roughly 5–10 ms (Ohala and Riordan 1979). This is not actually the case, though; the vocal tract automatically enlarges, primarily due to compliance of the soft tissue making up the inner walls of the cavity. This should allow for approximately 60–70 ms of closure voicing for a male speaker (Westbury 1983), varying depending on e.g. the point of occlusion. Voicing is maintained longest for a fronted occlusion (e.g. bilabial) due to the large cavity between the glottis and the supraglottal occlusion, which yields a slower build-up of air pressure and crucially yields a larger total area of soft, expandable cavity walls. Keating (1984b) shows that bilabial stops naturally retain voicing for roughly 30% longer than velar stops. When some languages show yet longer closure voicing, it is due to active vocal tract enlargement during the occlusion, implemented by e.g. lowering the jaw or raising the velum.

Westbury and Keating (1986) investigate the issue of articulatory naturalness in detail, using a model of breath-stream control with the vocal folds appropriately adducted and tensed for voicing. They show that syllable-initial closure voicing is articulatorily unnatural, since subglottal and supraglottal air pressure rises roughly synchronously, unless the vocal folds are initially fully abducted to allow for preparatory build-up of subglottal air pressure. Closure voicing is also unnatural syllable-finally; Westbury and Keating hypothesize that this is due to an inspiratory force that gradually but quickly counteracts the initially high subglottal pressure from the preceding vowel. However, it is natural for intervocalic stop closures to be voiced throughout most of their duration due to the high initial subglottal pressure following the preceding vowel.

Articulatory naturalness sometimes translates directly into typological patterns in phonology, but this is not always the case. In accordance with articulatory naturalness, there is a strong implicational hierarchy regarding voiced stops in phonological inventories. In almost all cases, languages with voiced stops also have voiceless stops (e.g. Ohala 1983a; Maddieson 1984). Furthermore, final obstruent devoicing is a very common typological pattern, partially because syllable-final segments tend to be lengthened, resulting in longer stretches of voicelessness in coda stops, and as such a lesser chance of closure voicing being interpreted as an important phonological cue (e.g. Blevins 2004: 103ff.). However, in spite of their unnatural status, syllable-initial voiced stops are actually quite common. Furthermore, voicing is most natural in medial position, but languages with no laryngeal distinction in stops generally have voiceless stops in all positions (Keating et al. 1983). This illustrates an important point: there is more to phonological patterns than ease of articulation.

Below, I characterize three approaches to the representation of laryngeal contrasts in the phonological literature, as well as the predictions they make with regards to intervocalic closure voicing. Some of these were also introduced in Section 2.4.3.1, but the discussion here will be somewhat broader, and not limited to the representation of contrast in Danish. There is a huge literature on the representation of laryngeal contrast, so some approaches will necessarily be missed here, while others may be grouped together even if they differ in some respect. I will refer to these approaches as 'concrete [voice]' approaches, 'abstract [voice]' approaches, and 'gesture-based' approaches.

The phonological feature [voice] has been conceptualized in different ways. It sometimes refers narrowly to the presence of voicing during closure, which is what I refer to as concrete [voice]. This is how [voice] is conceptualized in the laryngeal feature geometry of Lombardi (1995), the 'laryngeal realism' approach of Iverson and Salmons (e.g. 1995), and the 'laryngeal dimensions' model of Avery and Idsardi (2001). These are approaches that assume a direct relationship between physical laryngeal constellations and phonological laryngeal features. Such approaches usually assume that languages with aspiration-based contrasts employ an active feature like [spread glottis] or [glottal width] to distinguish laryngeal contrasts. It is common to assume that sonorant sounds are unmarked for [voice], since vocal fold vibration is the natural state of affairs in these sounds (Lombardi 1995). This creates a problem in determining the phonological origin of intervocalic closure voicing; surrounding vowels are unmarked for [voice], so voicing cannot spread from those. One possible solution to this is a non-laryngeal [spontaneous voice] feature node, which can spread from sonorants to obstruents, as proposed by Rice and Avery's (1989). Another solution is to simply relegate intervocalic voicing to phonetic implementation, placing it outside the purview of phonology. This would predict that laryngeally unmarked intervocalic stops always follow the phonetically natural pattern.

Jessen and Ringen (2002) and Beckman et al. (2013) argue that the intervocalic behavior of stops is relevant for determining whether [voice] or [spread glottis] are active in a language. Beckman et al.

show that Russian /b d g/ are voiced throughout the closure in intervocalic position with very few exceptions, while German /b d g/ are variably voiced in intervocalic position (roughly 60% of tokens are voiced throughout). They take the consistent voicing in Russian as evidence for an active [voice] feature, and the variable voicing in German as evidence for a gradient phonetic process of passive voicing. Following Chomsky and Halle (1968), they assume that at some stage in the phonological derivation, segments are assigned numerically-valued features; the degree of intervocalic voicing in a [spread glottis] language will depend on the value assigned to [spread glottis] at this late stage in the derivation.[1] The findings of Beckman et al. can only be taken as evidence for underlying features if one assumes a transparent relationship between phonology and phonetic implementation; see e.g. Keating (1984a) for a general critique of this stance.

In abstract [voice] approaches, [voice] in stops does not necessarily imply closure voicing. Chomsky and Halle (1968) and Keating (1984a) both assume that [voice] can be cued with either closure voicing or voicing onset approximately at the time of release, depending on which contrast the language in question employs. Kingston and Diehl (1994) similarly assume a feature [voice] that does not always entail closure voicing in stops. This argument partially relies on the finding that [voice]-induced F_0-perturbations behave similarly regardless of how the feature is phonetically implemented. In their account, the feature [voice] lowers F_0 on the following vowel.[2] Kingston and Diehl recognize that there is a discrepancy between initial and intervocalic position when it comes to the naturalness of closure voicing; they argue that an 'automatic phonetics' will output voiceless initial stops and voiced intervocalic stops, and that a 'controlled phonetics' is necessary to divert from that pattern.

If we assume a relationship between phonology and phonetics, then there should be a correspondence between which patterns are unmarked in phonology and phonetics. Given the aerodynamic account of stop voicing presented here, this entails that an unmarked

[1] See Kirby and Ladd (2018) for a critical discussion of the predictions that follow from this account, in particular as relates to laryngeally induced F_0-perturbations.

[2] Recall from Section 2.4.3.1 that the cause of this pattern is disputed; Hanson (2009) argues that F_0 is raised locally by voiceless stops.

stop should be voiceless initially and voiced intervocalically. It also entails that phonetic reduction is positionally defined: devoicing of [voice] stops is lenition syllable-initially and syllable-finally, whereas voicing of stops without [voice] is lenition intervocalically.[3] This is difficult to account for in a feature-based framework but seems to hold up for intervocalic position, where voicing of stops without underlying [voice] is crosslinguistically common (Kaplan 2010).[4] In an optimality-theoretic analysis of this problem, Smith (2008) proposes constraints that militate against voiced obstruents in onset position and voiceless obstruents in intervocalic position, which compete with faithfulness constraints (see also Hayes 1999). Katz (2016) points out some typological shortcomings of this account: Smith's account predicts languages which neutralize a laryngeal contrast in initial position due to devoicing, but no such language is attested. Likewise, Smith's account predicts languages where a laryngeal contrast is systematically neutralized intervocalically due to voicing; such languages are also surprisingly rare, and Katz proposes reanalyses of the attested languages.

Gesture-based approaches to phonological representation can straightforwardly account for these positional markedness relations. One such approach is Articulatory Phonology (Browman and Goldstein 1986, 1992). In Articulatory Phonology, articulatory gestures are taken as the primary units of phonological representation rather than segments or features. A consequence is that the duration and magnitude of glottal gestures can be represented separately from other gestures that make up traditional segments. The unmarked state of the glottis is adducted and tensed, which will not cause voicing initially but will cause voicing intervocalically, as discussed above.

Some predictions about the patterning of intervocalic closure voicing follow from these different conceptualizations of laryngeal representation. In concrete [voice] approaches, closure voicing is a necessary and sufficient criterion for [voice] and a different feature like [spread glottis] is needed to account for aspiration. Concrete

[3] See Steriade (2009) for a discussion of positional markedness and laryngeal contrasts focusing on final position.

[4] It does not, however, hold up for initial position; there is no common process of initial devoicing, at least not resulting in positional neutralization (Katz 2016).

[voice] approaches predict essentially categorical voicing of all intervocalic stops in 'true voice' languages; [voice] ensures voicing in one set of stops, and there are no available phonological mechanisms to counteract voicing in the other (unmarked) set of stops.[5] In 'aspiration languages', we predict varying degrees of intervocalic voicing of unmarked stops, and very little voicing in [spread glottis] stops (following Beckman et al. 2013). In abstract [voice] approaches, where [voice] can have different phonetic interpretations, it is less straightforward to predict intervocalic behavior. Following Kingston and Diehl (1994), a 'controlled phonetics' is necessary to divert from the natural pattern of intervocalic voicing. A gesture-based approach such as Articulatory Phonology also predicts the natural pattern of intervocalic stop voicing if no underlying glottal gestures are present; however, Articulatory Phonology allows a great deal of flexibility in how glottal gestures are represented, making it a very powerful representational framework.

4.3 Predictions for Danish

In this section, I will recap some relevant phonetic facts and phonological arguments concerning voicing and laryngeal representation in Danish stops, and discuss how they relate to the predictions posed in the preceding section. In the following, when I refer to /b d g/ and /p t k/, I refer to the surface contrast: /b d g/ are stops that would be unaspirated in distinct speech, and /p t k/ are stops that would be aspirated in distinct speech. I will refer to the two series as *laryngeal categories.*

I discussed in Section 2.3.5 how Fischer-Jørgensen (1972b) has argued on the basis of closure duration and EMG studies that /b d g/ are fortis, and /p t k/ are lenis. It has later been argued (by e.g. Grønnum 2005) that the difference in closure duration and articulatory tension between the laryngeal categories is insignificant, and both are lenis (see Section 2.3.5). This has affected transcription practice, such that /b d g/ are usually narrowly transcribed as [b̥ d̥ g̊] and /p t k/ as [b̥ʰ d̥ˢ g̊ʰ].

[5]Recall that languages displaying this pattern are actually quite rare (Katz 2016).

