Phonetic and dialectal variation in phonologically contrastive laryngealisation:
a case study of the Danish stød

Andrea Siem

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Linguistics

Department of Linguistics and English Language

Lancaster University

April 2023

Declaration

I declare that this thesis is my own work, and it has not been submitted in substantially the same form for the award of a higher degree elsewhere.

Andrea Siem

14th April 2023

Phonetic and dialectal variation in phonologically contrastive laryngealisation:

a case study of the Danish stød

Andrea Siem

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy in Linguistics

Department of Linguistics and English Language

Lancaster University

April 2023

Abstract

A growing body of research is documenting the variability of the phonetic manifestation of phonological contrasts in voice quality. This thesis is an empirical addition to the typology of cross-linguistic phonetic variation in the production of phonological laryngealisation. Investigating this variation is done through a comparative lens of dialectal differences within a language and with the Laryngeal Articulator Model (Esling et al 2019) as a guiding framework for investigating laryngealisation. Dialectal differences are analysed using the Danish stød as a case study, comparing speakers from Copenhagen and Aarhus. The stød has previously been reported to be highly variable in its production, making it an excellent source of phonetic variation. Further, the dialect of Aarhus has been speculated to contain a tonal stød variant under certain conditions (Kyst 2004) which has not yet been studied phonetically, making the inclusion of this dialect novel both in a typological perspective and in the context of the phonetics of the stød specifically. Data from 10 speakers of Modern Standard Copenhagen and 11 speakers from Aarhus is analysed using both acoustic and articulatory measurements. Acoustic measures include fundamental frequency, intensity, the amplitude difference between the first and second harmonic (H1-H2), Cepstral Peak Prominence, Harmonics-to-Noise Ratio and Subharmonic-to-Harmonic Ratio. The articulatory measures are obtained via Electroglottography, a novel method in the context of investigating the stød, which enables the vocal fold contact patterns to be modelled for analysis. Three main

research questions are explored: (i) how much gradient phonetic variation in voice quality occurs during the Danish stød (ii) which acoustic and articulatory measurements correlate with the subtypes of the stød and (iii) how the voice quality changes are timed. The different phonetic types of laryngealisation are categorised in reference to five subtypes of creaky voice described in Keating et al (2015). To accommodate the expected variation in the Aarhus dialect, the stød in divided into two types for analysis, regular and tonal, both elicited in a contrastive minimal pair. The differences are analysed dynamically using two different statistical methods, Generalised Additive Mixed Models and random forest models. The findings generally confirm that investigating dialectal differences when exploring types of contrastive non-modal phonation is a rich resource to draw upon in widening our empirical understanding of phonetic variation in phonological voice quality across different languages. They demonstrate that stød in Danish is not just one type of stød, and even the standard Copenhagen variant exhibits differences in acoustics and articulation based on its stød basis. The study of timing finds high variability but uncovers some general patterns according to stød type and dialects, solidifying timing as a rich source of phonetic differences. Lastly, the findings from this study support the notion that the larynx is an active articulator and that various structures within it can affect voice quality independent of activity in the glottis. This encourages more research into how these different laryngeal structures interplay and in what ways this affects commonly used acoustic and articulatory correlates of these mechanisms.

Acknowledgements

First I would like to thank the Faculty of Arts and Social Sciences at Lancaster University for generously awarding me with a financial scholarship allowing me to conduct my studies.

I would then like to profusely thank my supervisor Sam Kirkham for encouraging me to apply for the PhD and for his support and guidance during the process of both starting and completing it, being extremely helpful with coding advice and feedback on drafts and ideas along the way. Thank you so much for your time, expertise and patience Sam!

I would like to thank my 21 participants for generously offering their time to allow me to conduct this research – without them there would no data to analyse and I am very grateful for their contribution. A huge thanks also goes to Míša Hejná and Nicolai Pharao for letting be borrow the Phonetics labs at Aarhus and Copenhagen University so I could obtain high quality recordings.

I would like thank some other people at Lancaster University who have provided both wonderful academic and non-academic support in the process of writing the PhD, namely Emily Gorman, Claire Nance, Raffaella Bottini and Alexandra Krendel. I would also like to thank the lovely people in the Phonetics Lab research group which has been inspiring to be part of and learn from.

I would like to thank my viva voce examiners for their very generous and helpful feedback that improved the quality of this thesis immensely.

I would like to thank my amazing partner Michael who has been the unwavering support system I would never ask for but definitely needed, especially in the last critical stretch of the writing up phase under extreme time pressure while also lecturing full time.

I would also like to thank my family for their support on this journey and for opening up their homes for me in Denmark during lockdown, waiting for the labs to re-open during the pandemic so I could collect the data. Thank you to Mariane Siem, Thomas Brink Siem and Brian Brink Plagborg, and a special thanks to my wonderful sister Josefine Brink Siem who was also going through the crazy process of writing a PhD and could relate to all the ups and downs better than anyone else. Having someone in the know was absolutely golden.

A special thanks also goes to the wonderful communities at CrossFit Central Lancaster and Kaizen Academy for keeping me sane and letting me connect with my body during this

extreme cerebral feat. It has been invaluable to have places like this to go to for a fun and challenging recharge.

Table of contents

List of tables

List of figures

Phonetic and dialectal variation in phonologically contrastive laryngealisation: a case study of the Danish stød

1. Introduction

This thesis is an investigation of phonetic variation in contrastive non-modal voice quality. The purpose is to provide an empirical addition to the developing typology of the ways in which such variation can pattern cross-linguistically and to provide a novel intra-language perspective to the field. The thesis reports on an acoustic and articulatory study of dialectal differences in laryngealisation during the production of the Danish stød. Using the stød as a case study was motivated by previous studies suggesting high variability in its phonetic manifestation (Riber Petersen 1973, Fischer-Jørgensen 1989, Hansen 2015) and the observation that there are dialectal differences not yet systematically studied (Kyst 2004) that would add a rich perspective on such variation. Data from 10 speakers of Modern Standard Copenhagen and 11 speakers of the dialect of Aarhus is analysed using acoustic and articulatory measurements to explore the phonetic variation with a focus on characterising the range in the respective dialects. Acoustic measures include fundamental frequency, intensity, the amplitude difference between the first and second harmonic (H1-H2), Cepstral Peak Prominence, Harmonics-to-Noise Ratio and Subharmonic-to-Harmonic Ratio. The articulatory patterns of laryngealisation are obtained via Electroglottography which enables the modelling of vocal fold contact patterns. The analysis is centred around three main research questions:

A) How much gradient phonetic variation in voice quality occurs during the Danish stød, a binary phonological contrast?

B) Which acoustic and articulatory measurements correlate with the stød?

C) How are the voice quality changes timed?

To characterise the variation during the stød more systematically, the stød is divided into two types based on their *stød basis* which are comparatively analysed using the description of five subtypes of laryngealisation by Keating et al (2015) and informed by the underlying framework of the Laryngeal Articulator Model (Esling et al 2019). Exploring these research questions allow for the adoption of both topic-based and methodology-based perspectives. Topic-based perspectives include the nature of phonetic variation during laryngealisation and the range of dialectal differences within this variation, adding a valuable perspective to the cross-linguistic typology. The methodology-based perspective lies in exploring questions such as the relationship between acoustic measurements and articulation more broadly, how laryngealisation can best be captured and which statistical methods are appropriate to analyse dynamic differences in articulation.

The approach taken in this study is a novel research contribution in at least two ways. First, it is uncommon to find studies on dialectal variation *within* languages with voice quality contrasts, making this study highly original in including such a comparative dialectal perspective. Second, the method used to obtain articulatory measurements, Electroglottography, is a novel method within the phonetic research on the stød specifically, making the study methodologically original too. Additionally, the stød is divided into types in the analysis according to the stød basis which is a novel approach to studying the stød and accommodates the inclusion of other dialects differing from Modern Standard Copenhagen, the only dialect previously studied phonetically.

The thesis is structured as follows:

Chapter 2 first introduces some background information for the motivation behind the study to situate it within the existing body of literature. This is done via a description of the general physiology that allows for phonatory changes relevant to contrastive use, then a description of how non-modal phonation can be characterised with a focus on laryngealisation, followed by a cross-linguistic overview of how voice quality changes generally pattern in languages where non-modal phonation is contrastive. The stød is then described to introduce the case study and its role in Danish phonology, then giving a detailed account for previous studies on its phonetic variation. Chapter 2 ends with some research questions that the thesis explores based on the literature review. Chapter 3 outlines the methodology, including a description of the acoustic and articulatory methods chosen and a detailed description of the main statistical method used to dynamically compare the measurements in each dialect, Generalised Additive Mixed Modelling. Chapter 4 provides the acoustic and articulatory results from the analysis of research question 1. Chapter 5 presents an analysis aimed to answer research question 2. Chapter 6 is concerned with research question 3 and presents the results from an analysis of the timing of non-modal phonation during the stød. Chapter 7 discusses the results in dedicated subsections, the first four of which discuss the findings of the analyses in respect to different aspects of the methodology used. The findings are then discussed in light of previous studies on the stød followed by a subsection offering some new perspectives on the timing of non-modal phonation. The results are then compared to cross-linguistic patterns in contrastive non-modal phonation before the discussion is concluded with some perspectives on future work made relevant by the current research findings and suggestions for directions to follow from here. Lastly, Chapter 8 summarises the findings of the thesis in a conclusion.

2. Background

2.1. Vocal fold movement and physiology

As the current project is concerned with variation in phonation and settings of the larynx to create meaningful contrasts in spoken language, this section is dedicated to initially provide a brief overview of the articulatory mechanics behind their execution. This will lay the foundation for understanding phonological uses of voice quality and their relevant acoustic correlates described in later sections. The vocal folds are housed within the larynx where they, together with their surrounding structures, form the basis of all phonation and interact to create what we perceive as voice. The immediately surrounding structures are made up of cartilage which in turn connects to bone – however, this section only deals with the soft cartilaginous structures as these will suffice to investigate variation in phonation types for this project – for an overview of the hard structures of the larynx, see e.g. Gick et al (2013: 72-74).

The biomechanical aspects of the vocal folds have been discovered using methods such as tensile testing where controlled tension is applied to the vocal fold tissue to measure the response or via rheometry which was developed to measure how soft matters, such as the vocal folds, respond to applied forces, especially when the matter in question has more than one value of viscosity and thus requires more measurement parameters. As we will shortly see below, this is true for the vocal folds as they have more than one layer. However, because they are very small, around 11-15 mm in women and 17-21 mm in men and have to be measured ex vivo to be mountable to the measuring equipment, these methods potentially yield some inaccuracies (Zhang 2016). An alternative method is nanoindentation which can be used for in vivo testing – however, this method only provides high accuracy when using very small indentation depths as data interpretation otherwise becomes difficult and does not allow for proper assessment of the non-linear mechanics of the vocal folds (Zhang 2016). The mechanical properties of the vocal folds constitute a large part of creating different voice qualities but as will be elaborated below, it is not only the vocal folds in isolation that are responsible for variations in voice. Their surrounding structures are also active and these have been visualised via laryngoscopy, i.e. direct video recordings of the larynx and a great deal of our knowledge of the larynx comes from these recordings. Having outlined the methods used to discover the laryngeal anatomy, these structures are briefly described below.

2.1.1. Laryngeal structures

The vocal folds stretch across the larynx where they attach anteriorly to the thyroid cartilage and posteriorly to the arytenoid cartilages. The arytenoids are primarily responsible for ab- and adduction of the folds via activation of their attached muscles, particularly for posterior closing of the membranous part of the glottis (Selbie et al 1998; Hunter et al 2004; Yin & Zhang 2014). To achieve a complete closure, however, the thyroarytenoid muscle, which constitutes the majority of the folds' mass together with the vocalis muscle, also needs to be activated to ensure anterior closing and medial compression. The thyroid and arytenoid cartilages sit on top of the cricoid cartilage and can be manipulated through the cricoarytenoid and cricothyroid joint. These cartilages make it possible to adjust the position, geometry and mechanical properties of the vocal folds (Zhang 2016). The vocal folds themselves are further divided into layers, a body and a cover layer. This two-structured division was first proposed by Hirano (1974) who recognised that this was necessary to account for the vast array of different voice qualities produced by the vocal folds in vibration. The body layer consists primarily of the upper part of the thyroarytenoid muscle, called the vocalis muscle. The second part of the body layer is the elastic conus covering the vocalis muscle, which is particularly thick near the edge of the folds and this thick part is called the vocal ligament. The vocalis muscle and the elastic conus are tightly connected and gives the impression of moving as one during vibration. The second layer, the cover layer, consists of a thin layer of epithelium and the lamina propria, a loose connective tissue – together these make up the mucous membrane of the vocal folds. Importantly, the body and cover layer are only loosely connected and thus can be moved in different ways during vocal fold vibration (Hirano 1974). Later, however, Hirano (1988) updated the division to include a third layer named the transition zone where the vocal ligament was an independent layer rather than belonging to the body because the three layers were found to have distinct mechanical properties – the more superficial the layer, the more pliability.

2.1.2. Vocal fold vibration

To initiate vibration, and thus phonation, air needs to flow via lung contraction and the vocal folds need to be adducted, a process that narrows or closes the glottis, below which the air pressure from the lungs builds up and when it reaches a certain pressure level the vocal folds can self-sustain vibration to modulate the glottal airflow (Zhang 2016). Adducting the vocal folds is primarily achieved via the lateral cricoarytenoid muscles which pivot the arytenoid cartilages when they contract to move the folds together via an inward turning of the vocal processes (Gick et al 2013: 75). Sound is produced by a three-way sound source mechanism where first vocal fold vibration displaces volumes of air, then a fluctuating air force is applied to the airflow by the vocal folds and finally, turbulence is developed downstream of the glottal exit (McGowan 1988; Hofmans 1998; Zhao et al 2002; Zhang et al 2002a). The second source, i.e. the fluctuating air force to the glottal flow, is considered the most dominant and is what creates the harmonic component of the voice output. A fourth source may be created if the ventricular folds are adducted tightly and a jet of glottal air impinges onto them (Zhang et al 2002b). However, the ventricular folds contain no muscles and thus cannot be directly controlled but in some speakers, they may vibrate along with the true vocal folds to create a rough quality to the voice (Gick et al 2013: 75). Once phonation has been initiated and established, the vocal folds vibrate in a cyclical pattern and each cycle has an opening phase where the folds abduct, a closing phase where they adduct and a closed phase where the glottis is completely closed and glottal airflow is zero (Zhang 2016). Studies suggest, however, that women tend to have a small posterior glottal opening even during the closed phase and thus do not reach zero airflow and only an incomplete closure each cycle (e.g. Awan et al 2015). An increased breathiness baseline in females has also been confirmed for speakers of Danish specifically (Hejná et al 2021), which is of course relevant for the current case study.

2.1.3. Further ways to impact voice quality – the Laryngeal Articulator Model

Voice has largely been thought of as a purely glottal phenomenon where sound is produced at the glottis, i.e. via the vocal folds, and the sound is then moderated by articulators in the oral cavity to create speech, rendering vibration a more passive phenomenon and the oral modulation the active mechanism for speech. However, a more recent proposal has been that the whole larynx is in fact active. Esling, Moisik, Benner and Crevier-Buchman (2019) have developed the so-called Laryngeal Articulator Model (LAM) which proposes that the larynx is not just a source of vocal fold vibration but an active articulator in itself, utilising its many structures to manipulate the voice source output (Esling 2005). This model will be the guiding framework for investigating possible articulations of laryngealisation in this thesis and will be described in more detail here. Note, however, that the thesis will not attempt to validate the model as such but rather use it as a framework for describing possible patterns in the data and how they can be thought of when considering the larynx as an active articulator in addition to just patterns in vocal fold vibration. The major relevant structures are pictured below:

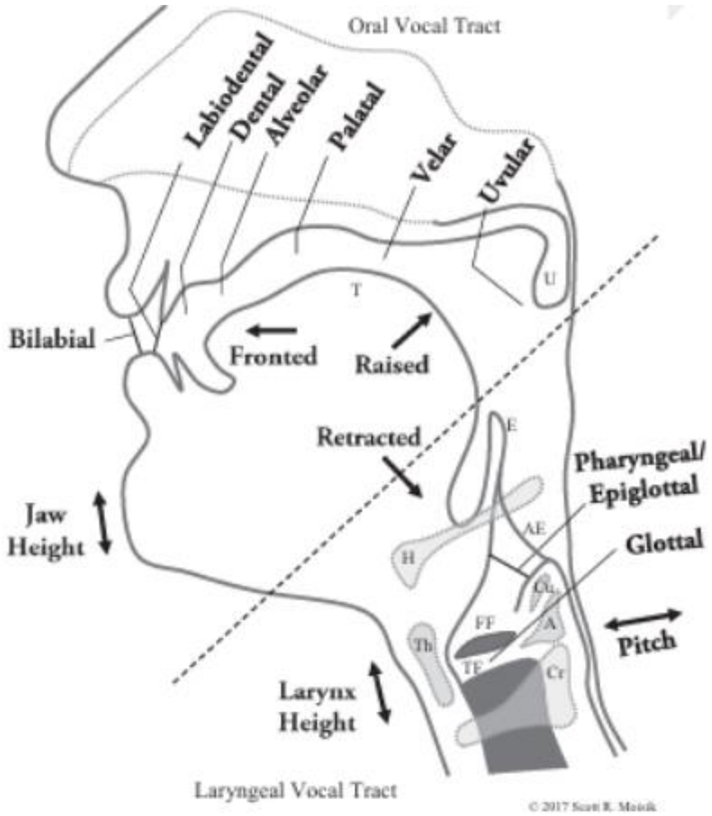

Figure 1. The Laryngeal Articulator Model, or the two-part vocal tract taken from Esling et al (2019: 6). Abbreviations include: T = tongue; U = uvula; E = epiglottis; H = hyoid bone; AE = aryepiglottic folds; Cu = cuneiform cartilage; A = arytenoid cartilage; Th = thyroid cartilage; FF = ventricular (false) folds; TF = vocal (true) folds; Cr = cricoid cartilage.

The multitude of vibrations the laryngeal articulator can generate is a function of its pattern of constriction, related to articulatory shaping of the lowest resonating chambers in the vocal tract. These resonances are a function of the volume of the epilaryngeal tube, of the degree of tongue retraction into the pharynx induced by aryepiglottic fold tightening and consequent vertical compaction, of the height of the larynx (distance below the hyoid bone) during these events, and of changes in the dimensions of the piriform fossae. Normally, the larynx is raised during laryngeal constriction, compressing the epilaryngeal tube and shortening the vocal tract through the pharynx. With accompanying tongue retraction, the pharynx is also compressed vertically, and resulting pharyngeal volumes are smaller. These effects are produced deep in the vocal tract, earlier than other resonance or noise components generated in the upper (oral) vocal tract. As a result, the laryngeal constrictor mechanism exerts a profound effect on the production and perception of voice quality throughout the entire vocal tract. In the LAM framework it is proposed that not one, but three different sets of folds can produce periodic signals (Esling et al 2019: 7). These are the vocal folds, the ventricular folds and the aryepiglottic folds. The vocal folds are the most efficient for the task but the other types of folds can also be considered sources of periodic energy.

The Laryngeal Articulator Model thus proposes that analogous to the tongue as the active articulator in the upper vocal tract, the aryepiglottic constrictor mechanism is active in the lower vocal tract, challenging the view of the pharynx as an empty space controlled only by the muscles around its walls. The model suggests that all effects that may be characterised as pharyngeal, epiglottal or, importantly for the present project, laryngeal are produced by the aryepiglottic constrictor mechanism, supported by tongue retracting and larynx raising. The border between the epilaryngeal tube and the upper vocal tract is the locus of stricture and the active articulator is the aryepiglottic folds, whereas the passive articulator is the epiglottis that they can vibrate against – these two articulators thus make up the aryepiglottic constrictor mechanism (Esling et al 2019: 5). The range of possible resonances is dependent on the volume of the epilaryngeal tube, the degree of tongue retraction and the height of the larynx, determined as its distance below the hyoid bone. The volume of the epilaryngeal tube can be made smaller by laryngeal constriction which compresses the tube, raises the larynx and shortens the vocal folds, combined with vertical compression by tongue retraction which reduces the pharyngeal volume. This effect thus originates deep in the vocal tract and consequently effects the sound as it goes through the upper vocal tract and contributes in no small amount to the perception of voice quality (Esling et al 2019: 7).

Further, it is proposed that there are in fact three possible sources of periodic vibration within the laryngeal mechanism. One is of course the vocal folds which are the most efficient at producing voicing, but also included are the ventricular folds and the aryepiglottic folds (Edmondson & Esling 2006). The ventricular folds can influence the mechanical action of the vocal folds and the aryepiglottic folds can perform a sphincter-like stricture in the upper part of the epilaryngeal tube. As such, sound production is not just a glottal phenomenon but also potentially a pharyngeal and epiglottal event. This expands the idea of what a laryngeal constriction is and can be because the LAM includes the folding and/or buckling of the aryepiglottic constrictor mechanism which leads to non-modal phonation production, or what could be classified as laryngealisation. This can either be in the form of so-called harsh voice or in the form of creaky voice caused by the laryngeal constriction which shortens the vocal folds and lowers the fundamental frequency by increasing their mass per unit length (Esling et al 2019: 14).

Thus, in the LAM framework, the states of the larynx include the components, and their many shapes, of the laryngeal articulator beyond the glottis itself, rendering the energy source for voice a multi-faceted mechanism involving different postures within the aryepiglottic constrictor and not just a function of glottal shape (Esling et al 2019: 37).

Having outlined the main laryngeal structures responsible for voice production the attention in now turned to discussing variations in what can be considered as ‘normal’ voice and how these can be used to create contrast as done e.g. during the Danish stød.

2.2. Characterising non-modal phonation

The literature on voice and phonation is extremely diverse when it comes to both terminology and voice quality categories, making strict definitions a challenge and cause of terminological dispute. Phonation is often used interchangeably with voice quality, which perceptually may refer to aspects of the voice other than pitch and loudness. However, authors often define voice quality differently, likely due to the subjective nature of its perception. In the LAM framework, the ‘designation of laryngeal activity as articulation reconciles the narrow notion of voice quality as phonation or vibratory register with the broader description of voice quality as a set of configurational postures throughout the vocal tract’ (Esling et al 2019: 7). The framework defines voice quality broadly as the long term, habitually recurring characteristics of a person’s voice; long-term in the sense that it spans over a larger time domain than e.g. prosodic and segmental phenomena and this is linked to accent and sociophonetic characteristics of a given community. More narrowly, however, the framework defines voice quality as voicing produced in the glottis, i.e. phonation, coupled with the short-term effects caused by manipulating the laryngeal articulator mechanism (Esling et al (2019: 1-2). This narrow definition will be adopted and focused on in this research as quick variations in laryngeal settings are relevant for the phonological contrast and phonetic variation under study.

Acoustically, voice quality is used for the amplitude and shape of the harmonics and the components of noise in the voice and their temporal variation (Zhang 2016). The articulatory aspect of phonation manipulation involves an interplay between laryngeal conditions, e.g. vocal fold stiffness, and the degree of subglottal pressure (van den Berg & Tan 1959; Xuan & Zhang 2014; Berry et al 1994). For example, an irregular, i.e. non-modal, vibrational pattern has been observed in conditions where the subglottal pressure was kept as in regular phonation but the stiffness of the cover layer of the vocal folds was reduced (Berry et al

2001). As such, non-modal phonation is a variable and broad cover term for any phonation type diverging from what can be defined as modal (Gerratt & Kreiman 2001).

The idea of modal phonation originally stems from the field of singing where a modal vocal register includes the range of fundamental frequencies normally used by an individual when singing or speaking, thus excluding e.g. falsetto voice (Hollien 1974). The concept of modal phonation has later been extended to the field of phonetics and in articulatory-acoustic terms it is characterised as what is produced when air pressure and vocal fold settings are optimised for maximum vibration (Ladefoged & Maddieson 1996), involving periodic vibrations and a rich glottal spectrum, that is, one high in overtones. In modal phonation the vocal folds do not simply open and close – rather, they are kept relatively thick to allow for a more complex cycle where they open from the bottom up and from the back forward. The closing phase is initiated from the bottom up but tends to move from the middle outward. The more complex vibrational cycle due to relative vocal fold thickness creates a rich sound we associate with modal phonation and makes the sound source easy to shape by manipulating the upper vocal tract (Mathieson & Greene 2001). In line with this, Zhang (2016: 2623) argues that an important feature of modal phonation is complete closure of the membranous glottis during vocal fold vibration which is crucial for producing harmonics at higher frequencies and thus enriching the spectrum. In turn, incomplete closure is often present in pathological voices with a weak or breathy voice quality. As mentioned above, however, women tend to produce even modal phonation with a glottal gap (Awan et al 2015) and thus a more breathy quality baseline. Therefore, a more inclusive way of defining modal phonation could be to say it involves maximum closure of the glottis relative to one's contact area baseline in the closing phase, which may be complete or near-complete depending on biological sex, but possibly also with an amount of inter-sex, inter-speaker variation.

In broad contrast, non-modal phonation is anything diverging from this and usually involves some degree of perturbations in the periodicity of a signal and a spectral shift. The acoustic effects of such a shift will be elaborated in section 5 below. Perturbations during non-modal speech may be cyclic in nature (Gerratt & Kreiman 2001), e.g. in the case of diplophonia - this term will be covered in depth further below.

A suggestion by Berry (2001) has been to characterise modal and non-modal phonation through the patterns of entrainment of the vocal fold vibration frequencies. The concept is that each vocal fold corresponds to a so-called eigenmode, which in a linear system is associated with a fixed characteristic frequency, an eigenfrequency, which is identical to the resonance frequency. But because the vocal folds do not vibrate at a single, characteristic frequency during phonation, their vibration is not entirely a linear phenomenon. Thus, a shift of the folds' frequencies occurs to align their oscillation during phonation, called entrainment, and the pattern of this entrainment is what characterises the phonation type. For modal phonation, a 1:1 entrainment pattern occurs where the folds oscillate at aligned frequencies. For non-modal phonation, exemplified by Berry (2001) by vocal fry produced with glottal pulses of alternating amplitudes, entrainment would instead be either a 1:2 pattern where one eigenmode vibrates at a lower frequency and completes cyclic vibrations alternating with glottal pulses and the other vibrates at a higher frequency and completes cyclic vibrations with each glottal pulse, or a 2:2 entrainment pattern where both eigenmodes vibrate at a lower frequency. If the trains of glottal airflow are irregular, the eigenmodes may become disentrained or do not entrain at any point. Thus, in this framework modal phonation results from a 1:1 entrainment of the vocal fold eigenmodes and non-modal phonation results from all other entrainment patterns.

Berry's (2001) approach provides a great framework for understanding patterns but is fairly abstract in terms of actually measuring voice. Another, and more measurable, way to describe non-modal phonation patterns has been through the respective lengths of the open/close phases of the vocal folds during a phonatory cycle. A cyclic pattern where the so-called Open Quotient (e.g. Slifka 2006) or Contact Quotient (e.g. Awan et al 2015) is skewed so the quotient is no longer around 50/50 is indicative of non-modal phonation types. A creaky type of voice would usually have a higher Contact Quotient whereas a breathy voice would have a lower Contact Quotient. This latter approach is typically adopted in EGG studies of vocal fold activity and is thus very appropriate for the present study as EGG will be used as one of two research methods. It should be noted, however, that in the LAM framework, it is not only vocal fold contact patterns that are responsible for changing voice quality, but rather the entire laryngeal articulator.

2.3. Creaky voice/laryngealisation

Having defined modal voice quality, it is now possible describe any divergence from this standard. This project is concerned with researching variations in non-modal phonation patterns, specifically variation on the continuum of laryngealised phonation types, the most extreme of which can be labelled vocal fry at one end and light compression at the other. This continuum can also be labelled hyper-compressed voice qualities (e.g. used in Hansen 2015) as opposed to hypo-compressed which would include breathy phonation types. A very common phonation type on this continuum is creaky voice which is phonologically contrastive in some languages, e.g. Jalapa Mazatek (Ladefoged & Maddieson 1996: 317), while in others it has a non-contrastive sociolinguistic function, e.g. in American English (Yuasa 2010). It can also function to aid turn-taking in conversations by signalling turn completion, e.g. in Finnish (Ogden 2001).

The terminology in labelling voice qualities on the laryngealised continuum is fairly diverse and most common labels include called creak, vocal fry, glottal fry, glottalisation, and laryngealisation. The detail in labelling generally depends on how many types on the continuum are considered and what the research focus is. These labels are not always interchangeable from an articulatory and acoustic point of view, however. Garellek (2013: 5) writes that glottalisation in a strict sense only refers to complete vocal fold adduction and that this may accompany a sound as a secondary articulation. However, as the vocal folds prepare to adduct, laryngealisation effects can occur on the adjacent segment and the acoustic consequences of this pre-adduction is also sometimes labelled glottalisation, meaning the label is used beyond this strict definition. For the label creaky voice, this is usually defined by properties of a low and irregular f0 and a constricted glottis, meaning a long closing phase in each vibratory cycle and low glottal airflow (e.g. Keating et al 2015). This definition, however, does not cover all subtypes of creaky voice which shall be made clear in the sections just below. The solution to this terminological variation has been to use umbrella terms intended to cover many types of voice qualities on the continuum of hyper-compression, e.g. Blankenship (2002) using the term laryngealisation to refer to a variety of phonation types on the scale, or Davidson (2020) using the term creaky phonation to 'include the other labels that have been used to refer to the acoustic and articulatory properties of creakiness' (p. 2). Garellek (2013) opts for the umbrella term 'glottalisation' for all acoustic and articulatory effects of glottal stops and/or laryngealisation (p. 5), despite also noting the fact that laryngealisation and glottal stops can occur in the same syllable which he suggests to indicate that they are produced differently but it is not known in exactly what way (Garellek 2013: 11). As a consequence, the Danish stød is labelled 'suprasegmental glottalisation'. This choice seems heavily informed by the focus of the research being the production and perception of glottal stops specifically, not phonologically contrastive use of any type of laryngealisation on the hyper-functional continuum.

Because the framework for the thesis is informed by the LAM, the terminology that will be used here is shaped by the insights of this model. As mentioned previously, the more traditional view of phonation types assumed a source-filter model where voice qualities were attributed to the vocal fold shape, properties and aperture at the glottis only. The LAM instead views the entire larynx as an organ capable of producing numerous variations of noise, resonance and periodicity, aided by airstream creating vibrations not only in the vocal folds but also at various points within the epilaryngeal tube and above, resulting in complex phonation. Because of this, the appropriate terminology should be 'states of the larynx' rather than 'states of the glottis' (Esling et al 2019: 38). Creaky voice is in the LAM viewed not just a glottal quality but also a highly salient aryepiglottic constrictor quality, resulting from tensing of the epilaryngeal tube resembling an epiglottal stop more than a glottal stop. As such, the structures above the glottis are highly relevant for the vibratory patterns. For this reason, using the umbrella term 'glottalisation' seems inappropriate as it indicates a glottal focus that is not fully compatible with the more holistic view of the larynx adopted in the LAM. The authors do note that the model's increased possibilities of impacting phonation pose even more difficulties for the classification of phonation types and their terminology. What the model contributes to the discussion of terminology, however, is a *definition* of laryngeal constriction as a 'progressive folding or buckling of the aryepiglottic constrictor mechanism' (Esling et al 2019: 14) which in effect changes modal phonation to non-modal phonation, either in the form of so-called harsh voice or in the form of f0-lowering and creaky voice, or indeed both, and the authors argue that either of these effects could be labelled laryngealisation.

It should also be noted that the LAM classifications of subtypes on the hyper-compressed continuum of voice qualities are categorised primarily based on their articulation derived from direct imaging of the larynx. This provides wonderful detail about the entire laryngeal articulator but makes it more difficult to use these classifications in research such as the current where these structures cannot be directly visualised, meaning it is a challenge to decide exactly which states of the larynx are being utilised for certain acoustic outcomes. However, that the larynx *is* being utilised seems clear from the LAM, even in producing full glottal stops described in Esling et al (2019: 65) as they observe creakiness/laryngealisation preceding or following glottal stops/glottalisation caused by the actions of the aryepiglottic constrictor mechanism.

In terms of the case study phenomenon, the Danish stød, the common terminology labels tend to be laryngealisation but Grønnum (2005: 215) describes how in its most emphatic form, it can be realised as a full glottal closure. Elsewhere, however, she argues that the stød is not a glottal closure but a special form of creaky voice with irregular vocal fold vibrations (Grønnum 2007).

To summarise this debate, there is much variation in terminology and despite some attempts at strict definitions for different labels, these are muddled too. In the remainder of this thesis the umbrella term laryngealisation will be used to cover phonation types on the hyper-compressed spectrum due to (i) being informed by the LAM framework and (ii) the tradition for doing so in some stød literature. This is with the exception that when the thesis discusses some clearly defined subtypes of laryngealisation, e.g. those outlined in Keating et al (2015), the specific labels from the literature will be used for clarity. In phonetic terms, however, these are all considered laryngealisations.

As stated previously, it is only fairly recently that laryngeal phenomena have been thoroughly described to include more than just the glottis (e.g. Esling et al 2019) and as such, most research on laryngeealisations has focused solely on vocal fold activity and its acoustic correlates, excluding other potential structures such as the aryepiglottic constrictor mechanism. Further, laryngealisation is not just one thing and there are interesting deviations from the standard type(s). Some of these various types are characterised in the following sections as they provide a useful basis of categorising different articulatory and acoustic effects when describing phonetic variation in contrastive laryngealisation.

2.3.1. Prototypical creaky voice

The most commonly described laryngealisation, called prototypical creaky voice, is articulated with a long closing phase in each vibrational cycle of the vocal folds, often with the adduction of the folds being quicker compared to modal phonation (Hollien et al 1966; Whitehead et al 1984; Zemlin 1981; Javkin & Maddieson 1983; Kitzing et al 1982; Ladefoged et al 1988). Glottal airflow is low (Keating et al 2015) as well as subglottal pressure and there may be ventricular fold contact (Laver 1980). The longer closing phase is achieved by a shortening and slacking of the vocal folds, drawing the arytenoid cartilages together via contraction of the interarytenoid muscles - this allows the folds to be held together for longer, maximising their mass per unit length and enabling only a tiny bit of air to escape between them in each cycle (Gick et al 2013).

To describe the acoustic correlates of prototypical creaky voice a few basic acoustic terms must be introduced first. A complex waveform that is repetitive consists of tones, or harmonics, at certain frequencies, differing according to the quality of the sound produced. The lowest tone is the fundamental frequency, f0, which is the rate at which the waveform cycles are repeated per second (e.g. Ladefoged 1996: 37-38). Thus, if they are repeated 100 times per second, the f0 is 100 Herz. The higher tones of the wave are called harmonics and defined as any whole-number multiple of the f0. The f0 is the first harmonic, H1, and the tone at double the f0 is the second harmonic, H2, and so forth. The component frequencies of a particular sound wave will have different amplitudes so that sounds with the same f0 can still be qualitatively different because of the relative amplitude of the individual harmonics. Frequencies and amplitudes can be represented on a spectrum which, unlike the waveform, also include how the components are combined and not just which components are present, and this unique combination essentially makes up the quality of a sound (Ladefoged 1996: 42-43). We can now describe the prototypical acoustic correlates of creaky voice to include a low and often irregular f0 (Laver 1980; Keating et al 2015) and a more prominent H2 relative to H1 (Gick et al 2013). Esling et al (2019: 63-64) describe that production of creaky voice requires a contraction of the thyroarytenoid muscles to create short and thick vocal folds which results in the closing phase making up around 90 percent of each vibrational cycle. Recall that the vocal folds have more than one layer; thyroarytenoid contraction affects them differently. The body layer is stiffened but the cover layer is slackened (Deguchi et al 2011). Moisik et al (2015) showed that the ventricular folds can be coupled with the vocal folds during creaky voice to create the following effects: (i) increase of the vibratory mass which lowers f0, (ii) addition of damping which makes vibration more likely to cease, (iii), addition of more freedom to the system which makes irregular vibration possible and (iv) perturbation of the transmission of the mucosal wave resulting in wave energy being reflected back towards the midline of the vocal folds earlier in each cycle compared to modal phonation. Esling et al (2019: 65) further argue that the aryepiglottic constrictor is important for creaky voice as it functions to tense the supraglottal tube to a degree where the glottis is almost obscured beneath it, resembling an epiglottal stop.

2.3.2. Non-typical laryngealisation

As mentioned above there are also types of laryngealisations that deviate from the prototypical description in both articulatory and acoustic features across the literature. Redi & Shattuck-Hufnagel (2001) distinguish four different laryngealised types based on acoustic traits of which the last two can be said to be non-prototypical subtypes; one is a glottal ‘squeak’ with a sudden, sustained high f0 and the other is so-called diplophonia – a type which will be elaborated below. Glottal squeak has been studied qualitatively in Hejná et al (2016), defining it acoustically as a sudden shift to a high sustained f0, usually with low amplitude. Articulatorily it can be explained by the high stiffness in the thyroarytenoid muscle that occasionally may manifest in sudden high frequency vibration (Esling et al 2019: 64). It should be noted, however, that it is studied in the context of glottalisation rather than laryngealisation, the former defined as instances of either an aperiodic f0 or a sudden drop in f0 relative to the immediate context (Hejná et al 2016), whereas laryngealisation may not involve aperiodicity or f0 drop, depending on type, more on this below. Sex differences were also found for glottal squeaks, with the phenomenon occurring frequently in females but only in one male in the data studied. Further, inter-speaker variation was also high, meaning not all individuals squeak, irrespective of speaker sex.

Returning to other subtypes of laryngealisation, Keating et al (2015) also distinguish four subtypes, namely: (i) vocal fry, where the glottis is constricted and f0 is low but the signal is periodic rather than irregular. The pulses are subject to high dampening which makes them individually distinct and audible, which in turn enhances the low f0 property of the signal, (ii) multiply pulsed voice, where f0 is irregular but not necessarily low, and characterised by alternating longer and shorter pulses which generally have a long closing phase in each cycle. If the pulse is doubled, there are two simultaneous periodicities, one fairly low and one about an octave higher, (iii) aperiodic voice, where f0 is so irregular that periodicity is lost and there is no perceived pitch; the irregularity of the signal is enhanced and 'noisy', and, finally, (iv) non-constricted creak, where f0 is low and irregular but the glottis is spreading and airflow is high. This type is described in detail by Slifka (2006) who reported on laryngealisation in utterance-final position where the vocal folds part before the utterance is complete, making sustained phonation difficult and thus resulting in open phased creak. Besides these types of laryngealisation there are some non-modal phonation types that are similar yet distinguishable on one or more parameters, e.g. so-called tense voice which refers to an articulation with a constricted glottis but without a low or irregular f0. It can function as creak phonologically in languages where this is contrastive and co-occurs with a high tone, e.g. in Mazatec (Keating et al 2015). Redi and Shattuck-Hufnagel (2001) demonstrate that creaky types vary both across speakers and across utterance positions. Laryngealised subtypes have further been shown to be not only acoustically distinct but also perceptually distinct (Gerratt & Kreiman 2001), making them meaningful for communication beyond pure phonological contrast and thus carry relevant sociophonetic information as well.

2.3.3. Diplophonia

As mentioned above, non-prototypical laryngealisation may under certain conditions occur as diplophonia. Diplophonia has been given most attention as a symptom of underlying voice pathology (e.g. Aichinger et al 2014) but it may also be used voluntarily to create contrast in voice quality. Diplophonia is defined by Redi and Shattuck-Hufnagel (2001) as any kind of alternation in pulse frequency, amplitude or waveform shape. This definition is fairly vague and could arguably refer to a number of phonatory phenomena. A more perceptual definition is given by Gick et al (2013: 112) who describe diplophonia as a ‘minor condition that typically makes a person’s voice sound deeper than their fundamental frequency’. Dejonckere & Lebacq (1983) give the somewhat simpler perceptual description that diplophonia is two distinct simultaneous tones. The authors studied the phenomenon in the context of vocal fold pathology in 74 subjects and identified it in relation to vocal fold lesions or functional voice disorders. In these conditions it commonly occurs as a transient bitonality, either at the onset or end of vocalisation. Using EGG Dejonckere & Lebacq (1983) showed that three different diplophonic patterns could be distinguished; (i) abnormal waves where the vocal folds joined in two phases during the closing, (ii) two separate alternating waves or, (iii) repeated groups of waves. Further, the closing phase is mostly small or not well defined. They conclude that diplophonia is a glottal phenomenon that result from either a double- or multi-phased closing or opening movement of the vocal folds or from regular sequences of unequal cycles. They reiterate that even though each vocal fold is considered an oscillator their respective oscillations are not independent, although they may be out of phase. This means that diplophonia, despite being perceived as two tones, never has two separate harmonic spectra based on two distinct f0s. Instead, the acoustic characteristic of diplophonia was found to be an addition of a second subharmonic series to the normal f0 structure, often at ½ of the f0. This was rarely permanent, however, but occurred transiently in most subjects. Further, the diplophonic spectrum was in most cases noisy which perceptually is associated with vocal hoarseness. Articulatory mechanisms leading to diplophonia can be a weakening in the coupling between the vocal folds due to reaching a sufficient level of separation allowing the folds to vibrate at different frequencies (Zhang 2016). As mentioned in section 2.3.2. above, Keating et al (2015) describe what they call multiple pulsed voice as having alternating longer and shorter pulses with a long closing phase in each cycle and with a pulse that may be doubled in which case there are two simultaneous periodicities, one low and one about an octave higher. This sounds very similar to the description of diplophonia, although the subharmonic f0 is described as an octave higher rather than half of the f0 and the closing phase is long rather than typically weak or non-existent as in the diplophonic subjects measured by Dejonckere & Lebacq (1983). Whether these should be regarded as separate phenomena or subvariants of the same phenomenon is not clear from the literature. For the purpose of analysis the main focus will be to establish the presence of subharmonics using the acoustic measure of Subharmonic-to-Harmonic Ratio (SHR).

2.3.4. Acoustic correlates of laryngealisation

Having outlined the main characteristics of laryngealisation, it is time to consider in more detail how phases of vocal fold movement correlate with spectral and waveform measurements. Generally, as mentioned above, a long open phase should yield relatively more energy at the first harmonic (H1) in the lower-frequency spectrum. If the open phase is short, the second harmonic (H2) may become dominant and thus the strength relationship between H1 and higher harmonics is a useful way to quantify phonation types acoustically (Zhang 2016). Assuming that each acoustic measure reflects a particular aspect of laryngealisation and that each type will have specific combinations of these aspects, Keating et al (2015) argue that the different kinds of laryngealisations are acoustically distinct from modal voice and outline some prototypical acoustic properties of different subtypes of laryngealisation, all assigned a main property and a main correlate, summarised below:

- Prototypical creaky voice will have a low f0, irregular vibrations correlating with high noise and glottal constriction correlating with a low H1-H2.

- Vocal fry which will have a low f0, glottal constriction correlating with a low H1-H2 and damped pulses, correlating with low noise and narrow bandwiths.

- Multiply pulsed voice which will have irregular vibrations correlating with high noise, glottal constriction correlating with a low H1-H2 and subharmonics correlating with a high Subharmonic-to-Harmonic Ratio (SHR).

- Aperiodic creak which will have no perceptible pitch and irregular vibrations correlating with high noise and glottal constriction correlating with a low H1-H2.

- Non-constricted creak which will have low and irregular f0 correlating with high noise but no indication of glottal constriction, i.e. non-low H1-H2.

These characterisations motivate three acoustic measures: f0, H1-H2 and SHR. Some additional measures are also relevant to include. One is the Harmonics to Noise Ratio (HNR) which, if low, will indicate low periodicity. Keating et al (2015) propose this measure as a correlate of an irregular f0 where glottal pulses are difficult to define in the signal. If periodicity is low and harmonics cannot reliably be measured, one further possibility is measuring the Cepstral Peak Prominence (CPP) which has been used to quantify dysphonia in aperiodic speech signals (Heman-Ackah et al 2003). The measure is calculated from a Fourier transformation of the voice spectrum and a low CPP indicates high aperiodicity. A low CPP also indicates a high proportion of noise in the signal which is relevant to distinguish certain subtypes of laryngealisation because although the HNR is also a measure indicating signal noise, the HNR will be high if the pulses are dampened (Keating et al 2015), even with the presence of noise. As such, CPP is a good acoustic candidate for distinguishing types characterised by high noise but dampened pulses.