The terms fortis and lenis are used in distinct ways in the phonetic and phonological literature. One use is as an arbitrary label for stop contrasts in languages which do not depend on closure voicing. Fortis–lenis has often been used in this sense when discussing Germanic languages, where the historic voiced–voiceless distinction has a diverse set of phonetic reflexes in the modern languages (Kohler 1984; Henton et al. 1992). Another use of fortis–lenis is as a phonetically substantial phonological feature referring to force of articulation. This feature may correlate with pulmonic force, muscular tension, closure duration, or indeed closure voicing (see Jaeger 1983 and references therein). Either use of the terminology is usually too imprecise for phonetic or phonological description.

Articulatory studies of carefully read speech have shown that intervocalically before stressed syllables, /b p/ are both accompanied by a glottal opening gesture in Danish (Frøkjær-Jensen et al. 1971), and EMG studies confirm that the posterior crico-arytenoid muscles are active in achieving this gesture (Fischer-Jørgensen and Hirose 1974; Hutters 1985); see Section 2.3.3 for more details. Similar studies of English found no such gesture during /b/ (Sawashima 1970; Hirose and Gay 1972); in Icelandic, which has a contrast between unaspirated and pre-aspirated stops in intervocalic position, both series of stops have a significant glottal spreading gesture (Pétursson 1976). Hutters (1985) proposes that the glottal spreading gesture in Danish is a measure taken to reinforce voicelessness in /b/, although she does not resolve this question; Möbius (2004) has shown that a glottal spreading gesture maintains voicelessness in German intervocalic stops, and Pape and Jesus (2014) have shown the same for European Portuguese and Italian.

Iverson and Salmons (1995) and Basbøll (2005) assume that the laryngeal contrast in Danish is managed with [spread glottis] (see Section 2.4.3.1). Motivation for this comes from the process of progressive sonorant devoicing, which has recently been shown by Juul et al. (2019) to be much less categorical than usually assumed (see Section 2.3.1 and Puggaard-Rode et al. forthc.). Kingston and Diehl (1994) assume that Danish stops are distinguished by an abstract [voice] feature. Motivation for this comes from the finding that /b d g/ trigger local F_0-lowering. Recall from Section 2.3.3 that Petersen (1983) does indeed find such an effect, but he crucially also finds that both laryngeal series trigger local F_0-raising relative to nasals. As Goldstein and Browman (1986) point out, this is consistent with an account where F_0-perturbations are the direct result of glottal aperture, something that Kingston and Diehl (1994) explicitly reject. Nevertheless, Kingston and Diehl's dichotomy between automatic and controlled phonetics can potentially account for both the presence and absence of closure voicing in Danish intervocalic stops.

As mentioned in Section 4.1 above, a number of facts about Danish stops make it difficult to predict the relative likelihood of intervocalic voicing. First of all, most of the relevant literature assumes that intervocalic voicing is categorical or near-categorical. Muscular tension is overall low in Danish stops, which should increase the chances of voicing, but all Danish stops are also accompanied by a glottal spreading gesture, which should decrease the chances of voicing. Closure duration is shorter and muscular tension weaker in the production of /p t k/ relative to /b d g/, but the glottal spreading gesture in /p t k/ has a greater magnitude and longer duration.

The results of this study allow us to compare some of the predictions from different approaches to phonological laryngeal representation, as laid out in Section 4.2. If [spread glottis] is indeed the only active laryngeal feature in Danish, we would predict at most variable voicing in /b d g/, and very negligible voicing in /p t k/ (following Beckman et al. 2013). If the laryngeal contrast is maintained with phonologized glottal gestures (as in Articulatory Phonology), it follows that the two series have underlying glottal spreading gestures of different magnitudes, both of which are expected to counteract voicing. A gestural account predicts that lenition leads to a reduction in the magnitude of these gestures, potentially causing voicing in both laryngeal series, but more readily in /b d g/. Neither of the two featural accounts (i.e. the abstract and concrete [voice] approaches) make any clear predictions about lenition and voicing.

4.4 Research questions and predictors

I set out with the following research questions in mind:

(1) Is there a difference in how frequently members of the two laryngeal series are voiced intervocalically?

The known facts about Danish stop production point in different directions. If the vocal folds were in a neutral, adducted position during the closure, one would expect a higher likelihood of continuous closure voicing in /p t k/, since they have shorter closure duration and a lower degree of articulatory force (Fischer-Jørgensen 1972b). However, for both series of stops, although to varying degrees, the vocal folds are actively spread during the closure. As such, I hypothesize that /b d g/ are voiced more frequently than /p t k/. This seems intuitively obvious and is explicitly predicted from both a concrete [voice] account and a gesture-based account of the contrast.

(2) Is closure voicing in intervocalic stops a lenition phenomenon?

From an aerodynamic perspective, voicing is natural in intervocalic stops, but there is evidence that voicing is actively blocked in all Danish stops. I test whether intervocalic voicing is more common in environments where lenition is generally expected, which would be predicted from gesture-based underlying representations.

(3) What factors predict closure voicing, and how large are their relative effects?

In addition to phonological laryngeal category and lenition, a host of other phonetic and extraphonetic factors are known to or can be expected to affect the probabilistic occurrence of consonant voicing (as established by e.g. Shih and Möbius 1998; Möbius 2004; Strycharczuk 2012). I aim to take as many of these as possible into account in order to explore their relative influence in Danish. These factors are presented in detail below.

4.4.1 Potential predictors

The detailed annotations of the DanPASS corpus (see Section 4.5.1) make it possible to test how a large number of (mostly categorical) predictors affect the rate of closure voicing. These predictors relate to segmental, prosodic, morphosyntactic, and other factors, which are discussed in the following sections.

4.4.1.1 Segmental predictors

The stops are coded according to LARYNGEAL CATEGORY and PLACE OF ARTICULATION. There is really no theory-neutral way to refer to the two laryngeal categories. Here, the terms ‘aspirated’ and ‘unaspirated’ are used as short-hand terms for the surface contrast between /p t k/ and /b d g/ in distinct speech, as mentioned in Section 4.3.

Place of articulation is expected to influence the likelihood of voicing, such that the chances of voicing are lower in occlusions further back in the oral cavity. This is aerodynamically motivated (see Section 4.2 for more details), and is reflected typologically: voiced velar stops are less common than alveolar ones, which are in turn less common than bilabial ones (Gamkrelidze 1975; Ohala 1992; Hayes and Steriade 2004; Brown 2006). Since bilabial and alveolar occlusions are physically quite close, and velar occlusions are significantly further back, a place effect should be most noticeable in velar stops.

The quality of surrounding vowels is expected to influence the likelihood of closure voicing; recall that Danish has an exceptionally complex vowel system (see Section 1.4). HIGH VOWELS are expected to decrease the chances of voicing, since high vowels have a tighter constriction in the oral cavity, making them less sonorous and more likely to devoice (e.g. Mortensen 2012). High vowel devoicing is caused by a constriction in the oral cavity, which makes it difficult to maintain voicing over time due to rising supraglottal air pressure. A preceding high vowel should decrease the chances of voicing more than a following high vowel. The following are considered high vowels: [i y ɪ ʏ e ø u ʊ o]. Note that these transcriptions are adapted to Danish (see Section 1.4); many of these vowels are higher than their conventional IPA counterparts, and they all have a mean $F1 < 400$ Hz in Modern Standard Danish (Juul et al. 2016).

In locating intervocalic stops, SEMIVOWELS were also considered vowels. I assume that semivowels occurring immediately before the intervocalic stop decrease the chances of voicing, simply because these are less sonorous than nuclear vowels (e.g. Parker 2002). The sounds in question are [ɪ̯ ʊ̯ ɐ̯ ɤ̯], as well as their syllabic counterparts [ɪ ʊ ɐ ɤ].

As discussed above, intervocalic voicing in Danish may be a lenition phenomenon resulting from passive voicing lasting throughout the closure. Therefore, voicing is expected to be more likely in environments that are associated with lenition. Surrounding NEUTRAL VOWELS should increase the chances of voicing, since the Danish neutral vowels [ə ɐ] generally occur in prosodically weak syllables (e.g. Basbøll 2005), where lenition is strongly expected. Vowel neutrality is strongly negatively correlated with stress: as a general rule, syllables with neutral vowels are always unstressed, but not all unstressed syllables have a neutral vowel. Preceding and following neutral vowels were coded separately, but their influence on closure voicing is expected to be essentially the same.

4.4.1.2 Prosodic predictors

STRESS generally reduces the chances of lenition phenomena occurring, so it is also expected to reduce the chances of closure voicing. If the preceding syllable is stressed, this is expected to increase the chances of voicing, as it is unlikely for two syllables in a row to be stressed. These two variables are coded independently, and one is not predictable from the other; there may be multiple consecutive unstressed syllables.

Adjacent STØD is expected to reduce the chances of voicing, no matter whether on the preceding syllable or the syllable in question.[6] Stød is phonetically akin to creaky voice; it is cued with low pitch and relatively aperiodic voicing during the final part of a long sonorant rhyme (Grønnum and Basbøll 2001, 2007). Stød is produced with laryngeal contraction regulated by vocalis and lateral crico-arytenoid

[6] At the morpheme level, primary stress is a phonological prerequisite for stød. In compounds, however, primary stress generally falls on the first member, while the second member has stød; some derivational processes also behave this way, and stød interacts with inflectional morphology in complex ways (e.g. Basbøll 2003). Furthermore, morpheme level stress is not necessarily realized at the sentence level, and morphemes can lose stress at the sentence level while retaining stød. As such, stød and stress are far from perfectly correlated in the DanPASS corpus; in fact, a small majority of syllables with stød are unstressed.

muscles (Fischer-Jørgensen 1987, 1989). Recall from Westbury and Keating (1986; see Section 4.2) that closure voicing is natural intervocalically, assuming that the vocal fold configuration is amenable to voicing; this is the case for vowels with modal voicing and less so for vowels with stød. As such, stød on the preceding syllable is expected to decrease the chances of continuous voicing. Although stød mainly affects the final part of syllables, it is also cued syllable-initially with many of the same articulatory and acoustic correlates as stress: increased airflow, pharyngeal pressure, intensity, pitch, and articulatory force (Smith 1944; Fischer-Jørgensen 1987, 1989). As such, stød on the syllable itself is also expected to decrease the chances of continuous voicing, but less so than stød on the preceding syllable.