2.3.5. Articulatory measures of laryngealisation

Having outlined some of the main acoustic correlates for laryngealisation, this section outlines the articulatory correlates of the phenomenon. Two parameters are relevant to characterise vocal fold movements in any phonation type; (i) the contact timing of each glottal cycle, i.e. the relative proportion of a contacting vs de-contacting phase and (ii) the amount of contact along the length of the vocal folds, i.e. if the contact phase involves a full glottal closure or only partial closure. These parameters can be observed most directly using video recordings of the vocal folds from above using a laryngoscope. However, this method is also highly invasive and requires a degree of medical training to be correctly executed. Instead, less invasive but more derivative methods can be used. Previous studies have used Electroglottography (EGG) which can indicate the vocal fold activity by determining the so-called Contact Quotient (CQ) from the first derivative of the signal (the dEEG) by identifying the landmark timepoints of the closing and opening instant of the vocal folds. However, for phonation types where the folds may not reach complete closure in any cycle the CQ is a simplified reduction of the actual vocal fold activity as the ad- and abduction of the folds are more accurately measured as an interval rather than a single timepoint (Herbst et al 2017). This is assumed to be the case for at least prototypical creaky voice given that its execution with largely adducted vocal folds but with a portion along their length open just enough to facilitate voicing (Gordon & Ladefoged 2001). For all subtypes, a dynamic interval approach is thought most appropriate to characterise their articulation. This means obtaining the EGG signal and calculating the CQ of the derivative for the entire stød basis and modelling it dynamically for statistical analysis.

EGG indirectly measures the contact area of the vocal folds via their impedance to an external electrical current passed between two electrodes placed on the front of the neck, bilateral to the thyroid prominence. The passing electrical current is low voltage to make it physiologically safe but high frequency to avoid large variations in the surface contact between the skin and the electrodes it passes between (Titze 1990). The method was first invented by Fabre (1957) and there is now a large body of research linking the shape of the

EGG signal to the vocal fold contact patterns by comparing it to various more direct visualisation methods such as stroboscopic photography (Fourcin 1974; Lecluse et al. 1975; Pedersen 1977), videostroboscopy (Anastaplo and Karnell 1988), high-speed cinematography (Baer et al. 1983a; Childers et al. 1990; Childers and Krishnamurthy 1985), photoglottography (Baer et al. 1983b; Berke et al. 1987; Dejonckere 1981; Kitzing 1983; Titze et al. 1984) as well as other non-visual measurement such as subglottal pressure (Kitzing et al. 1982) and inverse filtering (Rothenberg 1981; Rothenberg and Mahshie 1988).

A limitation of using EGG is that the exact glottal configuration is difficult to obtain as spatial discrimination of the laryngeal tissues is not highly distinct. This means that the signal records changes in the vocal fold contact area but does not reveal if these changes are anterior, posterior, high or low in the glottis. As such, it can mainly be used to measure degree of glottal closing via the correlation that more contact transfers more electricity and therefore closing will excite peaks in the signal (Gick et al 2013: 90). Further, it is not known to what extent other laryngeal mechanisms in the LAM framework affect the EGG signal. Nevertheless, EGG is arguably a useful measure for studying phonetic variation in contrative phonation as a method combined with acoustic measurements to add more detail to the phonatory picture in a way that is portable and non-invasive, original in the context of the stød and follows methodology from previous studies on vocal fold contact patterns.

2.4. Cross-linguistic studies on phonologically contrastive laryngealisation

Having outlined an articulatory and acoustic foundation on which to conceptualise and quantify laryngealisation this section provides an overview of the literature on some of the languages that utilise laryngealisation as a phonological contrast. The language under study in this thesis is Danish, and the literature conducted on the Danish stød is reviewed below in dedicated sections. Preceding these, however, this section provides an overview of the patterns of phonetic variation typically found within phonological laryngealisation in other languages as the sections thus far have identified characteristics of all laryngealisation, irrespective of whether its use is phonemic or phonetic. This section is primarily focused on phonation types on the hyper-compressed continuum, i.e. not breathy voice, and those which are suprasegmental. As such, segments like ejectives and implosives will not be considered since they, despite involving laryngeal activity, are not a suprasegmental property of a phoneme but rather phonemes in their own right.

The language-specific phonemic status of laryngealisation offers an interesting aspect of segmental timing of its realisation as phonological contrasts are assumed to be more segmentally restricted than non-contrastive, sociolinguistic uses of laryngealisation as the latter do not carry any exclusively semantic load in the sense of distinguishing lexical or grammatical entities. Sociophonetic uses of phonation do not suffer from these limitations and can span entire utterances. In support of the restrictive nature of phonological voice quality, Silverman (1995/1997) proposes that contrastive phonation and/or tone is sequenced within languages in a way that ensures the optimal perceptual recovery of these elements. The need for perceptual salience could manifest itself in a few ways such as placing the phonation contrast on the first portion of the vowel for early auditory detection for the listener, or such as having a transient portion of non-model phonation in between two stretches of modal phonation for maximum contrast. It might also be that non-modal phonation types are in themselves salient enough that they do not need articulatory timing as a perceptual enhancement and therefore can easily last the entire duration of a vowel. These options are explored in the following.

On the premise of investigating the presence of phonation-type contours analogous to tone contours, DiCanio (2009) investigated Takhian Thong Chong, or Chong, a language described as having phonological vowel contrasts in register. The so-called tense register is realised a laryngealisation but with a high f0 and the effect of register was found to be significant but not equally distributed throughout the duration of the vowel. Measured as Open Quotient values derived from Electroglottography DiCanio (2009) showed that in Chong, the first 50 percent of vowels with and without laryngealisation have similar values but that after this point they start to diverge to manifest the contrast in voice quality. Similarly, Esposito (2003) showed that in Santa Ana del Valle Zapotec, laryngealisation is realised towards the end of the vowel rather than in the onset. Avelino (2010) describes laryngealisation patterns in Yucatec Maya which exhibit three variations: (i) laryngealisation articulated as a full glottal stop separating two surrounding stretches of speech with modal phonation, (ii) vowels with modal phonation with a transient stretch of laryngealisation in the middle of the vowel and (iii) vowels starting with modal phonation and increasingly receiving laryngealisation towards the end of the vowel. The language White Hmong has a dual contrast with tone and phonation, one of which is low tone with creak contrasting with a breathy, falling tone. For the tone with creak it was found that the acoustic correlates of creakiness were present throughout the entire vowel (Garellek 2012). Of course not only vowels can have contrastive laryngealisation. Research on some Athabaskan languages found laryngealisation on so-called glottalised sonorants leading to a portion of creaky voice which patterned in two different ways: either as creaky pre-glottalisation produced simultaneously with the consonant onset but lasting less than the full consonant, or as post-glottalisation where the creak starts approximately in the middle of the consonant and lasts well into the following sound (Hargus 2016; Howe & Pulleybank 2001; Ladefoged & Maddieson 1996; Silverman 1995/1997).

In regard to the timing of laryngealisation during the Danish stød, it has been noted that the stød, when occurring on a long vowel, may extend into the following sonorant consonant and that, if occurring on a sonorant consonant, may be initiated on the preceding vowel and extend into the succeeding voiced sound (Grønnum & Basbøll 2001). It may also, despite being characterised as a syllabic prosody, extend into the following syllable (Grønnum & Basbøll 2007).

Relevant to the stød is also the relationship between laryngealisation and tone, particularly in this research project as the inclusion of the Aarhusian dialect hypothesises a tonal realisation of a laryngeal contrast.

Southeast Asian languages spoken in Vietnam, Myanmar, Thailand and Cambodia all use contrastive laryngeal properties that can be labelled tonation (Bradley 1982b). The most common form of tonation is contrastive changes in fundamental frequency used to distinguish either lexical or grammatical entities. In this context, tone can be defined as 'the presence of lexically significant, contrastive, but relative pitch on each syllable' (Pike 1948: 3).

While deliberate changes in f0 are the most commonly described tool used for tonal contrasts, the tonation systems of Southeast Asian languages are more complex and an important part of this complexity is the manipulation of laryngeal settings to change properties of the vocal folds and/or the height of the larynx to create changes in phonation type. Brunelle & Kirby (2016: 193) list the most common phonation types in these languages as creaky and breathy voice, but with other types used too. The languages of Southeast Asia combine f0 and phonation types to realise tonal contrasts in many different ways that are suggested to lie on a continuum of 'pure' tone on one end, i.e. changes in f0 signals the contrast only, and *register* on the other end, an umbrella term covering the use of f0, voice quality, vowel quality and duration to distinguish categories phonemically (Brunelle & Kirby 2016: 193). Examples of the former is Central Thai, a Tai-Kadai language with five lexically contrastive tones distinguished only by f0 height and contour (Edmonson 2010). An example of the latter is the Austronesian language spoken in Vietnam called Eastern Cham where f0 is combined with phonation and duration to signal phonemic contrasts, e.g. high vs low register words where the former have a high f0, modal voice quality, a low vowel and a short syllable rhyme, whereas the latter have low f0, breathy voice, a high vowel and a longer syllable rhyme (Brunelle 2005). Because of this complex interaction between phonemic changes in tone and phonemic changes in voice quality, Brunelle & Kirby (2016) make the point that while categorical labels like 'tone languages' or 'register languages' can be descriptively useful, the complexity of the interaction between tone systems, register and phonation types means it can be questioned if the distinction is even useful, particularly for purposes of linguistic typology (p. 191-192). Indeed, they emphasise this by pointing out that a language like Burmese has been typologically categorised as both a tone language and a register language, depending on the source, and that it is difficult to establish what is the most salient perceptual cue to the contrast in languages that make use of multiple properties of voice to signal phonemes (Brunelle & Kirby 2016: 194).

Another language that uses a suprasegmental contrast distinguished by laryngealisation or tone is Scottish Gaelic which interestingly has dialectal variation in the realisation of this contrast (Morrison 2019). The two phonological classes, labelled Class 1 and Class 2, are distinguished by tone in Lewis but by what is labelled as glottalisation in Islay. There is a third phonetic realisation labelled overlength too. In tonal dialects the phonetic realisation of the phonemic contrast is realised via a pitch peak on the first mora in Class 1, contrasting with Class 2 where the second mora receives the pitch peak. In so-called glottalised dialects this is done via laryngealisation on Class 2 forms between moras. Overlength is produced by elongating the second mora of Class 2 forms. Morrison (2019: 392) argues that this contrast is a direct reflection of the extent of the stressed syllable and further, that the metrical structure is predictable from the underlying segmental content, leading to derived metrical contrast between Class 1 and Class 2 forms. This is comparable to the stød in the sense that the segmental structure influences the stød type but is not comparable to the stød for all scholars as a mora-based analysis of the stød has not been accepted by all Danish phonologists. This point is elaborated in the section below on the Danish stød specifically. An interesting consequence of the metric analysis, however, is the observation that the Class 1 vs Class 2 distinction is dependent on the length of the vowel – if it is short, the form is Class 1, if it is long it belongs to Class 2 (Morrison 2019: 395). The relationship between vowel length and type is also relevant for the stød. Before diving in to the phonetics and phonology of the stød, however, this section will carry on with some general perspectives from other languages.

The language Isthmus Zapotec described by Picket et al (2010) also has complex interactions between tone and laryngealisation. It is described as having five vowels with three phonation types in stressed syllables – modal, checked and laryngealised. The difference between checked and laryngealised vowels is that checked phonation sounds like vowels ending in a glottal stop, and might be preceded by laryngealisation. Laryngealised vowels are longer and produced with what is labelled creaky voice, or sometimes with rearticulation of the vowel after weak glottal stops after a low tone. Elsewhere, the distinction is described as laryngealised vowels sounding like there is a glottal stop in the middle of the vowel whereas checked vowels end in a glottal stop which may be audibly aspirated (Martin 2010: 1). These descriptions are also interesting for the perspective of timing of contrastive laryngealisation – clearly these two types are distinguished by one type occurring in the middle of the vowel and one occurring at the end of the vowel. Tones are also phonemic in Isthmus Zapotec and Picket et al (2010: 369) propose that the language has five tone melodies available for noun roots: High, Low, High-Low, Low-High, and Low-High-Low. The phonetic realisation depends on the syllable profile of each morpheme they occur in, coupled with grammatical and phonological factors in the environment. The tones can co-occur with the three phonation types in various combinations, but the tones are all restricted to nouns. Similarly, the phonation types are described as occurring on vowels but with one type only on any given morpheme (Picket et al 2010: 367) – the labelling of just one type per morpheme (as opposed to per vowel) suggests some grammatical restrictions on the phonetic behaviour that are not found in most other languages in this section.

The combination of phonemic tone and phonation changes is relevant to this thesis but the description of the phonetic realisation of these are difficult to come by. Picket al et (2010) describes the auditory impression of a glottal stop for checked vowels but no acoustics correlates are given for neither checked nor laryngealised vowels, leaving little information for the reader about what type of creaky voice the laryngealised vowels might be classified as. Martin (2010: 3) provides a spectrogram of a laryngealised and checked vowel, indicating that the checked vowels end in a glottal stop with some pre-laryngealisation effects whereas the laryngealised vowels involve a change from modal phonation to laryngealisation with irregular vibrations and back to modal phonation, meaning the non-modal phonation is sandwiched in between two stretches of modal phonation. F1 and F2 are shown in the spectrogram but there is no indication of f0. The paper does not look at the interaction with tone, only the effects of stress on vowel formants which is somewhat independent of phonation, unlike the stød which requires stress to be realised. As such, the knowledge on Isthmus Zapotec adds nicely to the typology of timing of the contrasts in this section but less so in detail about laryngealisation types.

The next relevant question is the origin of the tone contrast in languages with phonemic tones. Brunelle & Kirby (2016) summarise how the loss of some coda consonants have evolved into tonal contrasts and that at some stage, they must have been laryngealised or breathy due to the fact that many e.g. Vietnamese dialects still preserve these phonation types and that they continue to be important perceptual cues. It is still debated, however, whether tonogenesis necessarily involves a developmental stage of voice quality contrast or if this is optional and language-dependent. Brunelle & Kirby (2016: 197-198) observes that languages with a complex tone system based on f0 usually develop from the loss of final laryngeals, both in cases where tone is combined with voice quality and in cases where the tonal changes are the only contrastive cue.

A language that seemed to develop a tone system somewhat independent from processes of laryngealisation is Korean. Korean has three voiceless stops labelled aspirated, lax and tense. They do have some phonation relevant to distinguishing them in that lax stops tend to be breathy and tense stops tend to be laryngealised, but their main distinction has been differences in Voice Onset Time (Yu 2018). This has changed in the modern dialect of Seoul, however, where it has been found that Voice Onset Time is no longer the primary cue to the contrast as aspirated and lax stops now overlap in values (Silva 2006). Instead, the perceptual primary cue is now tone (Kim et al 2002, Kim 2004) which has been taken as indication of a process of tonogenesis in Korean. The f0 differences appear to distinguish particularly the aspirated and lax stop contrast (Yu 2018) with f0 being lower after lax stops. In terms of voice quality, H1-H2 is higher in aspirated stops compared to lax stops, the former produced with breathy voice (Yu 2018: 92). As this section is focused on contrasts on the hyper-compressed continuum, breathy voice is not as relevant despite the tonogenesis being interesting. It is useful to note, however, that breathy voice and creaky voice are both reported to lower f0 in the general literature but in the context of Korean stop contrasts, breathy voice lowers f0 on the adjacent vowel more than the so-called tense voice quality (Yu 2018: 94). Whether this would also be true if comparing breathy voice to a more extreme form of laryngealisation, e.g. vocal fry, is not known, but with f0 effects being similar for both voice qualities, the important distinguishing parameters become other acoustic correlates, such as H1-H2 that is higher in breathy voice but lower in laryngealised voice qualities.

As this section has demonstrated, there are numerous possibilities of the timing and duration of contrastive laryngealisation cross-linguistically. There are also complex interactions with tone systems in some of these languages. Danish is not a language that is considered to have phonological tone but it has been suggested that there is a tonal realisation under certain circumstances of the laryngealisation contrast stød in the Aarhus dialect (Kyst 2004). As dialectal variation has been described for languages like Scottish Gaelic presented above, it has not received much attention for Danish. The stød is described further in the sections below to introduce the phenomenon and the previous findings on the phonetic variability of the contrast.

2.5. The stød in Danish - a case of contrastive non-modal phonation

As described above we can expect, and have observed, considerable variation in laryngealisation subtypes. Some subtypes, e.g. vocal fry, have received more attention while others, such as non-constricted irregular vibrations and non-pathological diplophonia, have received considerably less. This project aims to explore these sparsely described subtypes in vocally healthy speakers in a bigger data set. The phenomenon of stød occurring in Danish is a highly suitable candidate for capturing non-modal variation as studies have shown variability in subtypes during its production (e.g. Fischer-Jørgensen 1989, Hansen 2015). Studies have all been done on Modern Standard Copenhagen, however, and Hansen (2015) only attested subtypes in one speaker, making it highly relevant to expand the empirical pool to more speakers and include a dialect other than the standard to capture more variability in non-modal phonation. This chapter presents some background information on the stød, outlines previous findings and describes some dialectal variation in the realisation of the stød, the latter with the aim of making a case for expanding our knowledge of varieties of the phenomenon outside of Modern Standard Copenhagen as empirical data is limited.

2.5.1. Phonological function of the stød

The stød in Modern Standard Danish is a suprasegmental feature that occurs in stressed syllables with so-called phonological stød basis; either a long vowel or a short vowel followed by a sonorant consonant (Basbøll 2005; Grønnum 2005). Functionally, stød is an important part of the phonology as it is often lexically contrastive distinguishing meaning in words like *ven* [vɛn] ‘friend’ and *vend* [vɛnˀ] ‘turn’ (imperative) or signifying word class in e.g. *aftale* [ˈɑw̩tˢæːˀlə] ‘to agree, to make an appointment’ against *aftale* [ˈɑw̩tˢæːlə] ‘an agreement, an appointment’. Further, the stød is bound to certain word structures under particular morphological conditions and thus can also function to signal morphological structure rather than making semantic or grammatical distinctions. An example is that all originally monosyllabic words native to Danish have stød (Grønnum 2005: 210) and thus the stød can signal the original syllabic structure of a word post inflection.

Not all syllables with stød have a non-stød minimal pair contrast but because syllables with the required stød basis more often than not have stød, what is accounted for in the literature are the rules or mechanisms for when a syllable does not have stød which is thought of as the marked variant. These rules, formulated in the Non-stød model, however, are rather complex and essentially morphological in nature and given the articulatory and acoustic phonetic focus of this thesis they will not be outlined here. For a comprehensive account of them, see Basbøll (2005), or see Grønnum (2007) for a simplified run-through. In summary, the phonological function of the stød is to signal semantic lexical distinctions and to signal certain morphological structures. Basbøll (2005) proposes an analysis of the stød that includes morae and the rules for the presence of the stød are based on moraic structure in this proposal. Grønnum (2005) rejects the proposal of morae being relevant for the Danish stød because she has conducted research showing that there is no unambiguous acoustic or cognitive reality to the domain of the mora and thus it is an abstract phenomenon in the description of Danish. As such, Grønnum (2005) seems to demand a strong psychological reality to phonological analyses whereas Basbøll (2005) seems more generative in his approach. The issue of morae is raised in the discussion of dialectal variation below and will be paused here for now. The phonological scope of the stød is usually described as being the syllable, the stød being a syllabic prosody (Grønnum 2005). Despite being a property of the syllable, however, the stød basis does not always span the entire syllable. This will become more clear in the following sections.

2.5.2. Phonetic characteristics of the stød

The Modern Standard Danish stød realisation is generally described as an irregularity in vocal fold vibrations, or a laryngealisation, that affects both the fundamental frequency and the amplitude of a sound signal. As a phonological feature the stød is binary, either present or not, but as a phonetic phenomenon it is rather gradient. In its weakest form it may be just a slight vibrational change whereas in its most empathetic form it may be a full glottal closure (Grønnum 2005: 215). The following sections outline the current body of research on the stød in terms of its articulation and acoustic expression. It should be noted that even though the stød is a phonological property of the syllable there is a tradition to describe the stød as belonging to either the consonant or the vowel, depending on the stød basis (long vowel or short vowel + sonorant). This syllable vs segment division is reflected in the research where the object of measure may be the whole syllable, the vowel or consonant segments, or indeed all of these.

2.5.3. Previous research

All previous phonetic research has been done on the stød as it occurs in Modern Standard Copenhagen and thus the results presented here apply to this variant. The first instrumental study on the stød was done by Svend Smith (Smith 1944) and included electromyographic, oscillographic and kymographic measurements. This was later followed up by Riber Petersen (Riber Petersen 1973) with acoustic measurements of duration, intensity, f0 and inspections of irregularity in the signal. The most comprehensive study on the articulatory and acoustic aspects of the stød was done by Eli Fischer-Jørgensen (Fischer-Jørgensen 1989) who took electromyographic measures of laryngeal muscles, airflow, pharyngeal pressure, subglottal pressure, larynx position and fiberoptic recordings of the vocal folds. Audio was also recorded for creating mingographic curves of f0 and intensity as well as visual inspections of irregular vibrations. It is not specified what exactly a mingographic curve is, but it is inferred to be a visual representation of acoustic parameters similar to that of a spectrographic visualisation. The most recent research on the acoustic phonetics of the stød is a PhD thesis by Hansen (2015) who very thoroughly tested some acoustic correlates on stød tokens from one speaker and these results are very informative as well. Because of the technical and sometimes invasive nature of the studies mentioned, they each comprise a fairly small speaker set but their level of detail is valuable in uncovering the articulatory and acoustics aspects of the stød. These correlates are usually first and foremost found to be highly variable both within and between speakers, but some trends have emerged from the data and these are presented below.

2.5.3.1. Acoustic findings

Svend Smith's (Smith 1944) oscillographic registrations showed that the intensity and f0 decreased when the stød was present, often causing irregular oscillations.

Riber Petersen (1973) found that the syllable duration varied between different types of stød/non-stød contrasts. There was no consistent difference between syllables with a long vowel without or with stød, respectively designated V:C and V?C in the analysis. However, the difference between syllables without stød with a short vowel and with stød with a long vowel (VC vs V?C) were significant for all speakers, the latter being longer. Though duration differed between contrast pairs, when all tokens were pooled together, the duration difference was not significant. To check if the vowel and consonant length had an inverse effect, consonant duration was also measured but with the same result; no significant difference between stød and non-stød pairs. In cases where the stød manifested as irregular vibrations, the duration of the stød was consistently around 1/3 of the total vowel. Duration being generally non-significant between stød and non-stød pairs has later been supported by Grønnum & Basbøll (2001) who concluded that while stød segments (vowels or consonants) may be longer in some positions and in very distinct speech, this effect is negligible in normal everyday speech and across positions.

For f0, Riber Petersen (1973) observed three different patterns in the data; contours that were falling-rising, purely rising or purely falling. All patterns were present on both stød and non-stød words when all speakers were pooled. However, in syllables with stød on the vowel (V?C), the f0 fall starts later and reaches its minimum closer to the termination of the vowel compared to non-stød syllables with a long vowel (V:C) and is generally lower. The f0 peak was clearly highest in words with a stød vowel + a sonorant consonant (V?C).

For intensity measures, Riber Petersen reports a falling movement for syllables with stød and the intensity minimum is lower compared to non-stød syllables.

Fischer-Jørgensen (1989) found irregular vibrations in around 70% of the tokens with stød, occurring in both the intensity, f0 and airflow curves. She observed a tendency for high vowels to have more frequent irregularity. When reporting her findings Fischer-Jørgensen divides the stød into two phases based on the observation that the f0 curve tends to fall rather late in the syllable, on average 96 ms in. The first part of the syllable thus refers to the portion before the f0 fall occurs and is named the first phase of the stød. In this phase there is a strong tendency for f0 to be higher in words with stød, particularly in disyllables with a long vowel. In the second phase of the stød f0 tends to fall on approximately half of the stød tokens whereas for non-stød tokens f0 is even or slightly rising. The degree of f0 drop is very variable, however, and in some cases cannot be measured properly due to the occurrence of irregular vibrations causing f0 fall-out. Further, in the remaining stød tokens where an f0 drop is not observed this is not due to the curve being even; rather, the waveform is so irregular that a curve cannot be determined in any meaningful way. She concludes that the acoustic manifestation of the stød is very variable but if the stød is strong, typically in more distinct speech, it tends to have irregular vibrations and if it is weak, as in faster, more 'natural' speech, it tends to manifest as an f0 drop, perhaps with some irregularity. The intensity measures follow a similar pattern where the second stød phase generally has a steeper decrease in intensity than non-stød words and is deemed a very stable feature of syllables with stød.

Vowel duration measures showed that the vowel is shortened in syllables with stød but are still longer than short vowels, albeit probably only in distinct speech. If the stød basis is a short vowel + sonorant consonant, the consonant is slightly longer in syllables with stød. However, the difference is not significant and this finding is in overall agreement with Riber Petersen (1973) and Grønnum & Basbøll (2001) – increased duration is not a significant correlate with the stød.

Spectral measures showed that for most speakers, measuring the strength relation between the amplitude of f0 and F1 (F1-f0) with f0 being relatively weaker was useful for capturing the stød syllables, at least for open vowels where f0 and F1 are further apart. Most speakers also had an amplitude decrease in the lower harmonics in the last part of the stød vowels.

The last acoustic measure was deriving a glottal waveform without the formants via inverse filtering for a smaller subset of tokens, just five word pairs from two different speakers. The analysis showed that the glottal waveform amplitude decreased in the second part of syllables with stød, particularly in the second half of the vowel, reflecting reduced airflow in the glottis during this phase.

A more recent comprehensive study of the acoustic correlates of the stød was done by Gert Foget Hansen (2015) in a PhD thesis testing some hypotheses about the voice quality changes that occur during the stød. This was done on data from only one speaker but yielded some interesting results, particularly in regard to subtypes of the stød. The main hypothesis was that the stød is realised as a relatively short change in voice quality towards a compressed or creaky voice and back, thus it is a dynamic voice quality gesture that needs to involve a sufficiently large quality change during a sufficiently short time interval. Hansen (2015) argues that this hypothesis may explain why there are not always irregular vibrations during the stød and why stød realisations without these are not necessarily perceived as reduced in any way – what matters in terms of distinctness is the degree of movement up and down the voice quality scale, not how far up one gets as in an absolute value target, i.e. irregularity per se does not need to be reached if compression is executed in an appropriate time and manner. Hansen's (2015) acoustic measures focused more on harmonics and spectral information than duration, f0 and intensity which makes it an excellent addition to previous studies where these aspects have not been thoroughly measured, probably partly due to previous studies being from before speech science technology was well developed to conduct these analyses with relative ease.

The results showed that in tokens where irregular vibrations were present these did overwhelmingly not constitute the peak of a rising-falling compression interval based on measures of changes in H1-H2, periodicity and intensity. Because of this Hansen rejects the hypothesis that a short compression interval is the primary cue for the stød, although for some tokens this was the case. Instead he proposes that irregularity can occur independent of compression in the voice source because the compression peak and intensity dip (which he interprets as correlating with glottal constriction) can be timed in at least two different ways and thus high compression and glottal constriction cannot be codependently linked. He stresses, however, that he still believes that there is new knowledge to be gained by looking at the changes in compression in relation to the stød because the stød vs non-stød tokens still differ in the compression course and the compression degree, the former being more dynamic and to a higher degree. The stød tokens generally have a rising-falling compression course but the timing of it varies so that either (i) the compression peak and the intensity minimum occur close to each other and simultaneously with irregular vibrations, or (ii) the compression peak occurs before the intensity minimum (approximately 80-100 ms before) and before the onset of irregular vibrations. Hansen calls the former stød realisation type P and the latter type N and advocates for investigating these across a larger dataset to see if, as his data suggests, they are distributionally consistent.

In summary, the results from the four studies show that the prototypical stød has the acoustic correlates of an increased f0 and intensity in the first phase and a falling/lower f0, decreased intensity and often irregular vibrations in the second stød phase (deemed 'the stød phase proper'). Further, the airflow is decreased towards syllable termination. These findings indicate that the stød often involves a shift from modal to non-modal, creak-like phonation caused by a constriction in the glottis but that these do not always co-occur with irregularity and/or a compressed voice quality.

In addition to the acoustic measurements, Smith (1944) and Fischer-Jørgensen (1989) also obtained some physiological measurements and the results of these are outlined below.

2.5.3.2. Physiological findings

Smith's (1944) results suggested that syllables with stød are marked by a 'thrust-like' emphasis caused by a ballistic contraction of the respiratory muscles followed by an abrupt fall in their activity causing a fall in subglottal pressure. If this is not counteracted it may cause irregular vibrations. Smith thus concluded that the laryngeal activity he found in his acoustic recordings were a secondary effect caused by forceful muscle activity in the respiratory muscles. As he only did physiological measurements of these muscles, and not the larynx, this conclusion is not incompatible with the data – however, as more refined methods have become available, there seems to be little in favour of laryngeal activity being a secondary effect of respiratory activity during the stød (Fischer-Jørgensen 1989).

Fischer-Jørgensen (1989) did airflow measurements for a small set of speakers using an aerometer for oral airflow on one speaker and combined oral-nasal airflow using a pneumotachograph on four speakers. In the first phase of the stød measures were variable and inconsistent but for the second phase, airflow was significantly lower for syllables with stød, indicating constriction of the glottis. Pharyngeal air pressure was taken for three speakers and showed a slightly lower pressure for syllables with stød in the second phase – this was rather variable and not significant, however. To investigate if the articulatory force was greater in the first stød phase compared to non-stød vowels, palatograms were analysed of five different vowels produced in isolation by five speakers with and without stød and the results showed a larger contact area in vowels with stød, indicative of more articulatory force. Subglottal pressure was measured for one speaker and was higher in syllables with stød which was also the case for oesophageal pressure.

Fiberoptic recordings of the vocal folds were made for seven speakers for word pairs with the vowel, /iː/. The frames for analysis were chosen so they corresponded to the point of lowest f0 and intensity in the stød phase proper. All subjects except one (whose folds were covered by the ventricular folds during contraction and could not be seen) showed signs of contraction in that the distance between the folds was narrower when the stød was present and in some cases the folds were slightly shortened too. Subjects also had ventricular fold approximation to varying degrees.

Laryngeal muscle activity was also measured for a subset of two sepakers via electromyography (EMG) of the vocalis muscle, the cricothyroid muscle and the lateral crico-arytenoid muscle. A few recordings of the interarytenoid and the posterior crico-arytenoid muscles were also made but the preliminary recordings showed no difference in activity between stød and non-stød syllables for these and measuring them was not pursued further. The cricothyroid muscle showed greater activity in the first phase of syllables with stød compared to non-stød which matched the higher f0 measured in this phase – the cricothyroid muscle lengthens the vocal folds, making them thinner and more tense which in turn raises f0. In the majority of speakers cricothyroid activity did not, however, fall in the second phase where f0 also dips. Fischer-Jørgensen (1989) further discusses the role of the vocalis muscle in controlling f0 but from her data concludes that its function may be not to rapidly raise f0 but rather to help keep it high – however, vocalis activity is not necessary to achieve this as high f0s were recorded without vocalis activity in many instances. The vocalis muscle is of interest to the stød, however, because it has been found to be active in glottal constrictions such as so-called hard attack (Faaborg-Andersen 1957; Hirose & Gay 1973) or laryngealised sounds such as fortis stops in Korean (Hirose et al 1974). Fischer-Jørgensen found a consistent peak in vocalis muscle activity for stød syllables in 5 out of 7 speakers starting around 20-40 ms into the vowel. The effect was stronger in high vowels compared to low vowels and stronger in monosyllables. Another muscle that tends to be active in glottal constriction is the lateral crico-arytenoid muscle. This muscle also helps to facilitate voicing and the raising of f0. Its activity in Fischer-Jørgensen's (1989) data looks almost exactly like the vocalis muscle activity, i.e. a peak on syllables with stød, indicating a glottal constriction achieved by these two muscles combined. Fischer-Jørgensen notes that vocalis and lateral crico-arytenoid activity is expected to raise f0 in modal phonation but that the opposite happens during the stød – they are active but f0 falls in the second phase. She argues that this strong and sudden muscle activity can cause an overcompression of the vocal folds which dampens the vibrations and causes the irregularity observed in the acoustic measurements and thus is not so unexpected after all. This is supported by the timing of the muscle activity peaks vs the drop in f0 and irregular vibrations which always have their onset immediately before or right at the vocalis and lateral crico-arytenoid muscle activity peak. The resulting constriction would also explain the low intensity found in the last part of the second stød phase but strangely not the onset of the intensity decrease at this seems to happen well before the muscle activity peak.

2.5.4. Dialectal variation

The two dialects under study in the present project is Modern Standard Copenhagen and the Aarhusian dialect spoken in Aarhus, Eastern Jutland. The stød in the former has been described in the previous sections; phonetic stød research has been done almost exclusively on this variety so all of the above reported findings apply to the Copenhagen variant. This section thus aims at describing the variety spoken in Aarhus – there are, however, many more dialects where the realisation of the stød and its distribution differs from Modern Standard Copenhagen. For an overview of some of these varieties, see e.g. Ejskjær (1990).

Bodil Kyst (2004; 2007) has made a comprehensive description of the tones and intonational patterns of Aarhusian speakers based on three informants reading sentences aloud. The tones in Aarhus are relevant to the stød because in this dialect, the stød is hypothesised to be realised as a falling tone rather than a glottal constriction under certain conditions. Kyst (2004; 2007) uses the term regiolect rather than dialect in her study but does not comment further on the choice of terminology. In the following I will use the term dialect to match the Copenhagen variety which is described as a dialect, not a regiolect, and because a distinction between dialect and regiolect is not particularly descriptive or imperative for the present project which is only concerned with variation in stød realisations, not characterising the dialects as a whole. Kyst (2004; 2007) measured so-called stress groups ('trykgrupper') consisting of the first stressed syllable and all following unstressed syllables until the next stress falls which constitutes the beginning a new group. Three informants from Aarhus read sentences aloud and the analysis consisted primarily of comparative visual analysis of intonation curves created in Praat (Boersma & Weenink 2021). The results demonstrated that stress groups with stød behaved markedly different to stress groups without stød. If there is stød on the first syllable the tonal pattern is High-Low (HL); a high tone on the first (stressed) syllable and a low tone on the posttonic syllable. This in consistent across phonetic environments, whereas for the stress groups without stød the contour depends on whether there are voiced sounds between the vowels in the first and second syllable. If this is the case, the tonal pattern will be Low-High (LH). If there is a voiceless sound, however, the pattern tends to be HL, but this depends on the vowel length in the first syllable.

An example of a Low-High tonal pattern on a non-stød word and a High-Low tonal pattern on a word with stød in Kyst's material is given below:

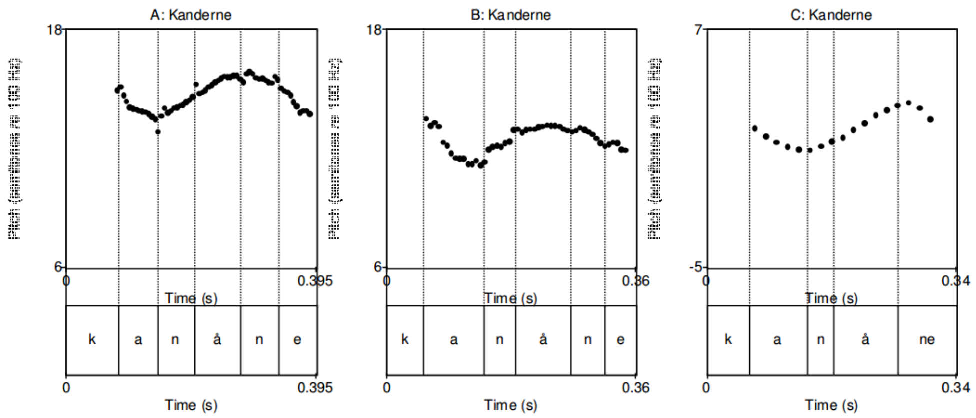

Figure 2. Low-High tonal pattern in the word kanderne 'the jugs' spoken by three informants from Aarhus taken from Kyst 2004: 11, Figure 6.

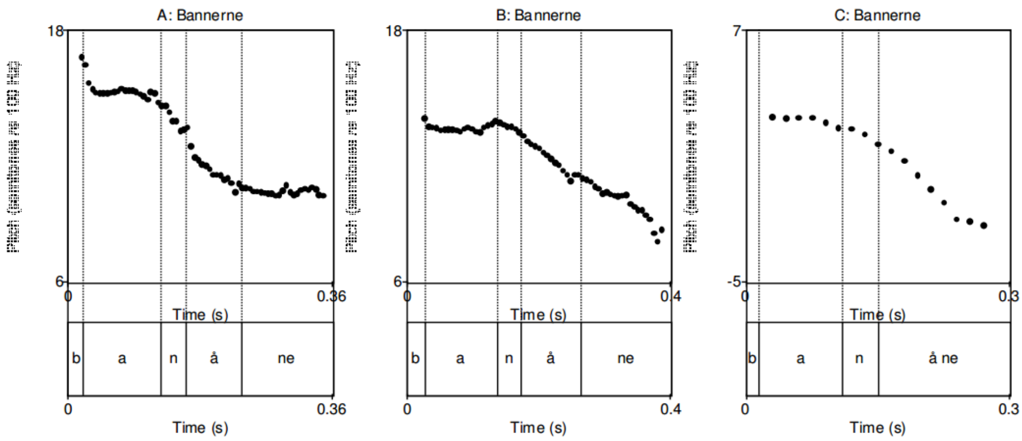

Figure 3. High-Low tonal pattern in the word bannerne 'the banners' spoken by three informants from Aarhus taken from Kyst 2004: 11, Figure 7.

Kyst (2004; 2007) argues that the respective tonal patterns in these stress groups are so distinct and regular that they function to signal if there is a stød on the same level as the actual stød – or, more precisely, together with the stød as they always co-occur. She argues that the posttonic tonal fall is part of the realisation of the stød in her data. She goes on to describe a subtype of the Aarhusian stød, a phenomenon observed on monosyllabic words where the stød is realised with a particularly long, falling tone which gives the auditory impression of a so-called bi-tonal pronunciation or even disyllabic realisation and Kyst perceives these words as carrying both tones of the stress group, i.e. the full HL pattern even though they are phonologically monosyllabic. This bi-tonal pattern occurs on closed syllables with long vowels, e.g. *lån* [lɔːˀn] 'loan', but can also occur on syllables with a short vowel followed by two voiced consonants, e.g. *film* [filˀm] 'film'.

Kyst speculates that the perception of two tones, or two syllables, may be caused by the fact that the tonal peak and trough of these monosyllables are as high and low as the rest of the tonal peaks and troughs of the sentence intonation and thus this one syllable contains both extremities that would usually characterise an entire stress group. An example from her materials are given below for the word *film* 'film':

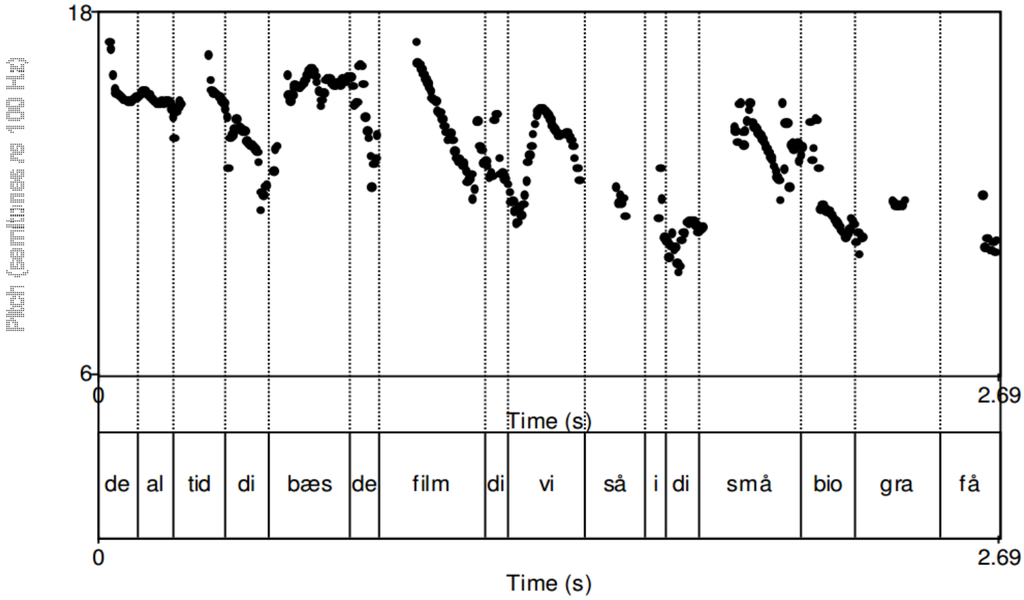

Figure 4. A demonstration from Kyst 2004: 41, Figure 43, of the 'two-toned' word 'film' in the context of the sentence melody. This figure demonstrates how the tonal pattern on film reaches the same high and low extremities as the overall sentence intonation.

Kyst (2004) discusses if this bi-tonality is related to something phonological or something articulatory phonetic in nature. She speculates that the bi-tonal pattern could be linked with morae, specifically bi-moricity, but ends up rejecting this explanation because if morae are counted as suggested by Basbøll (1988) for Danish then all monosyllabic words would have two morae because they have a long vowel or a short vowel followed by a sonorant, i.e. stød basis. Because bi-tonality only occurs on closed monosyllables with a long vowel and not on all syllables with stød basis (e.g. on lån [lɔːˀn] 'loan' but not on lund [lɔnˀ] 'grove') bi-moricity cannot be the cause of this phenomenon. Further, it is disputed whether the notion of morae is even appropriate for Danish – Grønnum (2005: 219) argues that there is no defined either acoustic or cognitive reality connected to the domain of morae and that it is thus an abstract unit in the description of Danish phonology and for this reason does not use it. It will not be attempted here to clarify whether morae are appropriate for Danish phonology as that is outside the scope of this thesis. That said, it is both relevant and convincing that Kyst (2004) finds that monosyllabic words (i.e. words that should have two morae if morae are accepted for Danish phonology) are not all articulated with the bi-tonal pattern observed in the data for the Aarhus speakers. As such, morae cannot be the defining factor in tonality here as that would mean all monosyllabic words would have a bi-tonal pattern which they do not. Only closed monosyllabic words with a long vowel have the bi-tonal pattern.

Kyst (2004) writes that on these long-vowel bi-tonal monosyllables she does not perceive any glottalisation or creakiness but rather the stød seems to be realised as the falling tone alone and calls for more research on whether the stød is realised differently in these syllables compared to open syllables and short-vowel closed syllables. I agree with her assessment that the stød on these syllables does not seem perceptually creaky in nature. In fact, a primary reason to study the Danish stød as a case when exploring phonetic variation in phonological contrasts is that Kyst's (2004) findings suggest these dialectal differences as a source of variation which not much is yet known about.

The suggestion of a potential tonal stød in Aarhus merits a note on the difference between tone and laryngealisation as f0 is also affected by laryngeal behaviour. In Section 2.4. above it was noted that e.g. the languages of Southeast Asia combine f0 and phonation types to realise tonal contrasts in ways that are suggested to lie on a continuum of 'pure' tone on one end, i.e. changes in f0 signals the contrast only, and *register* on the other end, an umbrella term covering the use of f0, voice quality, vowel quality and duration to distinguish phonemic contrasts. In the context of tonal languages, Pike suggested that tone can be defined as 'the presence of lexically significant, contrastive, but relative pitch on each syllable' (Pike 1948: 3). This definition does not necessarily exclude simultaneous laryngealisation, however. Brunelle & Kirby (2016) make the point that while categorical labels like 'tone languages' or 'register languages' can be descriptively useful, the complexity of the interaction between tone systems, register and phonation types means it can be questioned if the distinction is even useful. For the purpose of discussing Danish stød in this thesis, the tonal variant in Aarhus is expected to lie on the 'pure tone' end of the continuum of 'pure' tone on one end and *register* on the other end. This assumes that tone is the primary perceptual cue to the contrast based on descriptions in Kyst (2004) and my own auditory impression, as well as being a speaker of the Aarhusian dialect myself, thus having some intel into what the articulation of the tonal stød intuitively feels like. As such, tone is defined in the context of the stød as the absence of evidence of overt laryngealisation.

2.6. The origin of the stød

Before commencing the analysis of the stød it is worth providing some background on its historical origin as this background appears to involve an interplay between laryngealisation and tone which is very relevant to the current dialectal variation.