4.4.1.3 Morphosyntactic predictors

The type of MORPHOLOGICAL BOUNDARY at which the intervocalic stop occurs was coded. These include word boundaries, boundaries between roots and (derivational and inflectional) affixes, boundaries between separate parts of compounds, as well as *none*, if the intervocalic stop occurred morpheme-internally. Prefixes in Danish are exclusively derivational; suffixes are mostly inflectional, but can also be derivational. As consonants tend to be strong domain-initially (e.g. Keating et al. 2004), it would be more optimal to code the individual syllables for their position in a prosodic hierarchy (e.g. Nespor and Vogel 1986), but such a coding cannot be easily extracted from the existing DanPASS transcriptions. I hypothesize that the morphological boundary predictor is hierarchical in its influence on closure voicing, as it has been shown that intergestural articulatory timing is more stable within-word and within-morpheme than across words and morphemes (Byrd et al. 2000; Cho 2001). I therefore assume that morpheme-internal stops have higher likelihood of voicing than word-internal stops at morphological boundaries; and these in turn have higher likelihood of voicing than stops at word boundaries. Among morphological boundaries, the following hierarchy of morpheme boundary types is assumed: inflectional > derivational > compound.[7]

There are reasons to assume that stops at inflectional morpheme boundaries might be voiced at much higher rates than stops in other positions: inflectional suffixes are always unstressed and always have neutral vowels in Danish. Following a usage-based framework such as Exemplar Theory (e.g. Bybee 2001), stops in inflectional affixes may also be voiced more often simply because language users encounter voicing more often in affixes, and as such it is weighted higher in the underlying representation of affixes at a morpheme-specific level. Several phonological frameworks assume that morphology is invisible to phonetic interpretation and would thus predict that morpheme-specific underlying representations are impossible; this is the case in e.g. Lexical Phonology (Kiparsky 1985). However, recent studies show that specific morphemes can exhibit phonetic patterns that are not predictable from their phonemic makeup. Plag et al. (2017; see also Tomaschek et al. 2021) have found that the English 'homophonous' *s*-suffixes (third person singular present tense, plural, etc.) differ systematically in phonetic realization; consider e.g. the suffixes in the present tense verb [pʰɛts] 'pet+s' and in the plural noun [pʰɛts] 'pet+s'. The present tense ending is systematically longer than the plural ending. Heegård (2013) found similar results for variable rates of schwa-deletion in homophonous [-tə] *-te* suffixes in Danish; [peˈstɛmˀtə] *bestemte* can either be a past tense verb meaning 'decided' or a definite past participle meaning 'certain', both of which derive from the verb [peˈstɛmˀm̩] *bestemme* 'decide'; schwa-deletion is systematically more likely in the past tense form, meaning the two are only near-homophonous.

Additionally, words were coded as being members of either a CLOSED or an OPEN word class. Words from closed classes are often function words, and these often show significant phonetic reduction (e.g. Bell et al. 2003; Schachtenhaufen 2013).

[7] An alternative approach would be to use Basbøll's (2005: 351) complex hierarchy of graded productivity of morphological endings. However, Basbøll's hierarchy only covers inflectional endings, and the added complexity of Basbøll's hierarchy would potentially make the statistical results very difficult to interpret.

4.4.1.4 Other predictors

In addition to the predictors mentioned above, a lexical frequency measure was included. High lexical frequency is known to cause phonetic reduction, both in the course of language change (e.g. Hooper 1976; Bybee 2000a) and synchronically (Bybee 2000b; Pierrehumbert 2001; Pluymaekers et al. 2005), and has been shown to specifically increase voicing assimilation in Dutch (Ernestus et al. 2006). Although the speech in DanPASS is spontaneous, it is also nested within specific experiments, where lexical frequencies can differ substantially from the ambient language (see Section 4.5.1). Since contextual probability has also been shown to increase phonetic reduction (Jurafsky et al. 2001, 2002), LOCAL LEXICAL FREQUENCY was coded, i.e. lexical frequency in the DanPASS corpus itself. This measure is available in the online version of DanPASS (Grønnum 2016). This was compared to a more general measure of lexical frequency based on the much larger LANCHART corpus (which was also used in Section 3.3.3). Due to the experimental nature of DanPASS (the map task in particular, see Section 4.5.1), many frequent words in DanPASS do not occur in LANCHART at all. This means that modeling with general (LANCHART) frequencies rather than local (DanPASS) frequencies would require excluding just over 300 items, around 8% of the total number of tokens, particularly in the morphological compound category. The two frequency measures are further strongly correlated (r = .78). Given the strong correlation and the disadvantages of using general frequencies, only local frequency is used in the statistical modeling.

INDIVIDUAL WORDS were also coded, since Pierrehumbert (2002) mentions a number of cases where word-specific phonetic encoding goes beyond simple lexical frequency and contextual predictability; this relates directly to the discussion of Exemplar Theory in Section 4.4.1.3. Word-specific effects are not explored in any detail.

A local measure of speech rate was also included. LOCAL SPEECH RATE should affect the chances of voicing for aerodynamic reasons: unless inhibited, post-vocalic voicing should automatically continue for a certain amount of time during a stop closure (see Section 4.2). A higher speech rate also causes decreased closure duration (as demonstrated for Danish by Andersen 1981a), which increases the chances that voicing continues throughout the closure phase. Local speech rate is measured here as the combined duration in seconds of the two syllables flanking the intervocalic stop.[8]

Finally, a few extralinguistic factors pertaining to the speakers were coded. SEX has been shown to influence the degree of closure voicing, such that men are more likely than women to produce fully voiced stops (Swartz 1992; Ryalls et al. 1997). This could be aerodynamically motivated; on average, men have larger supralaryngeal cavities than women, causing a slower rise in supraglottal air pressure during closure. An alternative explanation for the same outcome is that women generally speak more 'clearly' than men (as demonstrated by e.g. Ferguson 2004 for vowel intelligibility), and show less of a tendency for lenition.[9] I am not aware of studies connecting AGE with closure voicing directly, but it has been shown that speech rate decreases with age (Seifert 2009), suggesting that lenition will also decrease with age.

Finally, the INDIVIDUAL SPEAKERS were coded. Sonderegger et al. (2020) recently showed that the implementation of closure voicing in Glasgow Scots is highly speaker-specific even when controlling for a large number of other factors, and Tanner et al. (2020) found similar results for Japanese. Speaker-specific effects are not explored in any detail.

The potential predictors and the directionality of their expected influence on closure voicing are summarized in Table 4.1.

[8] This is admittedly a rough measure of speech rate, chosen mostly out of convenience (it was easy to extract from the existing data). It is not unheard of, though; Bohn (2013) also measures duration of target syllables in his study comparing Danish infant directed speech and adult directed speech. The measure is presumably not too rough for the purpose of the current study, as the result show a very strong effect of speech rate.

[9] The latter explanation is almost certainly an effect of gender rather than biological sex.

Table 4.1: *Potential predictors and the expected directionality of their influence on closure voicing.*

Variable	Predicted likelihood of voicing	Notes
LARYNGEAL CATEGORY	unaspirated > aspirated	
PLACE OF ARTICULATION	bilabial > alveolar > velar	strongest effect for velar stops
ADJACENT SEMIVOWEL	decreased	
ADJACENT HIGH VOWEL	decreased	strongest effect preceding the stop
ADJACENT NEUTRAL VOWEL	increased	
STRESS	unstressed > stressed	
PRECEDING STRESS	stressed > unstressed	
ADJACENT STØD	decreased	strongest effect preceding the stop
MORPHOLOGICAL BOUNDARY	internal (no boundary) > inflectional > derivation > compound > word	
WORD CLASS TYPE	closed > open	
LOCAL LEXICAL FREQUENCY	increases with frequency	
LOCAL SPEECH RATE	increases with speech rate	
LEXICAL ITEM	random	
SEX	men > women	
AGE	decreased with age	
INDIVIDUAL SPEAKER	random	

4.5 Methods

4.5.1 The DanPASS corpus

The DanPASS corpus (Grønnum 2009, 2016) is used to answer the research questions posed in the previous section. The corpus consists of native speakers of Danish solving a number of unscripted tasks, either alone or in pairs. An original motivation behind creating the corpus was to counteract the bias for highly controlled scripted speech in studies of Danish phonetics. The recordings in the DanPASS corpus are unquestionably also laboratory speech, but they are much more spontaneous than what was previously the standard. Grønnum (2009) rightly points out that formal laboratory speech is well-suited for some phonetic studies; some phenomena are rare enough that they can be difficult to find sufficient examples of even in very large corpora, and sometimes it can be important to carefully control for interacting phenomena. For a phenomenon such as intervocalic closure voicing, where the triggering environment is very frequent, and informal speech can perhaps in itself be considered a triggering environment, basing the analysis on informal speech is crucial. The corpus has already been the basis for major contributions to our understanding of Danish speech, in the areas of consonant reduction (Pharao 2009), and phonetic reduction in general (Schachtenhaufen 2013), as well as intonation and prosody (Tøndering 2003, 2008; Grønnum and Tøndering 2007). As mentioned in Section 2.3.1, it is also basis for the most extensive investigation of voice onset time in Danish stops (Mortensen and Tøndering 2013).

The full DanPASS corpus consists of monologues recorded in 1996, and dialogues recorded in 2004. While the dialogues probably constitute a more natural speech setting, they are also somewhat more challenging to analyze. For this reason, the current study only makes use of the monologues. Monologues were recorded from 18 speakers: 13 men and 5 women. The speakers were between 20–68 years old, with a mean age of 29. Overall, the monologues constitute 171 minutes of speech, with a mean duration of 9*m*27*s* of speech per speaker (range 6*m*13*s*–15*m*49*s*). Technical details are reported by Grønnum (2009). The speakers were recorded performing three different tasks. *Network* is a description of various shapes and colors, based on a design by Swerts and Collier (1992). *City* is a description of a number of routes through a drawn city map, based on a design by Swerts (1994). *House* is a description of how to build a house model using a number of buildings blocks, based on a design by Terken (1984).

The recordings are accompanied by quite detailed annotations in Praat (Boersma 2001; Boersma and Weenink 2021). Segmentations are made at the levels of prosodic phrase, word, and syllable, whenever these could be segmented with reasonable certainty. While the syllable is a well-defined phonological unit in Danish (see e.g. Basbøll 2005), it is often difficult to find well-defined phonetic units corresponding to those (Schachtenhaufen 2010a, 2010b), particularly because syllabic sonorants and consecutive syllables consisting of homorganic vowels are abundant in Danish speech due to schwa assimilation. The recordings are annotated orthographically, phonemically, and phonetically. They are also coded for morphology and accompanied with parts-of-speech tags and annotations for pitch movements and stress. The phonetic transcriptions use Grønnum's (1998) standards for transcribing Danish and are generally rather narrow except where stops are concerned. [p t k] are used where aspirated stops would be expected in distinct speech, and [b d g] where unaspirated stops would be expected in distinct speech, regardless of phonetic implementation.[10] In other words, closure voicing during stops is ignored in the transcriptions. Grønnum (2009) does not motivate this, and perhaps as a result, previous studies using DanPASS to examine e.g. stops (Pharao 2009, 2011; Mortensen and Tøndering 2013; Schachtenhaufen 2013) have also ignored the distinction between stops with and without closure voicing.

4.5.2 Acoustic analysis

I used a Praat script (written by Dirk Jan Vet; see Puggaard-Rode et al. 2022b) to locate intervocalic stops in the DanPASS monologues, i.e. stops that do not occur initially in a prosodic phrase and are flanked on both sides by (semi-)vowels. Approximants were included because there are well-defined phonological processes whereby they syllabify and become phonetic nuclear vowels (Brink and Lund 1975; Basbøll 2005; Schachtenhaufen 2010b). The DanPASS transcriptions are segmented at the level of phonetic syllable, and only stop-initial intervals were included in the study. This is in line with the studies of glottal activity cited in Sections 2.3.3 and 4.3 above, which also focus on syllable-initial intervocalic stops. This decision makes the coding and interpretation of predictor variables considerably easier. For each stop, the surrounding syllables are isolated, i.e. the stop-initial syllable and the preceding syllable. The script creates a sound file and TextGrid file containing all such syllables from the DanPASS monologues. This sound file lasts just under 24 minutes and contains a total of 3,744 inter-

[10]Except in the case of tapped realizations of /t ~ d/, which are transcribed with with [ɾ].

Table 4.2: *Intervocalic stops in the DanPASS monologues by phoneme.*

Phoneme	Number	By-speaker range
/b/	189	3–25
/d/	1,278	28–167
/g/	752	26–65
/p/	327	8–32
/t/	431	16–39
/k/	767	24–67
Total	3,744	117–341

vocalic stops, with an average of 204.7 stops per speaker (range 117–341). They are broken down by phoneme in Table 4.2.

The variance in lexical frequency is rather extreme. There are 303 unique lexical items in the data, with an average of 12.4 observations per item, albeit a median of just 2 observations (range 1–303). 125 lexical items occur just once, while the 10 most frequent items occur a total of 2,025 times, i.e. they make up more than half of all tokens.

For each of the intervocalic stops, I manually checked if it was voiced throughout the closure. This was done by visually inspecting the waveform: constant periodicity up to the burst was taken as continuous closure voicing.[11] This method was relatively straightforward to implement, although it is certainly a simplification of the complexity in the phonetic signal. Figure 4.1 shows waveforms of stops from both laryngeal series that show continuous voicing and interrupted voicing, respectively.

Recent comparable studies by e.g. Davidson (2016), Sonderegger et al. (2020) and Tanner et al. (2020) all use a three-way distinction between *voiceless*, *partially voiced*, and *fully voiced*. However, none of these studies focus particularly on intervocalic stops.[12] There are two main reasons for not adopting a three-way distinction in this study: 1) Multi-valued categorical dependent variables are much more

[11] Whenever stops from the /p t k/ series were fully voiced, they typically also had breathy voiced release. As a result, intervocalic voicing is very unlikely to lead to neutralization.

[12] Davidson (2016) focuses exclusively on phrase-medial position, so it is likely that many stops in that study were also intervocalic.

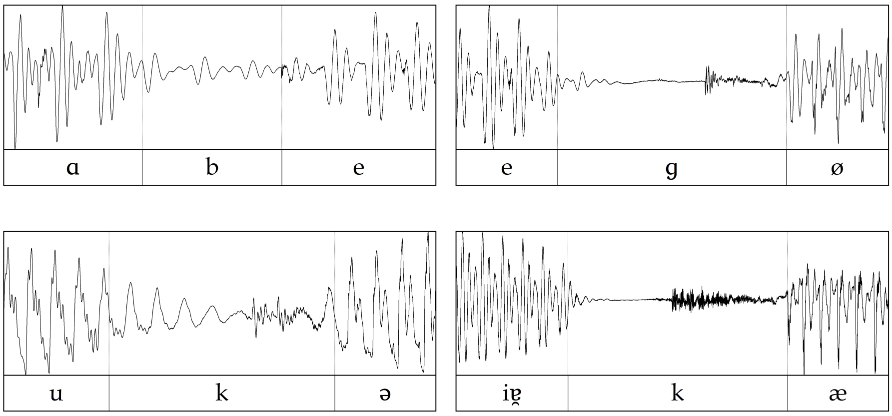

Figure 4.1: *Waveforms exemplifying: 1) A fully voiced token of* /b/ *in the phrase* [fʁɑ beˈkønˀl̩sn̩] fr(a be)gyndelsen *'from the start'. 2) A mostly voiceless token of* /g/ *in the same phrase as 1,* <fra b(egy)ndelsen>. *3) A fully voiced token of* /k/ *from the phrase* [tu gʱə] d(u ka)n *'you can'. 4) A mostly voiceless token of* /k/ *from the word* [fie̯ˈkʰænˀt] f(irka)nt *'square'.*

difficult to model than binary variables, and 2) fully voiceless intervocalic stops are known to be uncommon, so a *voiceless* category would likely have added little explanatory value. In intervocalic position, passive voicing essentially affects the first part of the following closure, regardless of the laryngeal category of the stop. This has been shown for at least Standard Chinese (Shih and Möbius 1998), German (Möbius 2004), and American English (Davidson 2016) aspirated stops, and for voiceless stops in several other languages (Shih et al. 1999). In an unpublished conference paper, Puggaard et al. (2019) showed that in Danish carefully read lab speech, there was no significant difference between the two laryngeal series in the relative duration of voicing during closure in stressed intervocalic position; both /b/ and /p/ were voiced for approximately one fourth of their closure. We compared this to Dutch, a so-called 'true voicing language', where the majority of intervocalic /b/ tokens were voiced throughout their closure duration.

This was a further argument for considering 'fully voiced' to only mean continuous voicing throughout the closure here.

Ideally, I would be working with a continuous measure of closure voicing, possibly measuring both intensity and relative duration of voicing. It is quite possible that true effects of some lower-level variables on voicing are masked in this study because of the relatively rough voicing measure. However, this would require much more fine-grained segmentation of the sound files.

4.5.3 Statistical analysis

All statistics used in the current study were calculated using the R statistical environment (R Core Team 2021; RStudio Team 2022) with a number of add-on packages.[13] I am interested in both exploratory data analysis and confirmatory statistics, although the analysis is not confirmatory in the strict sense of Baayen et al. (2017; see Section 7.7). The precise methods for the statistical analyses are described in Sections 4.6 and 4.7.1 below, respectively.

4.6 Exploratory analysis

In this section, I take a closer look at the data and explore correlations between the individual predictors and the presence of intervocalic voicing. This is a useful first look at patterns in the data, but in the next section, I proceed to build a regression model which provides a better picture of the complexities found in the data.

[13]Logistic mixed effects model were fitted using the `lme4` package (Bates et al. 2015b, 2021). The `car` package used for calculating variance inflation factors (Fox and Weisberg 2019; Fox et al. 2021), the `MuMIn` package for calculating model effect size (Barton 2020), and the `moments` package for checking distributions of continuous variables (Komsta and Novomestky 2015). The `ggplot2` package was used for generic visualizations (Wickham 2016; Wickham et al. 2021) and the `sjPlot` package for visualizing model coefficients (Lüdecke 2021). More details can be found in Puggaard-Rode et al. (2022b).

4.6.1 Categorical predictors

Table 4.3 and Figure 4.2 show the proportions of voiced tokens for each level of each of the categorical variables. The majority of categorical variables show at least some degree of correlation with closure voicing in the direction predicted in Table 4.1 above.

4.6.1.1 Segmental predictors

LARYNGEAL CATEGORY shows a clear correlation in the expected direction. As predicted above, intervocalic voicing is quite rare in /p t k/, where it was only found in 5% of all tokens. Intervocalic voicing is more common in /b d g/, where it was found in 38% of all tokens. Hence, voicing is not the norm for /b d g/, even though it is sometimes described in the literature as being essentially categorical. In total, continuous closure voicing is found in 24.6% of all intervocalic stops in the corpus.

PLACE OF ARTICULATION does not pattern as predicted from our aerodynamically motivated expectations; as expected, bilabials are voiced more often than velars, but unexpectedly, alveolars are voiced at a much higher rate than either of the other places of articulation. Presumably, there are non-aerodynamic reasons for this. Alveolar stops are generally more frequent than other places of articulation (see Table 4.2), and they are found at a higher rate in function words. While the transcriptions do in principle indicate tapped realizations of the alveolar stops, this is likely somewhat inconsistent, such that some realizations transcribed as alveolar stops are in fact taps [ɾ]; these are presumably always voiced.

PRECEDING SEMIVOWELS, as expected, are less likely than nuclear vowels to correlate with voicing in the following stop.