A few theories have been proposed of how the stød has developed into its current form. The vast majority of scholars agree that it arose from an originally tonal contrast (Sweet 1873; Verner 1881; Jespersen 1897; d'Alquen & Brown 1992; Fischer-Jørgensen 1989; Gårding & Lindblad 1973; Oftedal 1952; Riad 2003a; Brink 2000; Skautrup 1944). A few have argued, however, that the stød came first and then brought way to a tonal accent contrast (Kroman 1947; Andersen 1962; Malling 1982; Liberman 1984). Most recently this so-called stød first-hypothesis has been reproposed (Wetterlin & Lahiri 2015), arguing that an encliticisation occurred of the demonstratives *hinn* and *hit* to nouns around year 1000 AD. These developed into definite suffixes and to avoid resyllabification when they attached to monosyllables, the /h/ strengthened to the glottal stop /ʔ/. From here it was reanalysed to belong to the monosyllabic stem and developed into a stød/laryngealisation rather than a full glottal stop. With time this extended to some disyllabic words via epenthesis, i.e. the insertion of a *schwa*, which blurred the transparency in the relationship between the presence of stød and syllable count, making the stød a marker of prosody in words that were prosodically unpredictable. The stød was then reanalysed as tone, particularly via contact with Norwegian and Swedish as it spread to these languages and manifested as a tonal contrast. However, as argued by Goldsthein (2020: 27) this hypothesis is unlikely given (i) the lack of documentation for /ʔ/ replacing /h/ at any stage, (ii) the articulatory and aerodynamic unlikelihood of /h/, which has a spread glottis and low subglottal pressure, developing into /ʔ/ which has a constricted glottis and high subglottal pressure and (iii) the claim that the stød has its origin in Denmark and then spread to Norway and Sweden as a tonal contrast is not supported by neither the history of the definite marking distribution in Scandinavia nor by the current distribution of the stød in the dialectal landscape. I will thus assume the majority position that the tonal accent on words came first and that the stød developed from this. The question is, then, how the word accents emerged and what mechanisms let to the development of the stød from these tones. The following section briefly highlights the synchronic relationship between word accents and the stød before looking at their diachronic development.

2.6.1. Scandinavian word accents

As established above, the general consensus is that the stød is related to the word accents found in Norwegian and Swedish and that these preceded it. The word accents seem to have arisen from diachronic monosyllables (Accent I) and polysyllables (Accent II) in Old Scandinavian (Bye 2004; Bye subm.; Iosad 2016), more on this below. Today they distinguish lexical groups by tonal behaviour (Bruce 1977). Gårding & Lindblad (1973) further distinguishes some dialectal variation within the word accents; Type 1 dialects have one tonal peak in Accent II words, Type 2 dialects have two tonal peaks in Accent II words. The timing of the peak also differs, prompting a further subdivision of an early syllable peak (Type A) and a late syllable peak (Type B). An example of dialectal variation on the word /lame/ in two Norwegian dialects are given below, borrowed from Bye (2011). The boxes represent disyllables and the middle line the syllable boundary:

Figure 5. Accent I (left) & Accent II (right) tonal pattern in the Nordland dialect of Norwegian.

Figure 6. Accent I (left) & Accent II (right) tonal pattern in the Oslo dialect of Norwegian.

The Accent I vs Accent II distribution was demonstrated and described by Bruce (1977) who found that the primary governing factor in word accent distribution in non-compound words was stress placement (1977:17). Both the number and timing of peaks, however, are affected by the word when embedded in a sentence and as such, the categorisation into types is largely based on tonal behaviour in standalone words. Bruce does, however, also account for Accent distribution in more complex compound words, stating that the majority of these will receive Accent II independent of whether the two words in the compound would be Accent I or II in isolation (Bruce 1977: 19).

The Standard Danish stød distribution roughly corresponds to Accent I and some non-stød words to Accent II. However, there are some differences in the phonological requirements; Accent II production requires two syllables whereas the stød is contrastive in monosyllables but needs either a long vowel or a short vowel followed by a sonorant consonant to be realised, cf. stød basis. Further, only one accent can occur per word in Norwegian and Swedish, whereas the stød can occur more than once in a word as long as the stød basis is retained. To summarise, the stød contrast roughly corresponds to Accent I words, but there are differences in their phonological demands.

Two main hypotheses have been proposed of how the word accents developed. The Proto-Nordic hypothesis aims to explain the fact that Accent II is found primarily on Proto-Nordic words that were trisyllabic but underwent syncope around 800-850 AD to become disyllabic (Kock 1885; d'Alquen & Brown 1992). Riad (1998a; 2003b) proposes that trisyllabic words with more than one heavy syllable had two stresses, each of which correlated with a tonal peak, but that syncope of the unstressed syllable resulted in two adjacent stressed syllables. Cross-linguistically, this sort of stress clash tends to be dispreferred and thus resolved (Nespor & Vogel 1989) which in Proto-Nordic leads to reinterpretation of the final syllable as unstressed. However, the tonal peak of the final syllable was still retained, though no longer correlating with any syllable stress, creating what instead became a two-peaked word accent. An example of this process is the Proto-Nordic *ˈdōmiˌjan* 'to judge' > *ˈdōmˌjan* into Modern Swedish *döma* /dø:ma/ with Accent II. This theory thus suggests that Accent II preceded the one-peaked Accent I which emerged from a process of tonal retraction or a leftward accent shift (Riad 1998a; 2003b).

This hypothesis assumes a rhythmic principle that Proto-Germanic words with a non-initial heavy syllable receives secondary stress. But as Goldsthein (2020) rightly points out, such a rhythmic principle is not found in any Germanic languages or Germanic contact languages (Bye 2004: 44; Kristoffersen 2004). Further, the suggestion that Accent II arose through tonal retraction is typologically unlikely as tones tend to delay rather than retract (Goldsthein 2020). These facts make the Proto-Nordic hypothesis less likely as it is based on assumptions that lack documentation and seem implausible based on evidence from modern languages.

An alternative account of the emergence of word accents, named the Old Scandinavian hypothesis, instead relies on syllabic pitch peak alignment in Old Norse (Oftedal 1952; Perridon 2006a; Bye 2011; Iosad 2016). It is proposed that the pitch peak on the stressed syllable aligned earlier in monosyllabic words than in disyllabic words. Around year 1100 AD and on, the distinction between mono- and polysyllabic words in Old Norse was blurred by the emergence of new syllables caused by encliticisation and epenthesis but the originally monosyllabic words retained their early pitch peak alignment although no longer being monosyllabic. Accent I words thus developed first and from them, the two-peaked Accent II emerged due to peak delay, creating a contrast between Accent I and Accent II based on tonal peak behaviour. An example given in Goldsthein (2020) is the Old Norse *búit* /bu:it/ 'inhabit-PTCP.NOM.SG.N' that had late peak alignment and *bú* /bu:/ 'dwelling' that had early peak alignment. When the latter encliticised with the definite suffix -it, it became disyllabic, *búit*

/buːit/ ‘the dwelling’, and converged with the originally disyllabic word and now contrasted only in word accent, manifested in their tonal patterns such that Accent I *búit* translated as ‘the dwelling’ and Accent II *búit* translated as ‘inhabit-PTCP.NOM.SG.N’.

Goldsthein (2020) renders this hypothesis plausible based on the overall tendency for tones to delay cross-linguistically, and the fact that this delay has been observed to generalise to pitch peaks in stressed syllables with a longer prosodic domain in a number of languages (Silverman & Pierrehumbert 1990; Grabe 1998; Dalton & Ní Chasaide 2005).

2.6.2. The emergence of the stød

Historical evidence of the stød is not as rich as would be desired but it is believed to be referred to in a somewhat infamous speech by the Swedish bishop Hemming Gad in 1510 AD. It describes the, according to the speaker, unbecoming way that the Danes ‘press the words forward as if they will cough’ (translated in Basbøll 2008) and how Danes turn the words in their throats, twisting and sneering them, and ‘thinking this to be a particular ornament and well standing’. Whether what is referred to is actually the stød is disputed and the stød is not explicitly mentioned until around 200 years later, in spite of quite thorough grammars written at the time (e.g. Madsen 1586). The first undisputed record is by Høysgaard (1747) who describes syllable types in Danish based on, among other things, the presence or absence of the stød, which is described phonetically as a little hiccup. This most certainly means it is the Standard Danish stød as only this would be auditorily akin to a hiccup. The stød is generally assumed to precede the written sources on Danish language and part of its diachronic evolution therefore relies on less direct evidence. Attempts to date its emergence have relied on the points in time where sound changes occurred that gave way to a larger number of words having stød basis, i.e. the sonorant requirement for the stød to occur at all. Skautrup (1944) puts emphasis on Danish plosive lenition where some unvoiced consonants became voiced and thus provided the necessary stretch of voicing when combined with a short vowel. This would put the development of the stød around 1100-1200 AD. Jespersen (1897) and Andersen (1962) instead date it around 1300 AD, arguing that the relevant development was a hyper-lengthening of consonantal codas in Danish monosyllables. Brink (2000, 2018) proposes a more specific version of this, namely that the nasal /n/ lengthened before the vowel /ø/ and this was the precursor for the stød, resulting in a later emergence around 1450-1550 AD.

While it is obvious that the stød cannot have emerged without the necessary sonorant stretch enabling its production, it is less than obvious that the stød will have emerged synchronously with these changes. The posited hypotheses further imply that the tonal accent system was replaced entirely and at once by the stød contrast which appears unusually abrupt given the fact that the distribution of the stød has gone through dynamic processes of expansion and retraction in the lexical domain during the last 150 years (Brink & Lund 1975: 480). Further, some occuring sound changes that would enable stød, e.g. /r/-vocalisation from *spark* [spɑʁ̥k] to [spɑːk] '(a) kick', do not guarantee its presence – indeed, there is no stød in *spark* (Goldshtein 2020). To summarise, we do not have any direct evidence for when the stød emerged as it is not mentioned explicitly until 1747 AD (Høysgaard 1747) but it has in all likelihood been around for longer. What has been dated instead is the development of the sound changes necessary to make the stød an articulatory possibility and this exact time period is also disputed, ranging from 1100 AD to 1550 AD.

How the stød developed has too been given different accounts, e.g. Sweet (1873) who suggests that the stød links to the falling tone of Accent I, where others propose it arose from a tonal rise in Accent I rather than a fall (Verner 1872; 1881, Storm 1874, Jespersen 1897). Others, still, hypothesise that it arose from an intensity concentration on the stressed syllable of disyllabic words (Kristensen 1899, Smith 1938; 1944, Skautrup 1944). What the different accounts have in common is tying the development of the stød in with the weakening of unstressed syllables occurring around 1200 AD and onwards. This weakening would prompt unstressed vowels to reduce to /ə/ or drop completely along with the lenition of consonants. To compensate, the stressed syllable was strengthened via greater articulatory force, leading to wider tone contours and/or higher intensity (Skautrup 1944). What renders these hypotheses less likely, however, is the lack of evidence for the strengthened stressed syllables on which they are built.

Another approach has been to relate the stød to a similar phenomenon in the language Livonian (Kiparsky 1995) combined with a phonological analysis of the stød positing that it is a phonetic realisation of an underlying falling tone on one syllable (Ito & Mester 1997) due to Danish having High-Low pitch contours associated with the stressed syllable and the right edge of the prosodic domain, which, when occurring on one syllable, will be realised as stød (Riad 1998b, 2003a, 2009). This, however, also seems unlikely. Although some dialects may have a High-Low intonation pattern in unmarked sentences, this is not true for Modern Standard Copenhagen which has a Low-High-Low intonational contour (Grønnum 1992, Vazquez-Larruscaín & Basbøll 2013) and as Goldsthein (2020) points out, the stød also appears to autosegmentalise on a separate tier from sentence intonation. Instead, Goldshtein (2020) proposes that the source of the stød was the phonologisation of a historically non-contrastive laryngealisation and that the weakening of unstressed syllables enabled the promotion of this laryngealisation to become the primary contrastive cue in former Accent I words. The argument is that articulatory biases rooted in aerodynamic correlates of prosodic boundaries, especially when implemented by a low tone, resulted in irregular phonation being perceived as laryngealisation which then entered the non-contrastive pool of variance for Accent I. Speech perception mechanisms favours laryngealisation when it cues boundaries between a syllabic enclitic and a stem and the development of the stød was thus grounded in phonetic substance, what Goldsthein (2020) refers to as a 'natural history'. Laryngealisation worked to enhance the prosodic boundary that was primarily manifested tonally until now and was then gradually reinterpreted as a phonological target rather than a phonetic variant due to a shift in cue weighting. This was made possible by the weakening of unstressed syllables in Early Middle Danish which consequently weakened the ability of these syllables to carry tonal cues to indicate their underlying prosodic contrast. Since the tonal accent contrast is understood as a difference in underlying prosodic structure, this devalued the tonal information in favour of a laryngeal cue. Relating this diachronic development to the synchronic interpretation of the stød, Goldsthein tentatively concludes that 'the [..] stød should be seen as the manifestation of an underlying contrast in prosodic structure, the cues to which are neutralized when the stød either does not have sufficient sonorous material in the rhyme or when it did not develop stød historically. Stød is therefore not to be understood as a fully predictable outcome of synchronic structure, e.g. moraic structure [..]. Rather the presence/absence of stød should largely be explained in diachronic terms' (Goldhstein 2020: 57).

2.7. Research questions

Having reviewed the main literature on phonological use of creak-like voice quality and the stød, some research questions are posed below. The main research focus is an exploratory analysis of phonetic variation in a phonological voice quality contrast, adding a dimension of dialectal variation rarely explored in languages with contrastive laryngealisation, and not previously covered by experimental phonetic research on Danish which has only studied the dialect of Modern Standard Copenhagen.

As the stød is described as a laryngealisation akin to creaky voice, it is expected to be produced as a variety of subtypes on the continuum from prototypical creaky voice to more divergent subtypes. As the dialect of Aarhus is expected to have a stød type that is tonal in nature, this will add a further variant to the possible articulation of contrastive laryngealisation.

To accommodate the variation expected to be found in Aarhus the stød is divided into two types for the purpose of the research. As this is a novel approach, the distinction between the types will be clarified again here. The term regular stød refers to syllables where the stød basis is either a long vowel in an open syllable in a disyllabic word or a short vowel followed by a sonorant consonant in a monosyllabic word. Examples of words with regular stød in disyllabic words include *viser* [ˈviːˀsʌ] ‘shows’ and *griner* [ˈɡ̊ʁiːˀnʌ] ‘laughs’. Examples of regular stød on monosyllabic words include *vild* [vilˀ] ‘crazy’ and *selv* [sɛlˀ] ‘self’. The tonal stød is referring to when the stød occurs on a long vowel in a monosyllabic word. Examples of this type include *hvil* [viːˀl] ‘rest’ and *sæl* [sɛːˀl] ‘seal’. As such, monosyllabic words can either have (i) regular stød as in [vilˀ] or (ii) tonal stød as in [viːˀl], depending on whether the stød basis is a long vowel ([viːˀl]) or a short vowel and a sonorant consonant ([vilˀ]). Long vowels in disyllabic words always have regular stød but the stød basis is different from the regular stød in closed monosyllables.

A secondary, but tightly related, interest of this thesis is methodological approaches to measuring laryngealisation. Previous literature has identified acoustic and articulatory measurements useful to capture creak-like voice qualities and their subtypes and it is highly relevant to build on this by investigating their utility across data from different languages as well as their utility in capturing language-internal variability. Studying Danish data from two dialects contributes to this empirical realm by analysing a less-studied language and documenting dialectal variability within the language, testing whether these measures are appropriate for different languages with laryngealisation contrasts. The research questions are divided into three parts:

1) How much dialectal phonetic variation occurs during the Danish stød, a binary phonological contrast?

2) Which measurements correlate with (subtypes of) the stød?

3) How are the voice quality changes timed?

3. Methods

3.1. Ethics

Ethics approval for this study was granted by the Faculty of Arts and Social Sciences and Lancaster Management School's Research Ethics Committee at Lancaster University. As the ethics application was submitted during the Covid-19 global pandemic, the procedure requirements for obtaining data from live participants were stricter than usual, including requirements of social distancing which had implications for how the data was collected, more on this below.

3.2. Materials, participants and recording procedure

A word list was created to elicit two different stød contrast categories; non-stød vs regular stød and regular stød vs tonal stød. The words were minimal, or in a few instances near minimal, pairs differing only in their respective stød conditions. For each stød pair category there were words with high, mid and low vowels. This design was done to allow for future studies of how the vowel quality affects the stød. This direction was not pursued here but the consideration in design meant that two near-minimal pairs were included to elicit the stød in a certain vowel context. The rest, however, were strict minimal pairs differing only in the stød. Each word was imbedded in a sentence which, due to the ambiguous nature of Danish orthography, could not be a standard carrier phrase as context would need to be provided to suggest which kind of production was relevant in the case of stød/no stød homographs such as *griner* (laughing/a laughing fit), *løber* (running/a runner), etc. The target word was never the ultimate or penultimate word of a sentence to avoid a potential confounding effect of the stød realisation by the occurrence of phrase-final creaky voice.

A total of 21 speakers were recorded, 11 speakers from Aarhus (6 male, 5 female) and 10 speakers from Copenhagen (5 male, 5 female) between 23-41 years of age, average age 30.14. Participants were recruited via social media platforms and word of mouth. Inclusion criteria were that they had grown up in their respective dialectal area and that they had lived in this area within the past 2 years. Further, they could not be diagnosed with any vocal pathologies. For the males they could not have a neck beard as this could possibly degrade the quality of the Electroglottographic recordings. Participants did not receive any compensation for their participation which was entirely voluntary.

The recording procedure was the same for both dialect groups, albeit carried out in different labs. The Aarhus dialect speakers were recorded at Aarhus University's Phonetics Lab and the Copenhagen speakers were recorded at Copenhagen University's Phonetics Lab.

Participants sat in a sound attenuated booth in front of a laptop screen showing the elicitation sentences. The stimuli were presented in random order using the software SpeechRecorder (Draxler & Jänsch 2004) with three repetitions of each sentence. Participants were instructed to read them in a conversational manner, controlling when the next sentence was shown by pressing a button on the screen. If they mispronounced a word they were instructed to start the full sentence again to limit the influence of unexpected stress placement due to repair on the corrected word(s). The participants were informed that the study was about vocal fold vibrational patterns in everyday spoken Danish but not that the word list elicited the stød specifically as to not bias them in their production.

Audio was recorded through an Audio-Technica AT803 omnidirectional microphone which participants clipped on to clothing at chest height. The audio was digitized using a Sound Devices USBPre2 audio interface, recorded to a desktop computer at 44.1 kHz with 16-bit quantization. The vocal fold contact patterns were recorded using a Laryngograph® Electroglottograph synched with the audio signal via the USBPre2 interface. Participants were instructed to fit the electrodes on to their own necks on either side of the Adam's apple (the thyroid prominence) as physical contact post Coronavirus pandemic was still to be minimised in order to follow the safety procedure outlined in the required risk assessment for the data collection. All data collection was carried out by one researcher (myself) and the signal levels were checked and corrected pre-recording if not within the normal range on the USBPre2 interface.

64 sentences (32 minimal pairs) were recorded per speaker, each sentence repeated 3 times, giving a total of 192 tokens per speaker. However, a few tokens did not record properly due to software errors or due to participants not pressing the button at the right time, a downside of having them control the sentence display themselves. All tokens of one word pair were removed because after recording, it was decided that they could be argued to not constitute a phonologically minimal pair as initially thought, namely *puder* 'pillows' vs *pudder* 'powder'. This is because the word *pudder* can be analysed with the consonant belonging either to the first or the second syllable, and if it is analysed as belonging to the first syllable, it is no longer a minimal pair with the first syllable in *puder*. Another pair, *piber* '(smoking) pipes' vs *piber* 'squeeks', was removed for participants who lenited the intervocalic consonant in the latter but not the former, making them no longer phonetically similar. The removal of these

two word pairs constituted most of the excluded tokens. A few tokens were discarded due to mispronunciation of the target word (e.g. *onder* 'evils' pronounced as *ånder* 'spirits/souls') or in two cases because there was non-stød creaky voice on the entire elicited phrase, making the voice quality not a function of a phonological basis for stød and thus not deemed appropriate for the analysis of variation in a phonological contrast. It could be argued to take a different approach and include tokens with creaky voice on the entire phrase as this occurs in natural speech. This approach was not taken because the focus of the study was contrastive use of non-modal voice quality and sociophonetic phrase-spanding creaky voice cannot be considered phonologically contrastive. Occurring only on two sentences, this choice did not affect the final tokens included to any noticeable degree. The final number of tokens for analysis was 3788, 94% of the total tokens planned for.

3.3. Acoustic data processing and analysis

3.3.1. Data annotation and measurements

The data was first annotated manually using Praat (Boersma & Weenink 2021) where tiers were created marking the phonological stød basis for each word. Recall that the stød basis consists of a long vowel or a short vowel plus a sonorant consonant. For homographs, the stød vs non-stød contrast was marked on the word with 1 for stød and 2 for non-stød. If there was additional creak occurring on segments after the stød basis this was marked on a separate tier. An example of an annotation is given below:

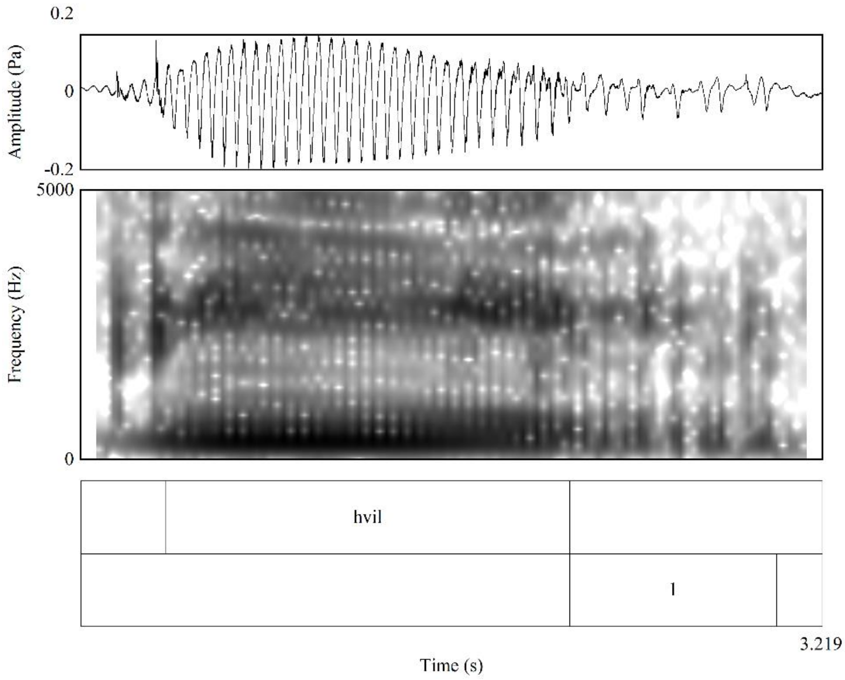

Figure 7. Annotation example of the word hvil 'rest' spoken by speaker FCPH2, marking the stød basis [iː] on Tier 1 with an orthographic word transcription and with the post-stød creak occurring on the [l] marked on Tier 2.