The behavior of HIGH VOWELS goes against the predictions; high vowels were expected to decrease the chances of voicing, in particular preceding the stop. High vowels preceding the stop show a weak correlation in the expected direction, and high vowels in the same syllable in fact correlate positively with voicing. This is contrary to the aerodynamically motivated predictions but could have a number of other explanations: high vowels are found in a number of very frequent function words, and the syllabic semivowels [ɪ ʊ] are both included in

Table 4.3: *Table of proportion of fully voiced tokens for each level of each categorical variable. Variables marked + show correlations in agreement with the hypotheses in Table 4.1, and ones marked – show disagreement with the hypotheses.*

Variable	Level	% voiced	no. voiced	
LARYNGEAL CATEGORY	Aspirated	5.05	77	+
	Unaspirated	38	844	
PLACE OF ARTICULATION	Bilabial	17.25	89	–
	Alveolar	35.75	611	
	Velar	14.55	221	
PRECEDING SEMIVOWEL	Absent	25.81	832	+
	Present	17.08	89	
HIGH VOWEL	Absent	22.38	584	–
	Present	26.69	337	
PRECEDING HIGH VOWEL	Absent	25.18	748	+
	Present	22.38	173	
NEUTRAL VOWEL	Absent	22.31	747	+
	Present	43.94	174	
PRECEDING NEUTRAL VOWEL	Absent	26	612	–
	Present	22.23	309	
STRESS	Absent	26.13	712	+
	Present	17.36	129	
PRECEDING STRESS	Absent	25.31	637	–
	Present	23.15	284	
STØD	Absent	26.39	792	+
	Present	17.36	129	
PRECEDING STØD	Absent	25.26	914	+
	Present	5.56	7	
MORPHOLOGICAL BOUNDARY	Internal	36.12	95	–
	Inflection	68.75	110	
	Derivation	38.81	26	
	Compound	9.76	73	
	Word	24.61	617	
WORD CLASS TYPE	Open	19.33	407	+
	Closed	31.38	514	
SEX	Female	20.96	192	+
	Male	25.75	729	

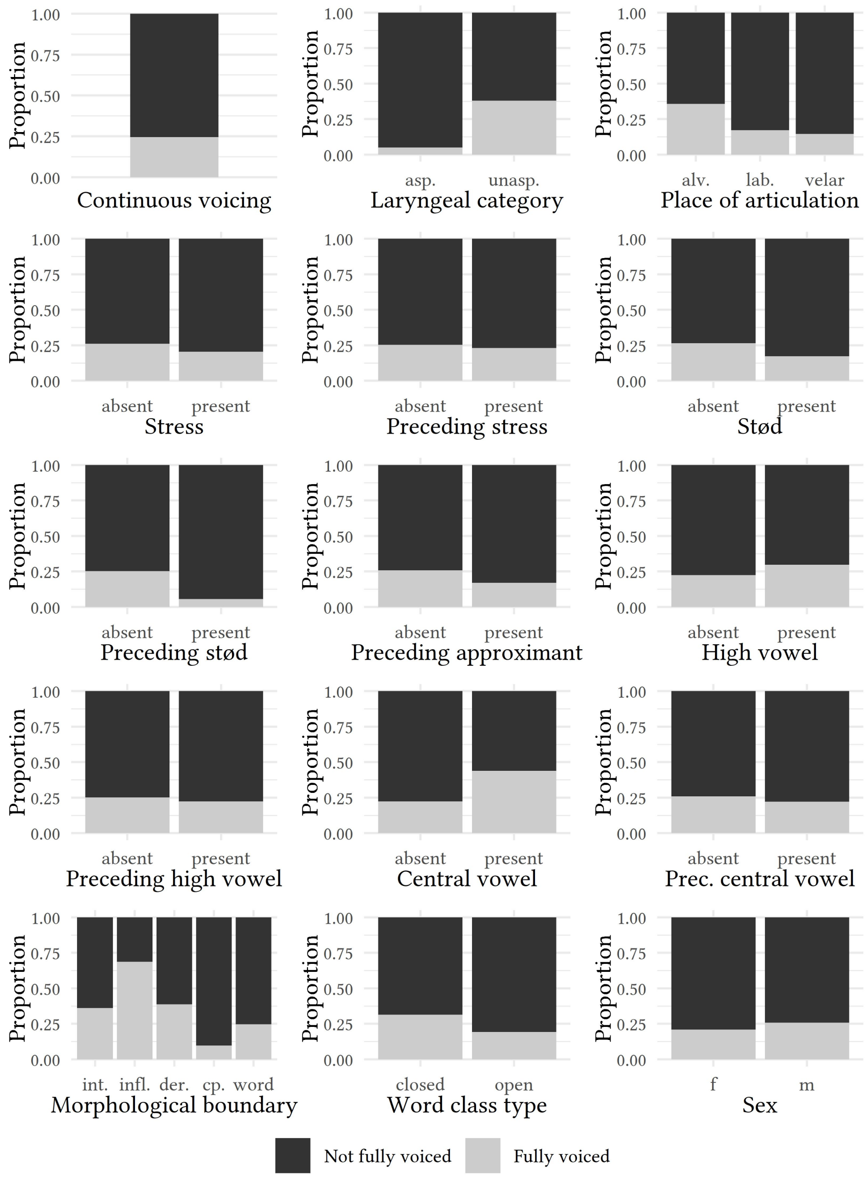

Figure 4.2: *Stacked bar plots showing the proportions of tokens with and without continuous voicing for each level of each categorical variable. (Morphological boundary levels: internal, inflectional, derivational, compound, word).*

this group. As such, there are predictable reasons why syllables with high vowels might frequently undergo phonetic reduction.

As predicted, NEUTRAL VOWELS in tautosyllabic position correlate positively with the presence of closure voicing. However, against expectations, neutral vowels in the preceding syllable show a slight correlation with the absence of closure voicing.

4.6.1.2 Prosodic predictors

As predicted, voicing is more common in UNSTRESSED than STRESSED syllables. Surprisingly, the presence of stress on the preceding syllable shows a (very weak) correlation in the unexpected direction. Also as predicted, voicing is less common in syllables with STØD and is exceedingly uncommon following syllables with stød.

4.6.1.3 Morphosyntactic predictors

The predictions regarding MORPHOLOGICAL BOUNDARY TYPE mostly do not pan out. By far the most voiced stops are at inflectional boundaries, with derivational morphemes and morpheme-internal stops being voiced at approximately the same rate. Stops at word boundaries, by far the most common category, show intervocalic voicing at around chance rate, i.e. the same rate as the data set at large. Finally, stops at compound boundaries are rarely voiced. Given the complexity of this factor, I will refrain from interpreting these results further until I present the results of the regression model.

As predicted, WORD CLASS TYPE interacts with closure voicing, such that members of the closed classes are voiced at a higher rate than members of open classes.

4.6.1.4 Other predictors

SEX correlates with voicing in the predicted direction, such that male speakers produce more voiced stops than female speakers.

4.6.2 Continuous predictors

Having discussed all categorical predictors, I now turn to the continuous ones. Figure 4.3 visualizes the proportion of stops with and without continuous voicing with density plots.

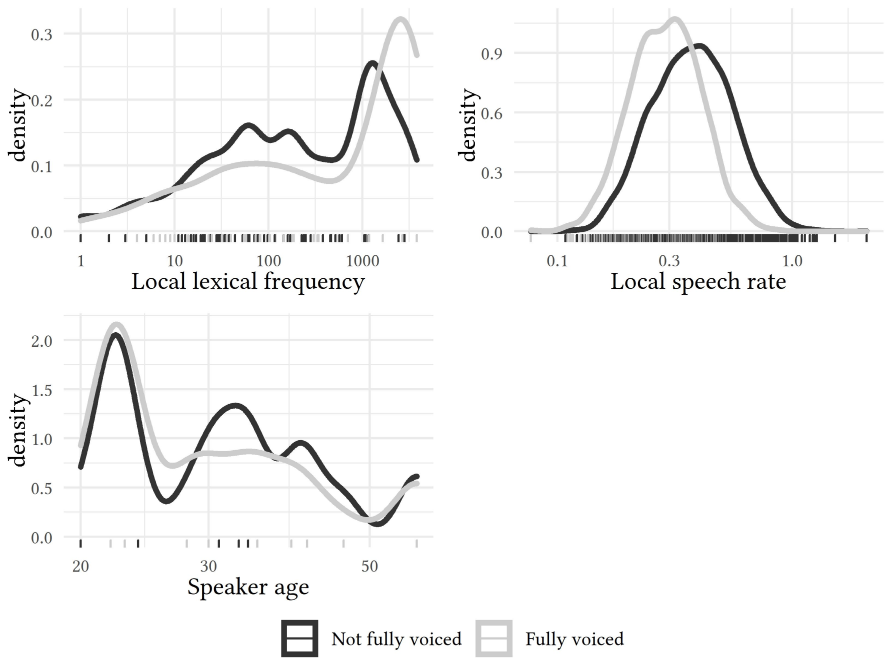

Figure 4.3: *Density plots showing the tokens with and without continuous voicing relative to continuous variables on a log-scale.*

It is clearly (and logically) the case that the most FREQUENT words also account for most tokens, both voiced and voiceless. It is also clearly the case that the words with very high frequency show a higher proportion of voiced tokens, and that the words with medium frequency, particularly between 50–500, show a higher proportion of voiceless tokens.

As predicted, SPEECH RATE correlates with voicing, such that voiceless tokens are more common during slow speech, and voiced tokens are more common during quick speech; recall that speech rate is coded as the duration of the syllables flanking the stop, so a low value (somewhat counter-intuitively) equals high speech rate. In both lexical frequency and speech rate, the distribution of fully voiced tokens is visibly more peaked than tokens which are not fully voiced.

We also see a correlation in the expected direction between age and voicing. Most speakers in the corpus are younger than 25 years old, so it follows naturally that most tokens, both voiced and voiceless, are also produced by this age group. It is, however, also the case that speakers in their thirties and forties produce a relatively higher proportion of voiceless stops.[14]

Having examined the correlations that are found in the data at face value, I now move on to analyzing the data with mixed-effects regression modeling.

4.7 Mixed-effects model

4.7.1 Model selection

The data come from a corpus that is not collected for studying intervocalic voicing, and I am interested in many independent variables. Given the lack of experimental control and the largely exploratory nature of the study, the data is presumably not structured in a way that allows for a maximal random effects structure; this is a common problem with mixed-effects models in linguistics (Meteyard and Davies 2020). This loss in optimal data structure comes with a corresponding gain in ecological validity, which is highly necessary when discussing potential lenition phenomena. There has been a lot of discussion of how to handle this issue in linguistics, with suggestions ranging from maximizing random effects (Barr et al. 2013) to balancing statistical power and Type I error by including only random effects that contribute sufficiently to the model's predictive power (Bates et al. 2015a; Matuschek et al. 2017). These papers generally assume 1) experimental data, and 2) a continuous dependent variable (i.e. linear mixed-effects models). This is important to keep in mind, since experimental data is generally more balanced than corpus data, and linear models are overall more likely to converge than logistic models with binary response variables (Seedorff et al. 2019). I opt for a data-driven

[14] The examples from above the age of 50 all come from a single speaker, so these can safely be ignored.

model selection procedure, inspired by the heuristics proposed by Sonderegger (2022).