Segment boundaries were informed by the spectrographic display and the waveform, supported by auditory impressions. However, in some cases the segment boundary was not clear, namely for words with a so-called 'soft d' and a vowel, e.g. 'taget' [ˈtˢæːˀð̞] 'the roof', where the two sounds are strongly merged due to the vowel-like nature of the alveolar approximant. In these cases the annotation was based on mostly auditory impressions of where the vowel and consonant boundary was supported visually by small variations in the waveform or spectrographic displays:

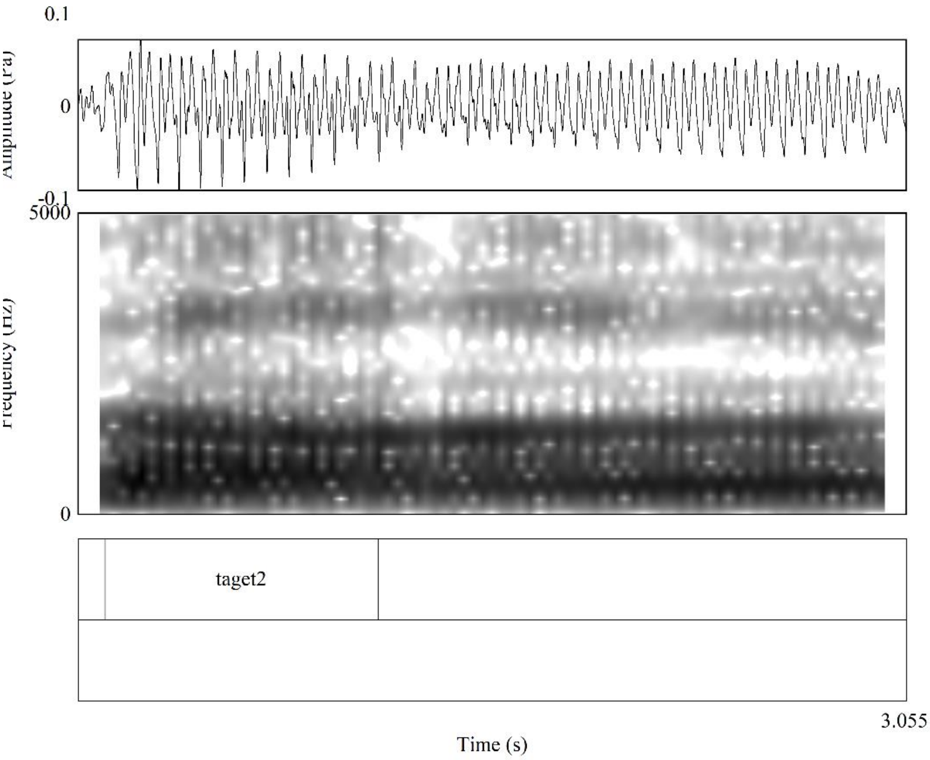

Figure 8. Example of an annotation boundary between a vowel (marked with 'taget2') and a following 'soft d'. There is a small change in the waveform and spectrogram at the segmentation line.

The Praat annotations were saved as TextGrid files. As outlined in section 2.3.4 the relevant acoustic measurements for laryngealisation in this study are mainly f0, intensity, H1-H2 and CPP, particularly the former two as they have been found to correlate most strongly with the stød (Fischer-Jørgensen 1989). Other measurements potentially relevant to the production of the stød are Harmonics-to-Noise Ratio and, if diplophonia is present, Subharmonic-to-Harmonic Ratio. As such, these acoustic measurements are analysed as a proxy of voice quality changes due to changes in laryngeal activity. How these measurements are expected correlate with the stød is outlined in the table below:

Measurement	**Description**	**Expected outcome**
Fundamental frequency (f0)	The lowest frequency of a period signal, perceptually what is perceived as pitch	The f0 is expected to be lower on tokens with regular stød compared to tokens without stød, particularly

		towards syllable termination. For the tokens with tonal stød, f0 is expected to fall gradually throughout the syllable, particularly for the Aarhusian speakers
Intensity	Measure of the amplitude/loudness of sound(s) in dB	The intensity is expected to fall during the realisation of the stød in the regular tokens and fall more steadily and slowly throughout the syllable for the tonal stød in the Aarhusian speakers.
H1-H2	The strength relationship between the first two harmonics of a signal which reflects whether the amplitude of the f0 or the first overtone is most dominant in the voice signal	The second harmonic (H2) is expected to be more dominant in the tokens with regular stød compared to non-stød. For the Aarhusian speakers, the tonal stød is expected to have a more dominant H1 rather than H2
Cepstral Peak Prominence (CPP)	Derived from the ceptrum. The cepstral peak indicates the periodicity of a signal via irregularities that translate into inter-harmonics energy which in turn increases noise and reduces the CPP	The Cepstral peak is expected to be lower in tokens with regular stød compared to non-stød in both dialects and tonal tokens for the Aarhus dialect

Harmonics-to-Noise Ratio (HNR)	Measure of the degree of aperiodicity and turbulence of glottal airflow in dB relative to the harmonic component of the signal	Laryngealisation should yield lower HNR values, meaning regular stød tokens are expected to have lower HNR values than non-stød and tonal stød tokens.
Subharmonic-to-Harmonic Ratio (SHR)	Characterises sound(s) with alternating pulse cycles/period-doubling, expressed as a gross error rate (GER) reflecting pitch doubling and pitch halving, defined as when the estimated f0 value is 20% higher or lower than the reference f0 value. For voice quality purposes, it is calculated as an SHR percentage/distribution.	If period doubling occurs on tokens with stød, possibly indicating diplophonia, the SHR percentage is expected to be higher.

Table 1. Acoustic measurements used and their expected outcomes for the stød types.

To obtain these measurements, sound files were imported into the software VoiceSauce (Shue 2010; Shue et al 2011). This is a program developed for acoustic analysis of voice quality which automatically obtains selected parameters from a WAV file with accompanying TextGrid information. The parameters outlined above were selected as output. This yielded VoiceSauce parameter estimation of the following 6 measures: H1-H2, CPP, HNR, SHR, intensity and f0. H1-H2 can be measured either corrected or uncorrected; in this case the corrected measurement option was chosen as this normalises the amplitude effects of different formant frequencies using an algorithm by Iseli et al (2006; 2007). This reduces the effect of the filter as nearby formants boost adjacent harmonics which should be avoided unless the tokens have the exact same filter configurations which is almost never the case for natural speech data. For the HNR, there are options to specify the HNR range measured. To exclude noise from higher-frequency segments such as fricatives the parameter HNR05 was selected which considers a range of 0-500 Hz. The f0 and formants were estimated using the standard Praat algorithms (Boersma 1993) implemented via VoiceSauce. The window size was 25 ms with a lower threshold of 3 periods for harmonic estimation and 5 periods for intensity, CPP and HNR which are the default settings in the software. The standard Praat settings track f0 in a range from 40 Hz to 500 Hz but these might not be optimal for all data, particularly if the data includes mixed sex speakers. Vogel et al (2009) did a study comparing some generic f0 settings to individualised ranges set for each speaker and found that for certain generic ranges these did not yield statistically different results from speaker-idiosyncratic settings, indicating that individual settings for each speaker is not necessary for a satisfactory result. These ranges did, however, need to be specified for speaker sex. For females the optimal generic range appeared to be 100-300 Hz and for males it was found to be 70-250 Hz (Vogel et al 2009). This is a good indication of what might be appropriate settings for modal voice quality – however, the study did not include any non-modal voice quality. As such, it would be reasonable to assume that for females, the range should be slightly lower to capture any prototypical creaky voice which has been found to lower the f0 below 100 Hz (Laver 1980: 122, Davidson 2019: 238-239). Further, Vogel et al's (2009) study did not include speakers with different languages and thus does not address whether these generic settings would be optimal for other languages. A study by Johnson (2006) gives reason to believe that the sex-aggregated tracking ranges are important for some languages, less so for others. The study found language-specific differences in how much men and women varied in their comparative F1, F2 and F3 values, corrected for overall biological information such as the average height by sex in the respective countries where the speakers were from. Johnson found that variation was substantial from language to language. Interestingly for the current data, Johnson also found that out of the included languages, Danish was the language with the smallest difference between male and female formants (Johnson 2006: 486). Assuming these trends are reflected in f0 too, sex-specific settings for f0 tracking do not appear imperative for the current study. To test this empirically, however, f0 tracking was done first on all the data with the standard Praat settings of 40-500 Hz. Then two other settings were applied, 70-250 Hz for the male speakers and 70-300 Hz for the female speakers, to see if f0 was tracked better with speaker-generic, sex-specific settings.

The 70 Hz lower threshold was applied for females to allow the algorithm to capture frequencies below 100 Hz for creaky voice as it has been found that instances of prototypical creaky voice will have an f0 value below 100 Hz (Laver 1980: 122, Davidson 2019: 238-239). The threshold was kept the same for male and female speakers as their creaky voice fundamental frequency has been found not to differ noticeably on average (Blomgren et al 1998).

To get a rough idea about differences in how the data was tracked using different settings, Table 2 below shows each acoustic measure with a calculation of the percentage of tokens containing rows with no values of 0 in the raw data, assuming a 0 is equivalent to untracked data. This means Table 2 expresses the tracked data as a percentage of tokens without any 0 throughout its trajectory over time. This is a rough estimate as any 0 present does not mean the token is untracked in its entirety, but it gives an initial idea about the differences when using different settings on the same data.

Sex and range settings	**Data tracked based on f0**	**Data tracked based on CPP**	**Data tracked based on Energy**	**Data tracked based on H1H2**	**Data tracked based on HNR**	**Data tracked based on SHR**
Female 40-500 Hz	64.3%	93.7%	95.4%	43.9%	93.8%	1.7%
Female 70-300 Hz	55.5%	32.2%	55.5%	8.5%	32.3%	1.7%
Male 40-500 Hz	60.0%	99.8%	99.7%	1.2%	99.8%	0.4%
Male 70-250 Hz	60.5%	37.7%	60.5%	0.0%	37.7%	0.7%

Table 2. Percentage of tokens tracked with no 0 values throughout their trajectory expressed as a percentage of total tokens for each acoustic measure, aggregated by speaker sex and tracking range settings for f0.

The generic settings from Vogel et al (2009) with a lower end threshold of 70 Hz produced poor tracking outputs compared to the standard VoiceSauce settings with a lower end threshold of 40 Hz. As Table 2 shows this is particularly noticeable for intensity, CPP and HNR. For female speakers, f0 also differs with approximately 10% less tokens containing untracked data with a lower threshold of 40 Hz. This may be explained by findings of very low f0 values in female speakers by Yuasa (2010: 324) who mentions observing ranges between 52.93 Hz and 83.36 Hz during creaky voice, meaning a threshold of 70 Hz would not capture all instances of prototypical creaky voice, even in female speakers. It is not well-established in the literature what a minimum f0 threshold should be to capture all types of creaky voice but the threshold of 40 Hz that is standard in the VoiceSauce algorithm seems reasonable based on Yuasa's (2010: 324) findings for females reported just above and is arguably suitable for males too based on e.g. Keating et al (2015: 1-2) who report a male speaker having an average f0 of 70 Hz during prototypical creaky voice and the finding that male and female creaky voice production has similar values in Herz (Blomgren et al 1998).

A pertinent question in regard to f0 tracking and the stød is whether the presence of 0s, i.e. untracked portions, during the trajectories is a consequence of irregular vibrations during stød production, meaning a finding related to the research in itself, or simply due to the algorithm not tracking f0 properly despite the presence of regular vocal fold vibrations. Given the amount of tokens, it is not possible to manually check each one to verify this so other methods were engaged that allowed the discovery of some general trends in the data.

If the tracking is affected by whether a token was produced with or without stød it is reasonable to assume that the presence of untracked portions are related to the production of stød – this assumption is motivated and supported by findings of approximately 70% of tokens with stød having irregular vibrations present in previously reported Copenhagen speaker data (Fischer-Jørgensen 1989) as irregularity in voicing patterns may cause the signal to become aperiodic and thus not trackable by the fundamental frequency. Conversely, if the tracking is unaffected by the presence or absence of the stød and simply reflects the algorithm failing to track for reasons not related to laryngealisation, it is expected that the amount of tracked data does not differ noticeably between tokens with and without stød.

With this in mind Table 3 gives an initial impression of whether the f0 tracking is affected by the stød or not:

Speaker sex	**Dialect**	**Non-stød fully tracked (%)**	**Regular stød fully tracked (%)**
Female	Aarhus	86.9%	65.1%
Female	Copenhagen	82.9%	33.7%
Male	Aarhus	76.0%	64.7%
Male	Copenhagen	72.2%	26.8%

Table 3. Percentage of tokens tracked with no 0 values throughout their trajectory expressed as a percentage of total tokens for the presence vs absence of stød, aggregated by speaker sex and dialect.

Table 3 shows that non-stød tokens do have significantly more f0 trajectories that are fully tracked compared to regular stød tokens for both dialect and speaker sex groups. This indicates that the tracking is indeed affected by the presence of the stød and thus reflects irregular vibrations during its production. This account was very broad, however, as any 0 present in a trajectory categorised the entire token not fully tracked in Table 3's percentages.

The calculations of tracking so far have been done on the raw data output extracted from VoiceSauce to test the difference between settings in tracking performance. The raw output contains an unequal number of rows per token because the algorithm takes a measurement of a variable every millisecond and creates a new row per measurement point. Thus, tokens with different durations will have an unequal number of rows. For testing the best range of VoiceSauce settings as was done above, the raw data is optimal. However, for running the statistical models, having an unequal trajectory length between tokens being compared is not optimal, more on this below. It has now been established that the standard VoiceSauce settings perform considerably better than the elsewhere recommended f0 settings (Vogel et al 2009) for modal voice with sex differences on tracking the current data. However, even with the standard f0 range settings, not all data is tracked and it appears that the stød is affecting the tracking of f0 which is unsurprising. A pertinent question now becomes how the remaining acoustic measurements are related to f0 tracking. The measures will be considered in turn below.

CPP is often used as a measure to quantify dysphonia in speakers with transient voiceless gaps in voiced speech (Heman-Ackah et al 2003; Awan et al 2010) because it does not rely on accurate pitch tracking due to being cepstral-based, calculated via a Fourier transform of the power spectrum (Patel et al 2018: 896). Further, it can be calculated on continuous speech rather than just sustained vowels (Parsa & Jamieson 2001; Watts & Awan 2015; Lowell 2012). This is ideal for the stød as the stød basis can be either a long vowel or a short vowel combined with a sonorant consonant, thus not always only a vowel and in faster speech, not sustained for very long. While not totally dependent on f0, CPP can still be influenced by f0 and intensity as well as the phonemes it is calculated on (Heller Murray et al 2022). The phonemes are controlled for by eliciting minimal pairs with the same phoneme with and without stød in the word list design.

For intensity, the VoiceSauce algorithm used to calculate it relies on the Root Mean Square (RMS) Energy which is equivalent to intensity but calculated at every frame over a variable window equal to five pitch pulses which normalises the intensity with f0 to reduce the correlation between the two. This means that despite some discrepancies in the amount of data tracked based on f0 and intensity as shown in Table 2 above, because intensity is calculated as RMS Energy and thus less correlated with f0, intensity is considered a viable measure to analyse in combination with other acoustic measures. F0 and intensity are not controlled for in the word list design as the need to elicit less formal speech was weighted high and because the stød and non-stød words are often homographs, meaning sentence context cannot be kept constant, i.e. via carrier phrases, as context is vital for participants to know which word to produce. The microphone distance was also not controlled well as the participants had to fit their own lapel mic to their clothes and it is possible that it might have moved slightly during the recording, despite being fixed to the chest. However, the statistical models later fitted are informed that the f0 and intensity trajectories come from separate tokens and speakers, meaning the model takes this into account in the statistical comparison, somewhat controlling for the differences due to sentence intonation. It is also worth noting that the current study measures relative changes between minimal pairs rather than absolute values of high/low intensity, making a highly controlled setup less pertinent, although worth bearing in mind when interpreting the results.

For H1-H2, the influence of higher formants like F1 and F2 is controlled in the VoiceSauce algorithm by choosing to extract the corrected measures rather than the raw values from the output spectrum. The relationship between f0 and H1-H2 is less straightforward. A study by Holmberg et al (1989) tested the correlation between pitch and glottal airflow, assuming various laryngeal settings impact airflow, e.g. increasing it by having a long closing phase of each vibratory cycle. The authors did not find a strong correlation between f0 and any glottal parameters tested such as H1-H2. Another study, however, used corrected H1-H2 values and separated the participants based on gender and age. This study found that H1-H2 increases as f0 increases when f0 values are lower than 175 Hz (Iseli et al 2006). This correlation was tested in American English speakers only and it is not known whether this generalises across languages. It can be noted for now, however, that there may be a relationship between corrected H1-H2 and f0 which should be kept in mind when interpreting the results of the analysis presented below.

For HNR, VoiceSauce uses an algorithm developed and described in de Krom (1993) which is cepstrum-based and found by liftering the pitch component of the cepstrum and comparing the energy of the harmonics with the noise floor in variable window length equivalent to five pitch periods. Because HNR is cepstral-based, it is less dependent on fo and thus less affected by fo tracking. The noise component can be influenced by the presence of other sources of noise in the signal but since the study uses a comparative method of minimal pairs, the noise components not strictly related to voice quality changes are relatively well controlled for. Further, HNR was set to be extracted at 0-500 Hz, thus filtering out higher-frequency noise components that may skew the ratio of noise and harmonics.

For SHR the subharmonics can be alternating oscillatory cycles that can affect both amplitude and f0, in the latter case doubling the fundamental period (Herbst 2021: 365) which is clearly strongly related to f0. The algorithm for SHR detection in VoiceSauce uses both a logarithmic frequency scale and spectrum shifting technique to find the modulation of the oscillatory cycles (Sun 2002, Herbst 2021), meaning both amplitude and frequency modulations will be tracked and extracted, theoretically meaning even if f0 tracking fails, some tokens could still have an associated SHR due to calculations on the spectrum shift rather than the frequency scale. It seems clear, however, that the performance of the algorithm is dependent on parameter settings for f0 (Herbst 2021: 371) but it is not clear if, within these settings as boundaries, untracked portions of f0 affect the tracking of SHR too. Presumably, the two are related to a higher degree than some of the other acoustic parameters, rendering f0 tracking a potential issue for the SHR measurements.

Table 2 generally shows that SHR was poorly estimated along with H1-H2 but it should be kept in mind that Table 2 was a very broad account of the tracking performance based on any 0 in the raw data trajectories. To get a more detailed and dynamic view of the distribution of tracked data across parameters, the final parameter estimation values were written to a .txt file based on selected tiers from the TextGrids and then imported into R Studio (R Core Team 2022) for visual exploration and statistical analysis.

As mentioned earlier, the statistical analysis requires the data to be time normalised to get an equal number of observations for each token. The raw data was time normalised by smoothing each token's trajectory to 11 equidistant time points from 0-10. The following paragraph aims to give a better overview of how the data was tracked using the time normalised data rather than the raw data containing one observation per millisecond. There are two reasons for this: 1) time normalisation reduces the number of observations, and thus also the number of 0s, to a relative number within the token, essentially making it easier to gauge where the larger gaps are and 2) the data analysis should be interpreted with transparency in where the tracking gaps are in the form in which they appear in the analysed data, i.e. the time normalised form. An alternative to this approach would be to completely exclude data with tracking gaps, but since parts of the untracked data appear to be linked to the presence of stød (cf. Table 3) excluding this data would mean excluding an important finding when looking at trends in the data. Further, there is no golden standard for when a token should be excluded or included based on the percentage of untracked data and thus no precedence for making informed decisions about a minimum threshold of tracking that would be reasonable. For this reason, the approach taken in the remaining acoustic analysis is to include all the data (except those excluded tokens mentioned in section 3.2) but to report the distribution of tracked data within each acoustic variable for transparency when interpreting the results of the analysis.

Before making distribution plots to show the amount of tracking for each acoustic parameter, all the 0s in the data were converted to NAs. This is important because the GAMMs fitted later can handle NAs when fitting trajectory smooths and also because the data is normalised in the steps below – if the 0s are not converted to NAs, the z-scoring method for normalisation will treat the 0s as data points and transform them, making them look like genuine values rather than an expression of lack of tracking.

For the distribution graphs below, the number of NAs in any token were calculated as a percentage of total observations within the token and plotted in a density plot aggregated by stød type and speaker dialect for each acoustic parameter. The y-axis in the density plots was computed by the *ggplot_build* function in R to extract the density values and determine the maximum density value, using this reference for an appropriate scale on the y-axis.

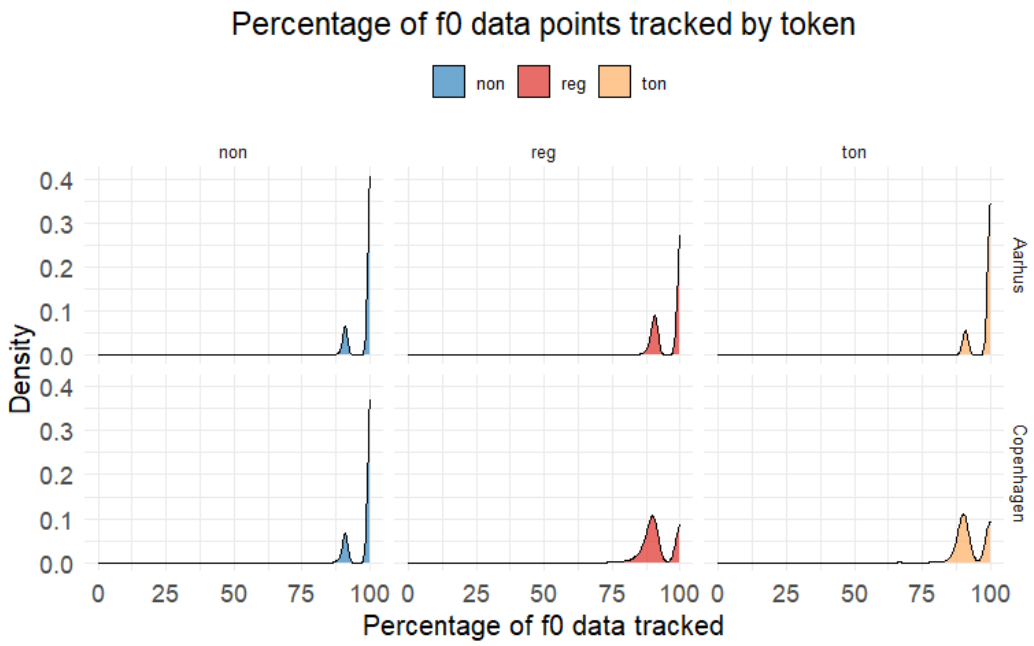

Figure 9. Density plot of percentage of data points tracked per token for f0 aggregated by speaker dialect and stød type.