The raw values of all continuous variables are positively skewed, so they were log-transformed in order to reach a normal distribution, and standardized to aid interpretation of the model.[15]

The categorical variables are contrast coded (see Schad et al. 2020). Sum contrasts were used for the binary variables. Variables corresponding to articulatory features are all coded as +½ ('present') and -½ ('absent'). laryngeal category is coded as -½ unaspirated, +½ aspirated; sex is coded as -½ female, +½ male; word class type is coded as -½ open, +½ closed. For the three-level variable place of articulation, two theoretically-guided Helmert contrasts were coded: one to test the distinction between velars and non-velars, and one to test the distinction between alveolars and labials, as in (4).

(4) *Velar contrast*: -⅓ bilabial, -⅓ alveolar, +⅔ velar
Bilabials vs. *alveolars*: +½ alveolar, -½ bilabial

The five-level morphological boundary variable is quite complicated. Here, four theoretically-guided Helmert contrasts were used: 1) *internal contrast*, testing the distinction between morpheme-internal and non-morpheme-internal; 2) *affix contrast*, testing the distinction between affix boundaries and other boundaries; 3) *affix type contrast*, testing the distinction between derivational affix boundaries and inflectional affix boundaries; and 4) *compound contrast*, testing the distinction between word boundaries and compound boundaries.

(5) *Internal contrast*: +⅘ internal, -⅕ inflectional, -⅕ derivational, -⅕ compound, -⅕ word
Affix contrast: +½ inflectional, +½ derivational, -½ compound, -½ word
Affix type contrast: +½ derivational, -½ inflectional
Compound contrast: +½ compound, -½ word

[15]Continuous variables were standardized by subtracting the mean and dividing by two standard deviations, following Gelman and Hill (2006).

The data were modeled using logistic mixed-effects regression.[16] The model selection procedure followed two steps: 1) Fixed effects selection with minimal random effects, and 2) pruning of the maximal random effects structure to achieve convergence with (almost) non-singular fit.

FIXED EFFECTS SELECTION. All independent variables which were theoretically motivated in Section 4.4 above are included in the model. There are no theoretical motivations for including interactions. However, we saw in Section 4.6 that the proportion of fully voiced /p t k/ tokens is near-floor, and this could be masking true effects in the data. For this reason, all possible interactions with LARYNGEAL CATEGORY were tested in a random intercepts-only model, in case some effects could be found only in /b d g/. Only significant interactions were kept.

RANDOM EFFECTS SELECTION. All meaningful by-speaker and by-item random slopes were subsequently added to the model; SEX and AGE can of course not vary by-speaker, and all by-item slopes for phonological or morphosyntactic variables are at least potentially problematic. I used strictly uncorrelated random effects; this leads to much higher convergence rates in logistic models, and Seedorff et al. (2019) show that it does not inflate Type I error rates even if the random effects are correlated in the underlying data (although it has a slight adverse effect on statistical power). This model converges with a singular fit, in this case because the model estimates zero-variances within some random slopes. In other words, the variance explained by these random slopes is not found to be different from that explained by random noise in the data. This is a symptom that the model is overparametrized, but should have no influence on the interpretation of the corresponding fixed effects (Brauer and Curtin 2018). All random slopes with estimated zero variances were removed, with the exception of LARYNGEAL CATEGORY. LARYNGEAL CATEGORY was kept since this is a variable of key interest in the study. This means that the resulting

[16]The model was fitted using the `glmer` function in `lme4`, using bound optimization by quadratic approximation (the `bobyqa` optimizer), with the maximal number of iterations increased from the default 10^5 to 10^6. These low-level mechanical details should have no effect on the results, but could be important for reproducibility. See fn. 13 for more details on the R packages used and Puggaard-Rode et al. (2022b) for code and data.

Table 4.4: *Summary of the final model.*

Simple fixed effects	Intercept, laryngeal category, place of articulation (velar contrast, bilabials vs. alveolars), preceding approximant, preceding high vowel, high vowel, preceding neutral vowel, neutral vowel, preceding stress, stress, preceding stød, stød, morphological boundary (internal contrast, affix contrast, affix type contrast, compound contrast), word class type, local lexical frequency, local speech rate, sex, age
Interactions with laryngeal category	Preceding approximant, preceding stress, local speech rate
By-speaker random effects	Intercept, laryngeal category (zero variance), velar contrast, high vowel, stress, stød, internal contrast, affix type contrast, compound contrast, local speech rate
Removed due to zero variance	Bilabials vs. alveolars, preceding approximant, preceding high vowel, preceding neutral vowel, neutral vowel, preceding stress, preceding stød, affix contrast, word class type, local lexical frequency
By-item random effects	Intercept, age, sex
Removed due to zero variance	Local speech rate

model is probably slightly overparametrized, since it is highly unlikely that there is *no* by-speaker variance for laryngeal category in the underlying data, but a reasonable interpretation is that the by-speaker variance for laryngeal category is very close to the random variation for laryngeal category in general. The entire model selection procedure is documented in Puggaard-Rode et al. (2022b), and the final model is summarized in Table 4.4.

None of the included independent variables show problematic collinearity; the variance inflation factor (*VIF*) is below 1.5 for all variables except those appearing in interaction effects.

The coefficients of a generalized linear model correspond to log-odds. These are suitable for regression modeling, as they are unbounded and normally distributed. Odds and odds ratios (*OR*s), on the other hand, are easier to interpret. In order to aid interpretability, I report both the model coefficients and standard error in the log-odds scale, and odds (ratio), which are computed by exponentiating the coefficients. The odds for the intercept can straightforwardly be interpreted as the odds of closure voicing with all other variables kept at zero. Since all variables are either contrast-coded or standardized, the *OR*s can be interpreted straightforwardly as the change in probability associated with that variable (see Sonderegger 2022: ch. 6). Odds and *OR*s are given as fractions; if $OR > 1$, the odds of voicing are higher in the variable level corresponding to + in the contrast coding, and if $OR < 1$, the odds of voicing are higher in the variable level corresponding to - . For the standardized continuous variables, *OR*s refer to the change in predicted likelihood of voicing associated with an increase of 1 standard deviation.

4.7.2 Results

The results of the logistic mixed-effects regression model described above is summarized in Table 4.5. No random effects table is included here, but it can be found in Puggaard-Rode et al. (2022b). The model has a reasonably high marginal effect size of $\Delta R^2 = 0.5$ and conditional effect size of $\Delta R^2 = 0.63$; this is the variance explained by the fixed effects alone and all effects combined, respectively.[17]

In some cases, the results of the mixed-effects model tell quite a different story from the exploratory analysis presented in Section 4.6. In these cases, the results of the mixed effects model should be taken as the best description of the data. The odds for the intercept means that the relative likelihood of a stop being fully voiced is predicted as 11.34 times lower than not being fully voiced if all other variables are ignored.

The significant effects overwhelmingly pattern as predicted. For the following categorical variables, this means that their effects

[17] See Nakagawa et al. (2017) for details of how this is calculated for generalized linear mixed-effects models.

Table 4.5: *Summary of logistic mixed-effects regression model. + indicates agreement with the hypotheses in Table 4.1, and – indicates disagreement with the hypotheses; no symbol indicates a null result. If nothing else is indicated,* OR *< 1 means that the odds of voicing is higher in the absence of a feature, and* OR *> 1 means that the odds are increased in the presence of that feature.*

Variable	**Odds (ratio)**	**coef**	**SE**	***z***	***p***		
(intercept)	1 : 11.34	-2.43	0.42	-5.72	<.001	***	
LARYNGEAL CAT., -asp +unasp	20.15 : 1	3	0.39	7.76	<.001	***	+
PLACE, velar contrast (+velar)	1 : 3.57	-1.27	0.39	7.76	<.001	***	+
PLACE, -bilabial +alveolar	1.16 : 1	0.15	0.34	0.44	0.66		
PRECEDING SEMIVOWEL	1.61 : 1	0.49	0.3	1.62	0.1		
PRECEDING HIGH VOWEL	1.1 : 1	0.1	0.16	0.59	0.55		
HIGH VOWEL	1 : 1.06	-0.06	0.25	-0.24	0.81		
PRECEDING NEUTRAL VOWEL	2.42 : 1	0.28	0.14	1.95	0.05	.	
NEUTRAL VOWEL	1.88 : 1	0.63	0.23	2.77	<.01	**	+
PRECEDING STRESS	3.7 : 1	1.31	0.24	5.41	<.001	***	+
STRESS	1 : 1.94	-0.66	0.21	-3.11	<.01	**	+
PRECEDING STØD	1 : 9.53	-2.25	0.52	-4.36	<.001	***	+
STØD	2.11 : 1	0.75	0.23	3.18	<.01	**	–
BND., internal contrast (+int)	1.28 : 1	0.25	0.32	0.78	0.44		
BND., affix contrast (+affix)	4.79 : 1	1.57	0.36	4.33	<.001	***	(+)
BND., affix type contrast (+inf)	3.13 : 1	1.14	0.61	1.87	0.06	.	
BND., comp. contrast (+cp)	1.57 : 1	0.45	0.41	1.1	0.27		
WORD CLASS, -open +closed	1 : 1.12	-0.11	0.31	-0.37	0.71		
LOCAL SPEECH RATE	1 : 18.29	-2.91	0.27	-10.86	<.001	***	+
LOCAL LEXICAL FREQUENCY	1.85 : 1	0.61	0.27	2.25	0.02	*	+
SEX, -f +m	1.7 : 1	0.53	0.44	1.22	0.22		
AGE	1 : 3.09	-1.13	0.41	-2.74	<.01	**	+
LAR.CAT. : PRECEDING GLIDE	1 : 2.59	-0.95	0.56	-1.71	0.09	.	
LAR.CAT. : PRECEDING STRESS	1 : 3.7	-1.31	0.47	-2.76	<.01	**	
LAR.CAT. : LOCAL SPEECH RATE	6.5 : 1	1.87	0.51	3.67	<.001	***	

are significant in the same (expected) direction as we saw in the exploratory analysis: LARYNGEAL CATEGORY, NEUTRAL VOWEL, STRESS, and PRECEDING STØD. The effect of laryngeal category is very strong, with the unaspirated set being more than 20 times more likely to be voiced intervocalically. The odds of voicing are approximately doubled in syllables with neutral vowels as well as in unstressed syllables, and the odds are around 10 times lower immediately following syllables with stød.