Figure 9 shows that the f0 tokens have density clusters at two major points on the x-axis, at around 90% and 100%. The Copenhagen data clusters at lower densities for regular and tonal stød types, being tracked slightly worse than the Aarhus speech data's stød tokens and showing more variability around 75-90%. However, even the least tracked condition, the regular Copenhagen stød, clusters at above 75% tracked data points in each token and the lacking 75% appears to be a result of the stød based on Table 3. In summary, some tokens have missing data points, seemingly connected to stød type and dialect which is interesting, but overall f0 is tracked reasonably well and in a comparable amount to previous work on Copenhagen stød (Fischer-Jørgensen 1989).

For intensity, CPP and HNR, all tokens had a tracking rate of 100% in the time normalised data, meaning density plots could not be created as the distribution was completely uniform. For purposes of demonstration, Figure 10, Figure 11 and Figure 12. below instead show bar plots for intensity, CPP and HNR, showing that all stød conditions in all dialects cluster at 100% on the x-axis:

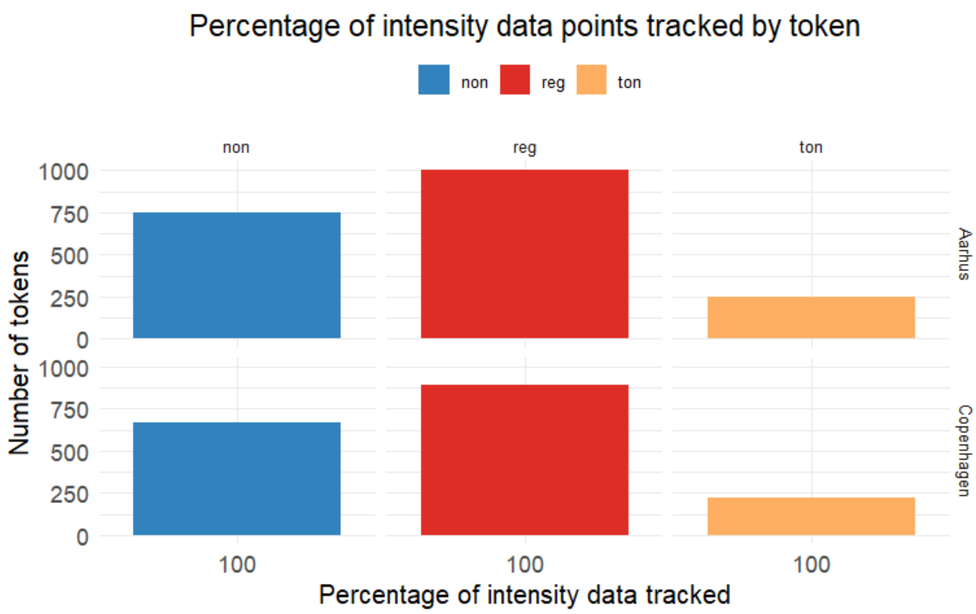

Figure 10. Bar plot of percentage of data points tracked per token for intensity aggregated by speaker dialect and stød type.

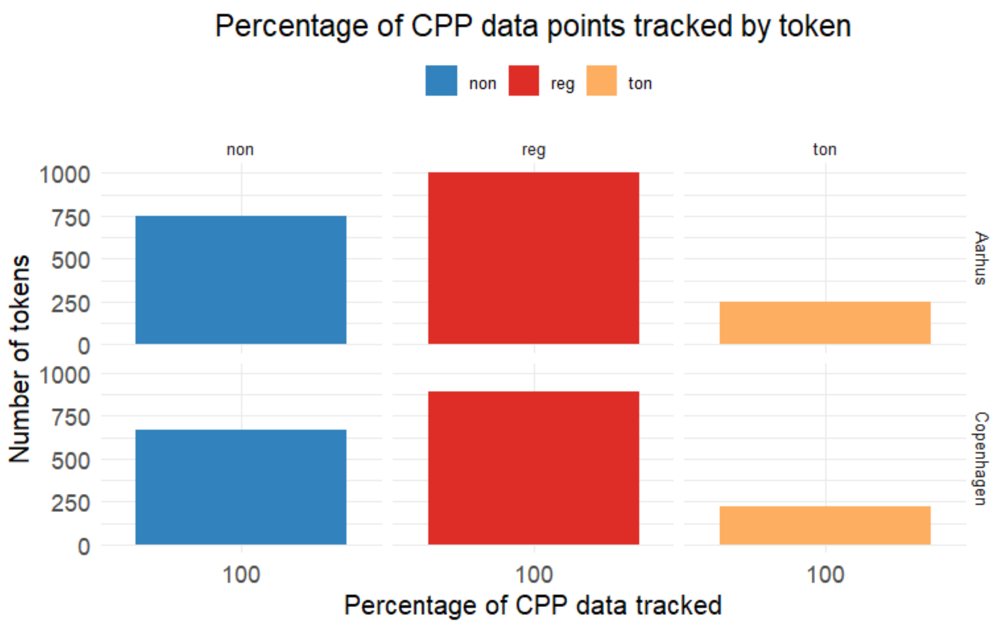

Figure 11. Bar plot of percentage of data points tracked per token for CPP aggregated by speaker dialect and stød type.

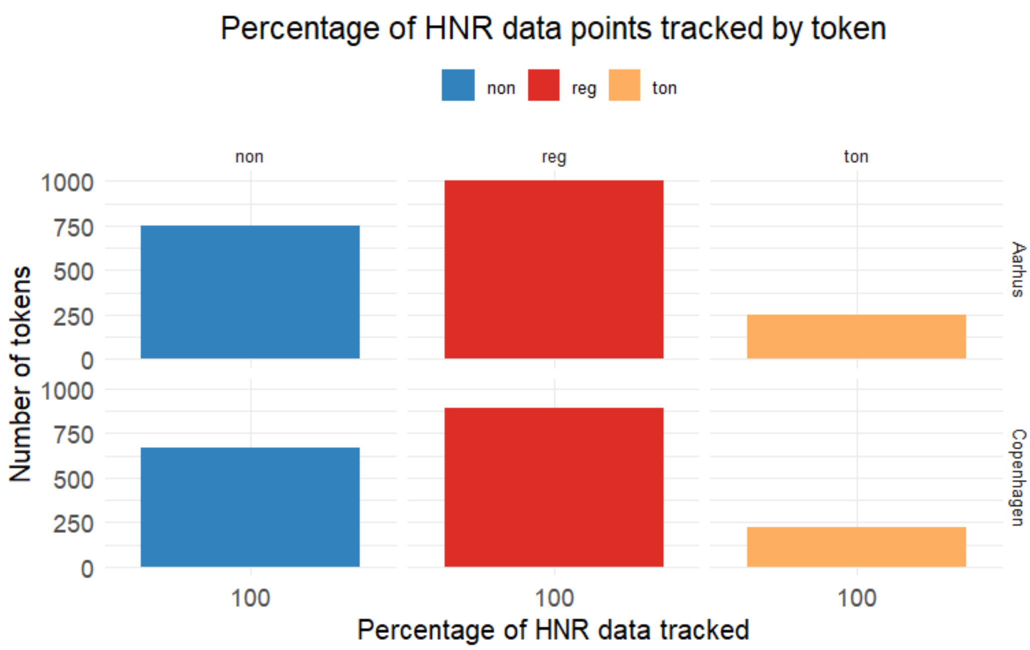

Figure 12. Bar plot of percentage of data points tracked per token for HNR aggregated by speaker dialect and stød type.

For H1-H2, Figure 13 below shows that most tokens had highest density clusters around 90% and smaller clusters at 100%. This means most tokens were reasonably well tracked but with one data point missing in many of the trajectories for both dialects. As the models fitted below can smooth over NAs when fitting splines, this is not considered a major issue for the integrity of the analysis but it is worth noting that H1-H2 was slightly worse tracked than the other measures so far. Based on Table 2, a large majority of these data points are caused by some NAs in the trajectories of the male speakers in particular.

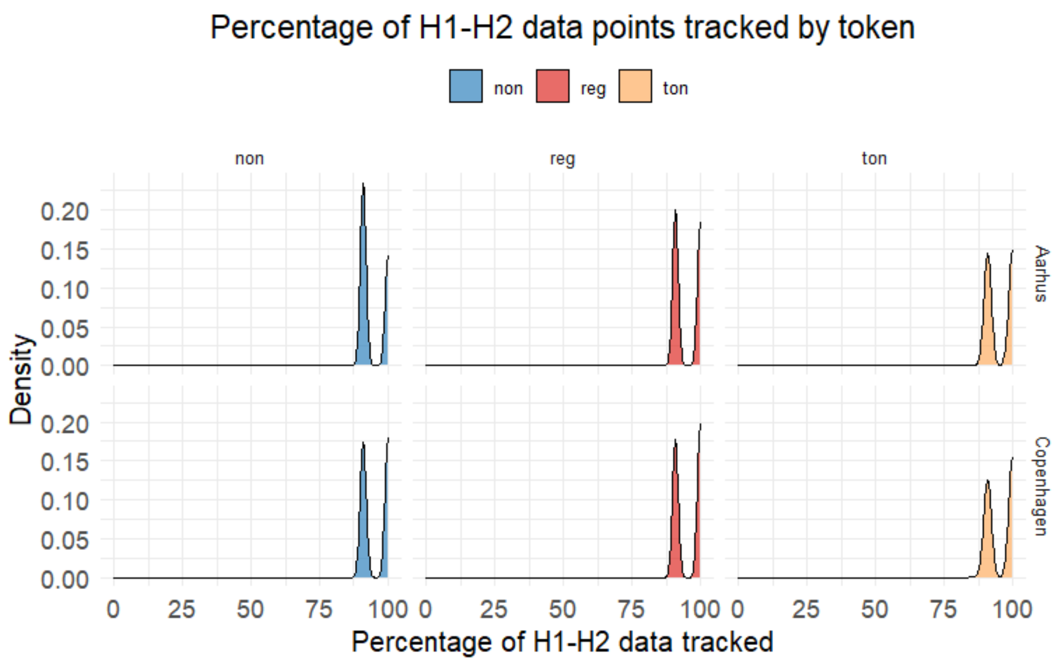

Figure 13. Density plot of percentage of data points tracked per token for H1-H2 aggregated by speaker dialect and stød type.

SHR was by far the poorest tracked acoustic parameter, as shown in Figure 14 below:

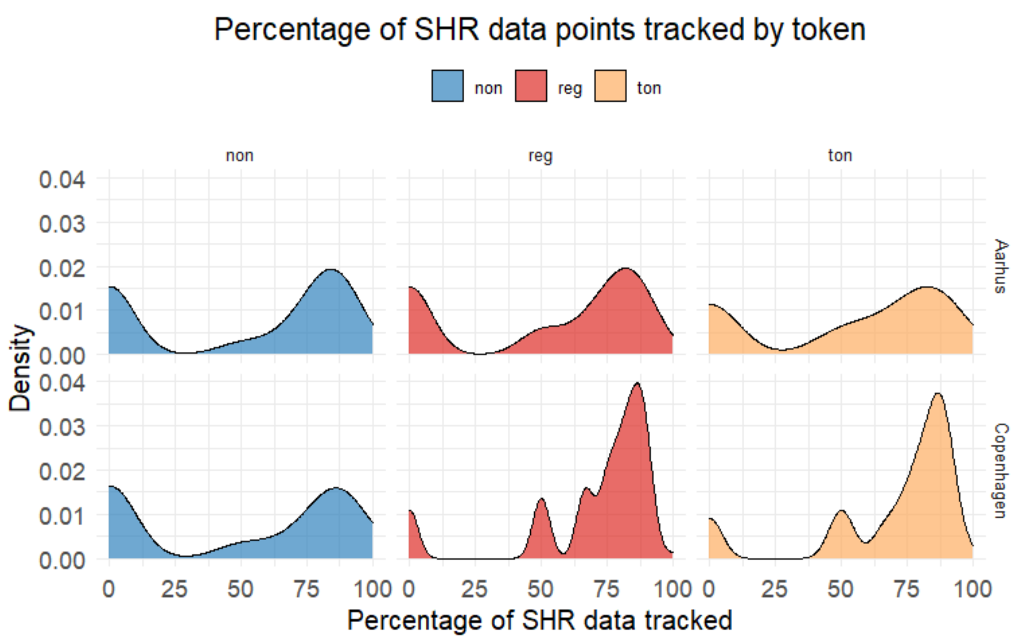

Figure 14. Density plot of percentage of data points tracked per token for SHR aggregated by speaker dialect and stød type.

The clusters are highly variable on both axes and, unlike the previous parameters, have clusters at 0% and no peaks at 100%. For both dialects, the two stød conditions have the highest cluster peaks above 75%, but higher for Copenhagen compared to Aarhus, indicating that the Copenhagen speakers' SHR data was better tracked for the two stød types. For both dialects, the non-stød tokens were tracked the poorest with a low density peak above 75% similar in height to the cluster peak at 0%, meaning just as many tokens were not tracked at all as were tracked well.

In summary, most acoustic parameters were tracked well with the chosen VoiceSauce settings. Missing data points occurred in parameters f0, H1-H2 and SHR. Interestingly, it was noted above that H1-H2 and SHR are the two acoustic parameters most dependent on accurate f0 tracking and indeed this correlation seem to be confirmed by the trends in the data tracking. However, it is not clear just why SHR tracking is considerably worse than f0 tracking here or why there seems to be dialectal and stød type differences in the tracking performance. In the analysis below, SHR will be considered significantly less reliable compared to the other acoustic parameters due to the smaller pool of data points tracked.

3.3.2. Statistical analysis

An initial analysis on pilot data from 4 speakers was initially carried out while waiting for labs to be available after a long period of lockdown due to the Coronavirus pandemic. A steady-state analysis based on stød basis interval means and midpoints was explored for the pilot data as a first step but it quickly became clear that this type of analysis did not capture the trends in the data very well. This is due to the dynamic nature of the intervals with regular stød – for a measure like f0 they tend to start high and then fall towards syllable termination, making a midpoint or mean value too reductionist as the high start and low end would essentially cancel each other out and skew the results toward insignificance. It is probably partly for this reason Fischer-Jørgensen (1989) divided the stød syllables into two phases in her study and analysed each phase separately. However, we now have more dynamic tools available to analyse full trajectories rather than static midpoints or averages which eradicates the need to divide the syllable for the present purposes.

An option for modelling dynamic trends that has been growing in popularity in recent years is Generalised Additive Mixed Modelling, or GAMMs, which fit so-called smooth terms to the data along with parametric terms, allowing for more 'wiggliness' in the model fit than traditional linear models (Wood 2006). GAMMs are an extension of Generalised Mixed Models (GAMs) which are models made up of two basic constituents: a number of so-called basis functions and a smoothing parameter. The number of basis functions determines the balance between under- and overfitting and should thus be neither too low nor too high. The smoothing parameter estimates the coefficients for the individual basis functions to control how much the fitted curve is allowed to 'wiggle', i.e. the degree of permitted non-linearity in the model. What GAMMs additionally do that GAMs do not is to include random smooths and/or error model specifications which can capture random effects alongside smooth terms and parametric terms. Random effects assume that the levels of a mixed model come from a larger sample which, although being independent estimates, are related to each other. An example is the effect of speaker – speaker A's tokens are more likely to be similar to each other compared to tokens spoken by speaker B and the model should thus be informed that a group of tokens is uttered by the same speaker by including speaker as a random effect. This is contrary to the influence of a fixed effect which is assumed to be predictable and lack idiosyncrasy. Random effects thus inform the model about dependencies between data points and that these should not be regarded as completely singular measurements (Winter 2019: 234-236). We can now appreciate how including random effects when modelling the current data is important because we have 10 measurement points for each token and if we do not include random smooths and/or error models, the model fit will ignore any grouping effect or temporal structure in the data, i.e. assume that the 10 measurements from each token are independent rather than taken from the same trajectory. As a consequence, estimates of significance will be overly confident as the model thinks there are more independent data points informing the fit than is actually the case and the parameter estimation will be highly anti-conservative.

The procedure used for fitting GAMMs in the following is largely based on that outlined in Sóskuthy (2017). First, however, each acoustic measure was Lobanov z-scored (Lobanov 1971). This adjusts the means and standard deviations to a common scale for all speakers which makes it easier to compare features across various speakers. Specifically, the goal of z-scoring was to minimise effects due to inter-speaker anatomical variation to help isolate the comparative effects of the stød on the acoustic parameters. Lobanov z-scoring was preferrable for this data because the focus of the research is on articulation and acoustics rather than the interaction with perceptual aspects, the latter typically requiring other types of normalisation such as log transformations to accommodate the innate frequency normalisation done by the inner ear. Further, the Lobanov z-scoring method has been found to be the most successful normalisation technique for language variation research (Adank et al 2004) and it does not assume a particular shape of the distribution of vowels (Lobanov

1971). A caveat is that Lobanov z-scoring is usually applied to vowels only and research on its effectiveness in language variation is tested on normalising vowel spaces between speakers. However, since the stød basis *always* contains a vowel, sometimes combined with a consonant, the method is still thought useful for the present data.

Before fitting any complex models it is useful to first simply establish whether there is a significant effect of stød type on the syllable trajectories for each acoustic measurement - only if this is the case is it meaningful to start exploring the exact nature of these effects. To demonstrate the method for doing this, a model comparison is given below for the f0 measurement - the steps involved will be the same for all acoustic measurements. As the stød was elicited in two different minimal pair contrasts (no stød vs regular stød and regular stød vs tonal stød) the data was first subsetted into separate data frames for each dialect accordingly to allow for separate model comparisons and then into subsets for each stød type. The example given here was created on the Copenhagen dialect non-stød vs regular stød subset, fitting GAMMs using the *mgcv::bam* function in R (R Core Team 2022).

To test whether the presence of stød has an overall effect on the f0 trajectory, a GAMM was fitted with stød type as a function of the normalised f0 values. Smooth terms included a reference smooth over the 11 equidistant timepoints with no grouping specification and a difference smooth grouping the time points by stød category (no stød vs regular stød). Random smooths were included by speaker and by word, respectively.

In the model summary, both the reference smooth and the difference smooth had *p*-values < .05, suggesting that the trajectories for the non-stød and the regular stød are significantly different. To confirm this a likelihood ratio test was executed by creating a nested GAMM without the difference smooth to compare the best model fit; a comparison that can be made using the *itsadug::compareML()* function. The model comparison summary table confirmed that including the difference smooth significantly improved the model fit, confirming the hypothesis that the presence of the regular stød in Copenhagen speakers significantly changes the trajectory of f0 compared to tokens without stød.

One important thing missing in the GAMMs fitted so far is error model specifications. Recall from above that these are implemented to avoid anti-conservative estimates caused by the model assuming the 10 equidistant timepoints are independent data points rather than taken along the same trajectory. The *mgcv::bam* function allows one type of error model, an autoregressive model with lag 1 (AR1). This model assumes that the errors for adjacent observations are correlated, i.e. the error at lag 3 is determined by the error at lag 2 and so on (Sóskuthy 2017: 30). The degree of correlation between errors needs to be manually specified so in order to get a rough estimate the function *itsadug::start_value_rho()* can be used. This calculates autocorrelation in the residuals at lag 1. Fitting a model with this error specification massively improved the model fit compared to the full model without an error specification and the nested model without the difference smooth. The model fits were compared by calculating the Akaike Information Criterion (AIC) for each model. The AIC score combines and penalises two quantities; poor model fits and unnecessary model complexity, expressed as the model with the lowest relative AIC being the model with the best fit. However, the autocorrelation calculated at lag 1 is a rough estimate and not necessarily the most suitable value for the error model. Bearing this in mind a few different rho values were tested to check the best model fit. The lag 1 estimate calculated the rho at 0.5 which was compared to models with rho set at 0.3, 0.4, 0.6, 0.7 and 0.8. The AIC steadily decreased as the value went up between 0.3-0.6, reaching the lowest at 0.7 and then increasing again with rho at 0.8. Based on this it appears that an AR1 error model with rho at 0.7 best captures the f0 trajectories for the Copenhagen speakers' non-stød vs regular stød subset without adding unnecessary complexity to the model. More complex interactions can be modelled to improve the model fit if there are still significant patterns left in the residuals. A plot of the autocorrelation residuals for this GAMM is printed below to explore whether more terms should be included:

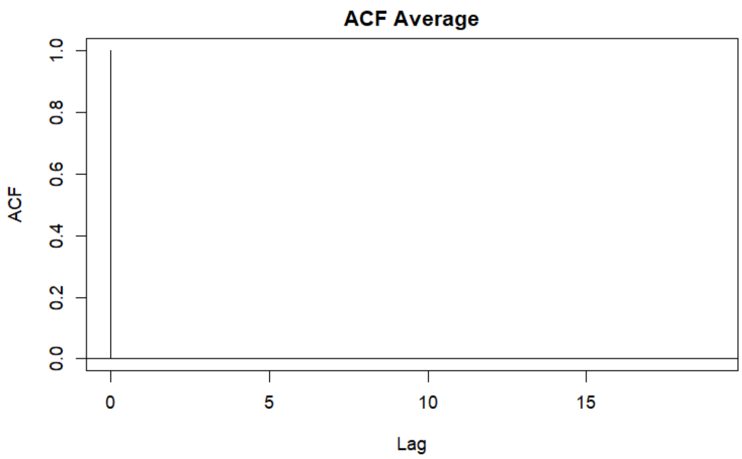

Figure 15. Autocorrelation plot of the residuals for the GAMM fit of the effect of f0 on normalised time by stød type.

The plot shows some negative autocorrelation in the residuals at lag 1 which is unavoidable but otherwise appear to predict the data extremely well reflected in no lag in the residuals at the remaining points on the x-axis. Adding more complexity to the model to improve the model fit is thus unnecessary and may even be penalised which would be evident in a recomputed AIC score.

The next step was to explore the nature of the difference as the model summary of the GAMM with the difference smooth does not tell us exactly how the trajectories differ, just that they are significantly different in some unspecified manner. An important part of interpreting GAMMs is therefore data visualisation. The plotted GAMMs will be printed and explored in the analysis starting in section 4. Before this, however, the methodology is described for the EGG portion of the data collection.