The PLACE OF ARTICULATION variable patterns differently from what we saw in the exploratory analysis. The model finds that voicing in bilabials and alveolars is around four times more likely than in velar stops, but there is no significant difference between bilabials and alveolars. This is in line with aerodynamically motivated predictions. Recall that alveolars were overall voiced at a much higher rate than other places of articulation; this effect disappears in a model that also takes e.g. stress, lexical item, and morphological structure into account.

There is a fairly strong effect of PRECEDING STRESS in the expected direction; voicing is around four times more likely following stressed syllables. This is interesting, as there was essentially no correlation between preceding stress and voicing in the exploratory analysis.

The STØD variable patterns in the opposite direction of the predictions and what we saw in the exploratory analysis. Closure voicing is found to be around twice as likely in syllables with stød. I return to this in the discussion in Section 4.8.3 below.

Only one of the contrasts for MORPHOLOGICAL BOUNDARY TYPE is found to have a significant effect on closure voicing: affix-initial stops are voiced at a much higher rate (around four times) than stops at other kinds of morphological boundary. There are good reasons to expect this at face value: /p t k/ are rarely found in affixes and never in inflectional affixes, affixes are almost never stressed, and affixes often have neutral vowels. However, these are all variables that are controlled for independently in the model, and because of this, I predicted that morpheme-internal stops would be voiced at a higher rate than affixes. I return to this in Section 4.8.2.

Other categorical variables – PRECEDING SEMIVOWEL, PRECEDING HIGH VOWEL, HIGH VOWEL, PRECEDING CENTRAL VOWEL, WORD CLASS TYPE, and SEX – have no significant influence on voicing in the model,

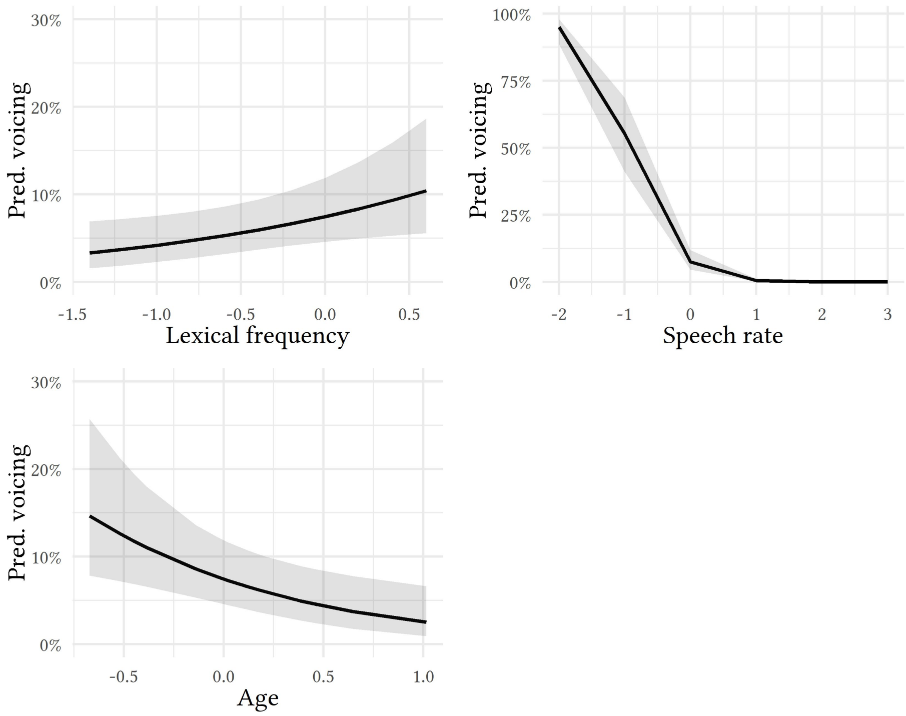

Figure 4.4: *Plots showing the likelihood of fully voiced stops of continuous variables as predicted from the mixed-effects model. The x-axes are standardized. Note that y-axis scales differ; due to the very high likelihood of voicing in very quick speech, keeping the scales identical would blur the effect in other variables.*

even though in some cases, there seemed to be clear correlations in the exploratory analysis. In all cases, we must assume that the correlation we saw at in the exploratory analysis can be better explained by other (potentially random) variables in the data.

The influence of continuous predictors is visualized in Figure 4.4. There is a clear increase in the predicted likelihood of voicing as lexical frequency increases, and a clear decrease in the predicted likelihood of voicing as age increases. The local speech rate variable in particular has an extremely strong influence on voicing, such that quicker speech leads to higher rates intervocalic voicing. In fact, the

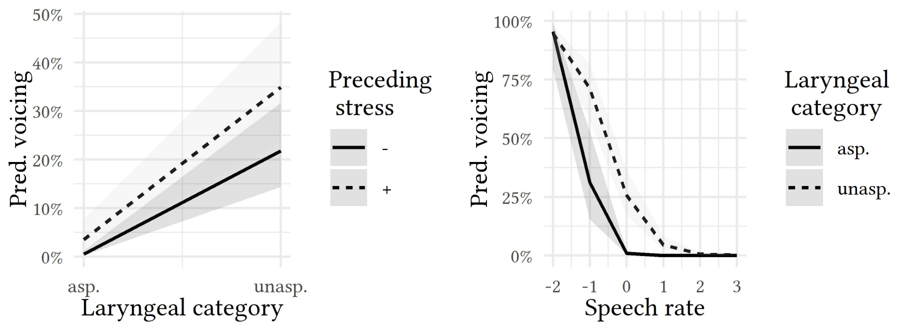

Figure 4.5: *Plots showing the likelihood of fully voiced stops of interaction effects as predicted from the mixed-effects model. The x-axes are standardized. Note that y-axis scales differ; due to the very high likelihood of voicing in very quick speech, keeping scales identical would blur the effect in other variables.*

predicted likelihood of voicing is near-ceiling for the quickest tokens, and near-floor for a large portion of the slower tokens.

Figure 4.5 shows the predicted significant interaction effects. The interaction effect between LARYNGEAL CATEGORY and PRECEDING STRESS patterns as predicted: there is a fairly marginal difference in predicted voicing after stressed syllables in /p t k/, whereas the effect is much more pronounced in /b d g/. The interaction effect between LARYNGEAL CATEGORY and LOCAL SPEECH RATE is similar: both laryngeal categories show near-ceiling voicing in the fastest tokens and near-floor voicing in the slowest tokens, but near-floor voicing is predicted in faster speech in /p t k/ relative to /b d g/.

4.8 Discussion

In this section, I discuss the results in relation to the research questions posed in Section 4.4.

4.8.1 Closure voicing and laryngeal category

The strongest predictor of closure voicing is laryngeal category. There are two main findings: 1) /p t k/ are voiced only very rarely, and much more rarely than /b d g/, and 2) /b d g/ are voiced commonly, albeit still at lower than chance rate. The three major accounts of laryngeal representation in Danish stops (and beyond) that I presented in Section 4.3 all straightforwardly predict the first finding, and all have mechanisms that can account for the second finding.

Abstract [voice] approaches predict the first finding: [+voice] stops are naturally voiced at a higher rate than [+voice] stops. With regards to the second finding, in Kingston and Diehl's (1994) abstract account of [voice], a controlled mechanism could be postulated that actively blocks voicing in [+voice] stops. Such a mechanism seems counterintuitive, but it is already independently needed for Icelandic, where intervocalic voicing of unaspirated stops is seemingly even more rare than in Danish (Pétursson 1976).

Concrete [voice] approaches also predict the first finding, but not necessarily the second finding. [spread glottis] generally blocks voicing, while unmarked stops are expected to be voiceless in initial position and voiced in intervocalic position.. In Beckman et al.'s (2013) account of [spread glottis], they assume that active privative features are reinterpreted as numerically valued features at some stage in the derivation. Since [spread glottis] is the active laryngeal feature in e.g. German, Danish, and Icelandic, /p t k/ are assigned high numeric values for [spread glottis], while /b d g/ are assigned lower values. They suggest that German /b d g/ are assigned [1sg], which allows for passive intervocalic voicing, and that Danish /b d g/ are assigned [5sg], which blocks passive voicing. This predicts the results quite well. Note, however, that other proponents of [spread glottis] in Danish (like Iverson and Salmons 1995 and Basbøll 2005) do not necessarily assume this mechanism; without such a mechanism, we would simply expect the unmarked /b d g/ to be near-categorically voiced, since this is the unmarked realization of stops in intervocalic position.

Ultimately, I believe the best explanation of the results is one that incorporates our existing knowledge of glottal activity in Danish stops from research by Frøkjær-Jensen et al. (1971), Fischer-Jørgensen and

Hirose (1974), and Hutters (1985). Recall from Section 4.3 that /p t k/ have shorter closure duration and are produced with less muscular tension than /b d g/. Either both series are phonetically lenis, or /b d g/ are in fact fortis. The shorter closure duration and lower muscular tension of /p t k/ would predict more closure voicing in the aspirated series if the vocal folds were properly adduced and tensed for voicing. In careful speech, however, all Danish stops are accompanied by a glottal spreading gesture, presumably to enforce voicelessness. The glottal gestures are different in magnitude across the laryngeal series. /p t k/ have a glottal spreading gesture of great magnitude that lasts throughout the closure and into the release, whereas /b d g/ have a smaller glottal opening gesture that peaks during the closure. Maintaining glottal spreading in /p t k/ is prioritized, because it is required for the aspirated release, which is the primary cue to the laryngeal contrast. Voicing is also actively blocked in /b d g/ through glottal spreading, but in these sounds, it is not crucial for maintaining the contrast. The differences in magnitude of the glottal gesture can explain both findings: the differences between the two series, and the fact that the majority of stops in spontaneous speech are not voiced throughout. Such fine-grained differences in duration and magnitude of gestures can be straightforwardly encoded in the gestural scores of Articulatory Phonology.

The results can be accounted for by all three major accounts of laryngeal representation, but not all theories predict the results equally well. Recall from Section 4.2 that an abstract [voice] account did not allow us to make any specific predictions. A concrete [voice] account only predicts the result with the added machinery of gradient phonetic interpretation of feature values. A gesture-based account predicts the results well with no additional machinery: the necessary 'ingredients', so to speak, are already built into the representational grammar.[18]

On a final note, recall that Schachtenhaufen (2022) has recently suggested abandoning the transcription standard using [b̥ d̥ g̊ b̥^{h} d̥^{s} g̊^{h}] in favor of [p t k p^h ts k^h], since fortis–lenis is not traditionally

[18]This is, of course, a direct result of the generative capacity of Articulatory Phonology being very powerful; this is an advantage here, but certainly also has drawbacks.

indicated in IPA, and IPA guidelines suggest using [b̥]-style transcription to indicate devoicing of sounds that are usually voiced. The study presented here further cements that Danish /b d g/ are *not* usually voiced: not only are /b d g/ categorically voiceless in most positions, voicing is also regularly blocked in the one syllabic position where it would actually be phonetically natural. I am therefore strongly in favor of Schachtenhaufen's proposal.[19]

4.8.2 Closure voicing as lenition

Closure voicing is mainly found in environments where phonetic lenition is expected: its occurrence increases with speech rate, and it is found at higher rates in unstressed syllables, in syllables with schwa, and in affixes. Based on these results, it seems sensible to consider intervocalic closure voicing a lenition phenomenon in itself.

This has some interesting phonological consequences. As discussed in Section 4.2, it is often difficult to account for intervocalic allophonic voicing with reference to a [voice] feature. In phonological representational frameworks relying on privative features, voiceless unaspirated stops are generally considered unmarked, i.e. they carry no laryngeal features. Similarly, voicing is generally not considered phonologically marked in sonorant sounds (e.g. Lombardi 1995). As such, [voice] is not specified in sonorant sounds and cannot spread from them. Besides, rather than being lenition, the addition of a [voice] feature to a stop entails an increase in markedness and a more complex underlying structure. Rice and Avery's (1989) non-laryngeal [spontaneous voice] feature, which can spread from adjacent sonorant sounds, may be able to represent the process; it does not, however, capture the lenition aspect, as it still entails the addition of phonological material. It also does not capture the probabilistic nature of the process' distribution. We can approach a statistical model of when continuous voicing is more or less likely to occur, but this does not allow us to predict its occurrence with any certainty.

The question remains: *why* is closure voicing a lenition phenomenon in intervocalic stops? A gesture-based approach to

[19] At least as pertains to voicing; see Section 1.4 for a comparison of the transcription standards used here and those proposed by Schachtenhaufen (2022).

laryngeal representation can account for this. I propose that closure voicing in /b d g/ follows from the loss of the glottal spreading gesture that is usually found in these stops. When the vocal fold configuration is optimal for voicing and subglottal pressure is high, some amount of closure voicing is natural and requires no extra effort (Westbury and Keating 1986; see Section 4.2). This vocal fold configuration is required for producing vowels both before and after intervocalic stops, so maintaining it throughout the closure will require the least articulatory effort. In contexts where gestural undershoot is generally expected, it is unsurprising that we also see the loss of a non-distinctive glottal spreading gesture (as in /b d g/), and to a much lesser extent, the loss of a distinctive glottal spreading gesture (as in /p t k/).

This type of lenition is not predicted from either of the featural representational accounts discussed above. If /b d g/ are abstractly specified as [voice], we would not expect lenition to be required for phonetic voicing; in fact, Kingston and Diehl (1994) explicitly use the presence of intervocalic voicing as an argument for why Danish has contrastive [voice]. If /b d g/ receive some value for [spread glottis] late in the phonological derivation, there is no explicit mechanism for reducing this number in environments prone to lenition. However, intervocalic voicing as a consequence of lenition follows directly from the established facts about glottal activity, and as such, can also follow from a representation relying on gestural scores. Recall from Section 4.3 that only a gesture-based account of underlying representation leads to specific predictions about lenition, namely that lenition would cause reduction in the timing and magnitude of associated glottal gestures. This is in line with the results presented here. The difference in lenition rates across the two laryngeal series follows directly from the difference in magnitude of the underlying gestures. This account also correctly predicts that voicing-as-lenition is only found intervocalically; the loss of a glottal opening gesture would not result in voicing in initial position, where voicing requires effort.

In Table 4.6, I summarize the predictions following from different theoretical approaches, and whether or not support for these predictions was found in the current study.

Table 4.6: *Summary of predictions from different theoretical approaches. + indicates that predictions are in agreement with the findings of this chapter.*

Approach	**Danish /b d g/** *Prediction*		**Danish /p t k/** *Prediction*		**Lenition** *Prediction*	
Concrete [voice]	Variable voicing	+	Negligible voicing	+	No predictions	
Abstract [voice]	All outcomes possible		All outcomes possible		No predictions	
Gestures	Limited voicing	+	Negligible voicing	+	Voicing in both series	+
					More voicing in /b d g/	+

4.8.3 The relative predictive power of variables

LARYNGEAL CATEGORY is a very strong predictor of voicing, as are a number of variables associated with lenition. Particularly strong lenition variables are LOCAL SPEECH RATE, PRECEDING STRESS, and affix boundaries, but overall, the majority of lenition variables have a significant influence on voicing in the expected direction. It is interesting that affix-initial stops have a particularly high likelihood of voicing. In Section 4.4.1.3, I hinted that this may have an exemplar theoretic explanation: affixes are so often encountered with closure voicing that it has seeped into the underlying representations at the morpheme-level in a way that is not predictable at the phoneme-level. This is obviously controversial, in large part because it is impossible to represent in many modular approaches to grammar (where phonetic information is invisible to morphology), and it is a very different conception of phonological representation than those discussed above. This remains an interesting problem for further research.

Many of the other variables that were expected to influence closure voicing are aerodynamic in nature, and these generally do not have an observable effect on closure voicing in the data. This may be either because these variables truly do not influence closure voicing, or because the influence of these variables is more gradient in nature. It is possible that aerodynamic variables affect the relative proportion of closure voicing within those closures that I simply categorize as 'not fully voiced'. This can only be tested in a study with more fine-grained coding of voicing.

I had a number of predictions for how the tongue position before and after the occlusion would affect the rate of closure voicing, which can be mostly be summarized as follows: a narrower constriction in the oral cavity before and after the occlusion was expected to decrease the chances of closure voicing, because such sounds are sometimes taken to be less sonorous (e.g. Parker 2002), and voicing follows more naturally from sounds with higher sonority; in fact, Chomsky and Halle (1968) define their distinctive feature [sonorant] exclusively with reference to whether voicing follows naturally from the vocal tract configuration. However, none of these predictions hold up; no effect of tongue body position was found except for the point of occlusion itself.

Place of articulation has a strong effect on voicing, and this has an aerodynamic explanation. The supralaryngeal cavity is relatively small during a velar occlusion, which provides little opportunity for passive expansion, and as such, velar stops are voiced at a lower rate. Alveolar and bilabial occlusions are more amenable to voicing, and the difference in size between the resulting cavities is negligible, which may be why they do not differ significantly in their amenability to voicing.

The influence of stød on the potential for closure voicing can also be thought of as an aerodynamic effect. The naturalness of intervocalic closure voicing crucially depends on high subglottal pressure at the time of occlusion and on the vocal fold configuration being amenable to voicing. Closure voicing following stød is very rare – this was a strong effect in spite of the total number of relevant tokens being quite small – presumably because laryngeal contraction in the production of stød causes a vocal fold configuration that is less amenable to voicing than that of modally voiced vowels. Tautosyllabic stød was found to increase the chances of voicing, which is surprising, given that stød has many of the same syllable-initial cues as stress. However, another initial articulatory correlate of stød reported by Fischer-Jørgensen (1987, 1989) is increased subglottal pressure, which in itself increases the likelihood of voicing.[20] This may serve to explain why tautosyllabic stød empirically shows a negative correlation with voicing (see Table 4.3), but correlates positively with voicing in a model that also controls for stress (see Table 4.5).

4.9 Conclusion

In this chapter, I have investigated intervocalic stop voicing in a corpus of spontaneous Danish speech. Although Danish stops are generally well-described, most of what has previously been written about voicing has been speculative. I have shown that intervocalic voicing is very rare in /p t k/ and occurs at lower than chance rate in /b d g/. In modeling the data, I controlled for a number of aerodynamically motivated predictors, most of which appear to have little influence on the occurrence of closure voicing. However, closure voicing was found at relatively high rates in environments where lenition is expected, i.e. quick speech, unstressed syllables, before neutral vowels, and in morphological affixes. This supports an analysis of intervocalic voicing as a lenition phenomenon. These findings can be accounted for with reference to previous articulatory studies showing that both laryngeal series of Danish stops are produced with glottal spreading gestures that counteract voicing, although these gestures differ in timing, magnitude, and functional load. Intervocalic voicing can be modeled as the loss of this gesture. The gesture is lost at a higher rate in /b d g/, where it is shorter, of smaller magnitude, and does not serve a critical distinctive function. There is a very extensive literature on the representation of laryngeal contrast, and I have necessarily discussed only a few perspectives here. If intervocalic voicing is indeed a lenition phenomenon, I suggest that this is best represented in a phonological representational framework which can directly incorporate the timing and magnitude of articulatory gestures, such as Articulatory Phonology.

[20] Although bear in mind that subglottal pressure was measured for only one participant, and no words with initial oral stops were measured, so this explanation must be taken with a grain of salt.

Few corpus studies of intervocalic voicing are available, and as such, it is difficult to compare these results to other 'aspiration languages' (or 'true voice languages' for that matter). This means that more studies are necessary, detailing how different variables influence the probabilistic occurrence of closure voicing in stops in other languages. This will help determine which effects should be associated with phonetic implementation only, and which should be considered grammatically encoded.