3.4. EGG data analysis methods

As described to in section 2.1.3., the Laryngeal Articulator Model (LAM) posits that there is more to the articulation of phonation types than the vocal fold contact patterns, or, in structural terms, the aperture between the arytenoid cartilages. Other structures are involved as shown by more recent methods of investigating phonation in ways that directly visualises the larynx (e.g. Esling et al 1998; Esling 1999; Esling 2005, Esling et al 2019). However, as direct visualisation requires specialised equipment and technique, EGG is a reasonable alternative that offers less detail but allows for easier recording, and thus more participants, and lends itself well to quantitative and dynamic analysis of phonation coupled with acoustic analysis. EGG as a method was described in section 2.3.5 above and thus this section will not contain this part but instead move on to describe how the data was analysed.

3.4.1. Data processing

The EGG signal was recorded on a separate channel from the audio channel which was extracted after the data had been annotated in Praat (Boersma & Weenink 2021) so that the EGG segments matched the TextGrids and both signals could be annotated simultaneously. The EGG files were processed in Praat using two automated scripts[1]. In order to calculate the Contact Quotient, one script extracted the first derivative of the EGG signal, the dEGG, as the derivative more accurately corresponds to the vocal fold contact patterns (Childers & Krishnamurthy 1985; Heinrich et al 2004). To visualise why, the EGG and the dEGG signals are compared below:

[1] The Praat scripts were very kindly written by Dr. Stefano Coretta, Edinburgh University.

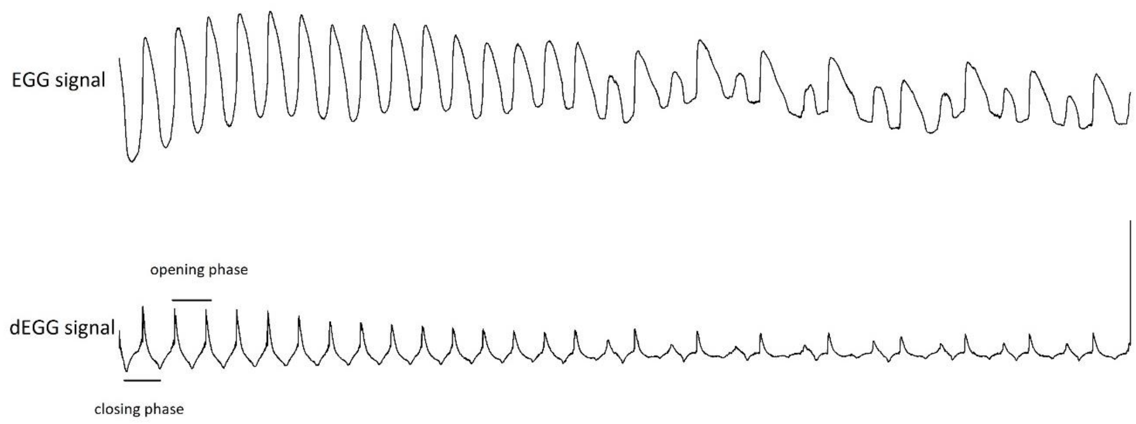

Figure 16. EGG (top) and dEGG (bottom) signal on the stød basis (165 ms) of the word 'byen' from speaker MCPH1.

As evident the dEGG signal on the bottom contains well-defined peaks and troughs corresponding to vocal fold contact patterns where a peak equals maximum contact and a trough means no contact. From this signal the Contact Quotient can be calculated as closing/contacting and opening/de-contacting intervals from the relative dEGG maximum and minimum in each glottal cycle. As such, the CQ is an estimate of the relative duration of the vocal fold contact per vibratory cycle found by subtracting the relative dEGG maximum from the relative dEGG minimum. These scores were obtained using a second script which calculated and wrote dEGG measures to a CSV file for further processing and analysis in R Studio (R Core Team 2022). The dEGG derived CQ scores should all be a relative measure between 0 (open) and 1 (closed) so any outliers outside of these values were filtered out.

There were quite substantial differences in how well the CQ values were tracked per speaker. The Copenhagen dialect had a total of 16359 observations racked whereas the Aarhus dialect had 26969 observations. The Aarhus dialect had 11 speakers recorded whereas Copenhagen only had 10, but even with this in mind, that is an average of 1635 observations per speaker for Copenhagen and 2451 per speaker in Aarhus which is a remarkable difference. The question is whether the lack of tracking is related to the articulation of the stød, and thus in itself a finding in terms of dialectal differences in stød realisation. This could be a plausible explanation given that the stød can make vocal fold vibrations irregular and possibly to an extent where the EGG cannot reliably be tracked as this only works on voiced stretches of speech. To explore this, plots were created for number of observations tracked per speaker segregated by stød type as this explanation would only be viable if the amount of cycles tracked were observably less for the tokens with stød compared to the tokens without stød:

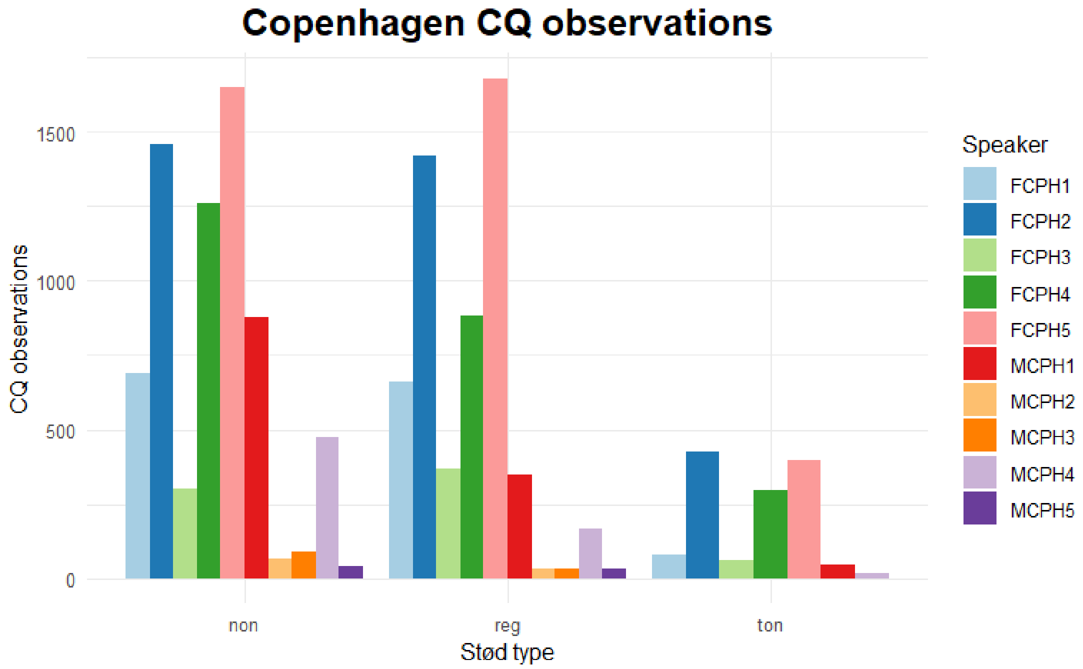

Figure 17. Bar graph showing the CQ observations per speaker for each stød type for the Copenhagen speakers. Non = no stød, reg = regular stød and ton = tonal stød.

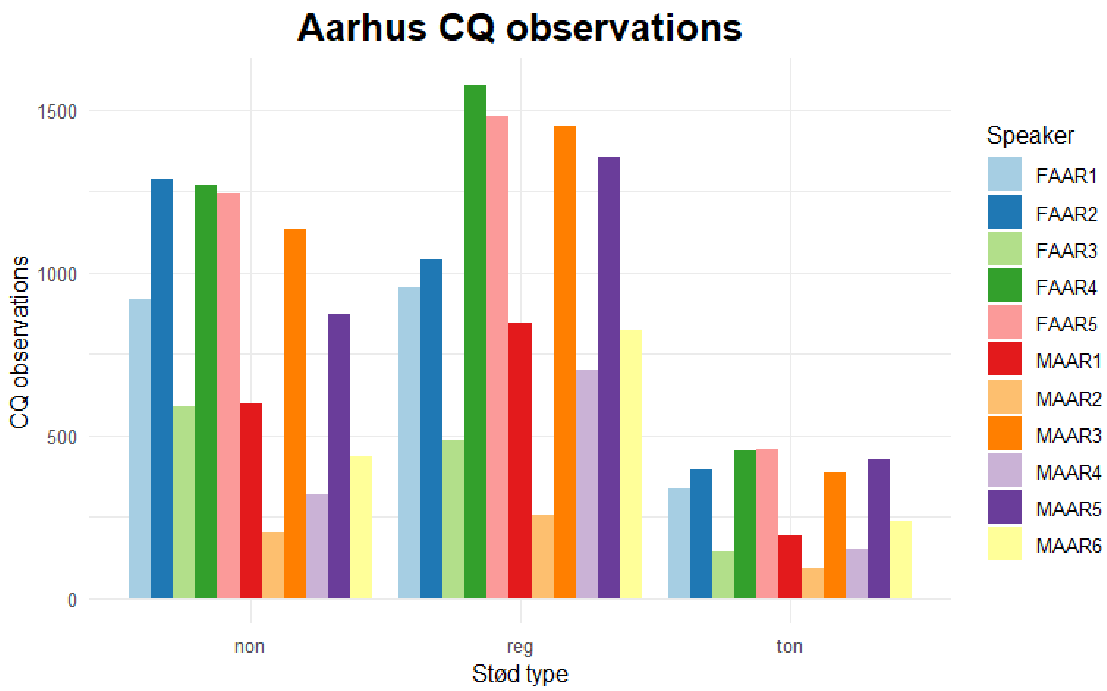

Figure 18. Bar plot showing the CQ observations per speaker for each stød type for the Aarhus speakers. Non = no stød, reg = regular stød and ton = tonal stød.

As the plots show there appears to be much more inter-speaker variation than variation between stød types. This means that the lack of tracking cannot be interpreted as merely a consequence of loss of regular voicing due to production of the stød. The charts also confirm that the Copenhagen speakers are particularly low in glottal cycles tracked compared to the Aarhus speakers and add the observation that the male speakers are the primary contributors to this low count in Copenhagen. The variability in tracking is likely to be a consequence of the circumstance that getting Ethics approval for the study during the Coronavirus pandemic meant not being able to touch the participants to fit the EGG equipment and they had to do this themselves. Further, there was no possibility of checking the EGG signal during the recording session to see if the tracking quality was degrading as the session progressed due to participants sitting inside a sound booth and the data collector sitting outside it. Additionally, some people simply track better than others due to idiosyncratic factors. Some of these factors were controlled for, e.g. ensuring male participants did not have a neck beard as that could lower the detection, while others were not controllable.. For the analysis and results in the coming sections, they should be interpreted with the knowledge that the amount of data for the two dialects differs widely and for some conditions, like the tonal stød, the Copenhagen male speakers contribute very little data. When fitting the GAMMs, the models contain a smooth to inform the models that the tokens come from different speakers and thus controls for inter-speaker variability to some extent, avoiding heavily skewed estimates. However, for comparing the acoustic and EGG data later, it is more problematic that the data sets contain dissimilar tokens for these speakers. This second issue will be dealt with in later sections – first, the EGG is analysed on its own.

3.4.2. Statistical analysis

The statistical analyses for the CQ values were executed fitting GAMMs to the data just as was done for the acoustic measures. This procedure is very appropriate for the EGG data as the CQ measures are extracted as glottal cycles over relative time, making them subject to dynamic rather than static statistical modelling. The CQ measure is a proportion, meaning it is measured as a proportion of each glottal cycle. As a consequence, the time measurement column in the data is the relative time of EGG and DEGG measurements within a glottal cycle. This means the data is already time normalised and the relative time can be used in the GAMM to smooth over instead of manually creating equidistant timepoints as was necessary for the acoustic data. Therefore the EGG data was not further time normalised at this stage of the analysis.

While making absolute judgments on vocal fold contact patterns from the CQ measures on its own can be subject to variability and error (e.g. Hampala et al 2016; Herbst et al 2017), having three different stød conditions to compare using the same methods of glottal cycle measurements allows for a relative comparison of the CQ, eliminating many of the issues associated with having to make absolute inferences about vocal fold contact patterns from just one vibratory condition with no comparative laryngeal states. As such, fitting GAMMs with and without smooths for the relevant parameter under investigation is a favourable method for the purpose of this study. This also allows for processing of a large amount of data – many previous studies have adopted a more qualitative, visual interpretation of individual EGG waveforms as means of analysis (e.g. Dejonckere & Lebacq 1983; Titze 1990) which would severely limit the amount of data feasible for inclusion in a time limited project with just one researcher. Thus, while the level of detail might be less with a quantitative approach, the amount of data is more robust and lends itself well to statistical analysis.

GAMMs were described in depth in section 3.3.2 which can be referred to again for a reminder of what these models are and can do. The description here will thus function as a shorter documentation for how the statistics were obtained for the EGG data as a refresher for the more detailed description in section 3.3.2.

First the data was checked for how the glottal cycles were tracked. As seen in the previous sections, some tokens only had very few glottal cycles tracked. From previous literature on EGG it is not clear what the lowest threshold for number of glottal cycles detected per token would be for it to be viable for analysis. Surprisingly, not even comprehensive overviews of previous studies such as Herbst (2020) make any mention of this issue. Avelino (2010: 277) describes how in their data on Yalálag Zapotec, cycles in a creaky EGG waveform are almost double the duration of modal cycles with five creaky pulses vs eight modal pulses in a 50 ms window. This gives at least some indication about the number of glottal cycles to be expected, albeit a voiced stretch of 50 ms is on the shorter side compared to Hansen (2015) who found a lag in the timing of the compression peak vs the intensity minimum, the former occurring approximately 80-100 ms before the latter. This indicates quite a longer time window than Avelino (2010) but Hansen (2015) used less naturalistic data that was ‘read aloud’, likely decreasing the speaking rate and thus also the duration. For the statistical analysis at this stage all CQ values were included for each dialect as there is no clear guideline as to when and how to exclude tokens with less data tracked from previous literature. However, the differences in tracking depicted in Figure 9 and 10 should be kept in mind when interpreting the GAMMs. The total number of glottal cycles for analysis were 45.301.

As mentioned earlier, the EGG data was not time normalised to equidistant timepoints throughout the duration of each token but rather a variable was created containing the relative time of the dEGG measurements within a glottal cycle for each CQ measurement. The lowest relative timepoint was set as the start event for each token for the AR1 autoregressive error model specification. The data was then split into subsets of both the two dialects and the respective minimal pair conditions for each dialect (non-stød vs stød, regular stød vs tonal stød). Then GAMMs were fitted using the *mgcv::bam function* in R (R Core Team 2022) and following the same procedure as for the acoustic data. For each dialect it was determined whether stød type had an overall effect on the CQ values by fitting a GAMM with CQ as a function of stød type including a reference smooth over relative time with no grouping specification and a difference smooth grouping each relative time trajectory by stød type. Two random smooths by speaker and by word were specified and an AR1 autoregressive error model was added, finding the best rho value by comparing the AIC score for different models and finding the lowest score. The overall effect of stød type on the CQ values was determined via a likelihood ratio test using the *itsadug::compareML()* function to compare a full model to a nested model without the intercept and difference smooth for stød by relative time. If the model summary table confirmed that including the difference smooth significantly improved the model fit that confirmed the hypothesis that the presence of the stød significantly changes the CQ value trajectory compared to tokens without stød or tokens with regular vs tonal stød. If this was the case, another nested model was fitted to test for the specific effect on CQ trajectory shape by excluding the difference smooth for stød by time but including the intercept. If any model comparison yielded a significant result, plots were made of the trajectories to examine the specific way in which the CQ value trajectories change according to which stød type is present, just as was done with the acoustic data.

Having outlined the methods for both the acoustic and articulatory data preparation and analysis, the results are presented in the following.

4. How much dialectal phonetic variation occurs during the Danish stød, a binary phonological contrast?

This chapter presents an analysis of the phonetic variation during the production of the stød. The analysis includes acoustic and articulatory data from two dialects to provide a novel perspective on intra-language phonetic differences in laryngealisation. The reference point for characterising the variation will be the subtypes of laryngealisation outlined in Keating et al (2015). While this is a short conference paper and thus could be argued to be less suitable to base an analytic categorisation on, it is also to my knowledge the only paper that systematically outlines subtypes of laryngealisation from the literature along with their acoustic correlates, referencing them to and with each other. This makes it an extremely useful guideline in categorising phonetic variation in creak-like voice quality, a continuum which the stød is assumed to fit on.

The subtypes are all defined by certain acoustic parameters being high or low. The analysis starts by determining the effect of stød type on each of these parameters for each dialect, then compares that with the EGG data for each dialect, ending with a comparative summary of the most fitting subtype characterisations for each stød type in the respective dialects.

4.1. Acoustic analysis

4.1.1. Copenhagen

As described in the methods section the acoustic data was first subsetted into the respective minimal pair stød categories for each dialect, leaving a (i) non-stød vs regular stød and (ii) regular stød vs tonal stød pair for analysis per dialect group. To each of these minimal pair categories GAMMs were fitted with the relevant (normalised) acoustic correlate as the outcome variable, a parametric term for stød type and added smooth terms for normalised time and stød type by normalised time as predictors. The models also included random smooths for speaker and word and an AR1 error model specification, the value of which was found by testing several different values of *rho* for each model and deciding the best fit based on the lowest model Akaike Information Criterion (AIC) score. The parametric term captures stød type effects on the height of the relevant acoustic correlate trajectory whereas the smooth term captures stød type effects on the shape of the trajectory.

4.1.1.1. f0

Effects for f0 are reported in Table 4 below; first for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this effect is significant, effects are reported for a model comparison excluding the smooth term for stød type by normalised time to specifically test for the effects of stød type on f0 trajectory shape.

Comparison	χ^2	*df*	$p(\chi^2)$
CPH non vs regular stød			
Overall	52.66	3	<0.001
Shape	9.89	2	<0.001
CPH regular vs tonal stød			
Overall	0.42	3	0.839
Shape	N/A	N/A	N/A

Table 4. Effects of stød type on f0 trajectory for the Copenhagen dialect. The overall effect represents the difference in height. Shape differences were not tested as overall height differences were not significant.

There is a significant overall effect of stød type on both the height and shape of the f0 trajectory for the non-stød vs regular stød contrast but not for the regular vs tonal stød contrast. An important part of GAMM interpretation is to visualise the fitted trajectories to evaluate exactly how they differ as the model summary does not reveal the nature of a difference in conditions, just whether it is significant or not. To visualise the GAMMs for f0, two plots were created, one of the model prediction f0 trajectory for each stød type (no stød vs regular stød, encoded 'non' and 'reg') based on the mean f0 trajectories from the data (bold line) and including the corresponding pointwise confidence intervals (shaded coloured area around the line), and a plot of the difference smooth. First, the non-stød vs regular stød condition is explored:

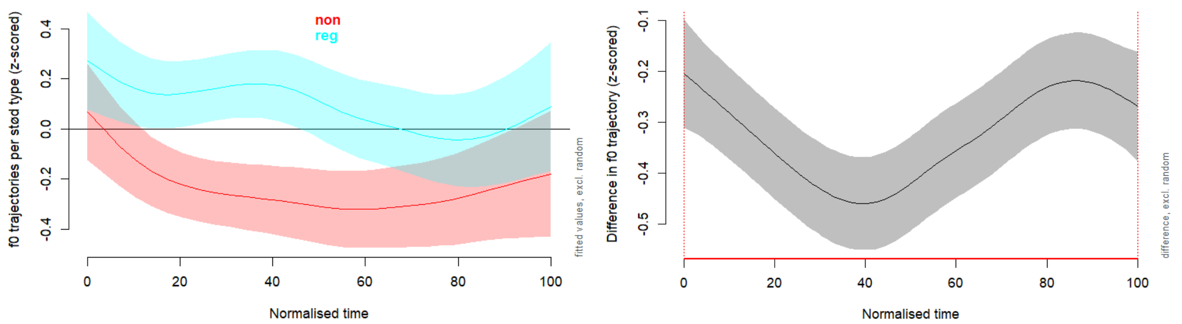

Figure 19. GAMM fit (left) and difference smooth (right) of the effect of stød type (non vs regular) on the f0 trajectories for the Copenhagen dialect with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red line on the difference smooth plot indicates the area where the two trajectories differ most.

The non-stød tokens' trajectories start relatively high with a gradual decline in f0 until around the 60% where the f0 slightly rises towards syllable termination. The regular stød tokens' trajectories are overall higher than the non-stød f0 trajectories and are characterised by a wave-like pattern of rise and fall – they start with an initial small dip, then a small rise peaking around 40%, followed by another dip with a trough around the 80% timepoint, finishing with a rise from there toward syllable termination. The significant effects of stød type on both height and shape of the f0 in the model predictions for the contrast appear to be influenced largely by the height and shape difference occurring in the 40% timepoint interval as shown by the dip here in the difference smooth, although they do differ significantly throughout their entire trajectories. There is, however, also some overlap in the confidence intervals towards syllable termination in the GAMM smooths. The peak in f0 at the 40% timepoint for the regular stød allows the f0 to decline which could be part of the stød realisation given the lack of such a trajectory peak in the non-stød tokens. It is surprising, however, that the stød condition generally has a higher f0 than the non-stød condition and that the f0 ends with a rise for the regular stød. The GAMM fit here suggest that what characterises the regular stød realisation is the relative f0 movement within the token rather than any absolute target value.

Table 4 showed that the regular vs tonal stød conditions had trajectories that did not differ significantly, unlike the non vs regular stød. The GAMM fit is visualised below:

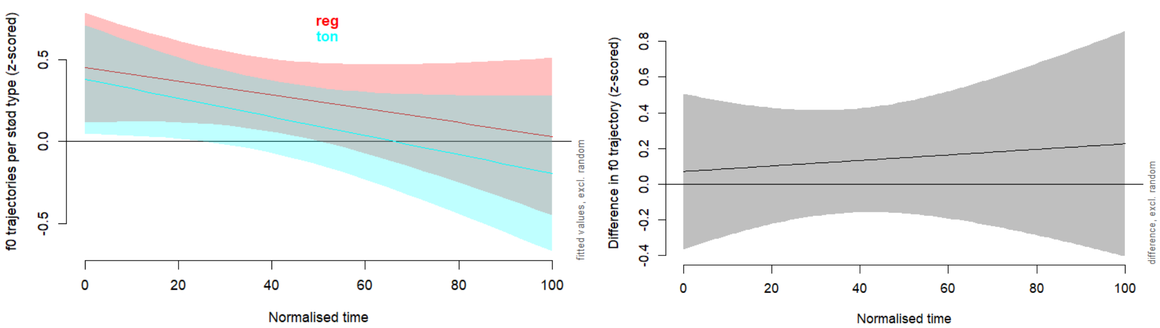

Figure 20. GAMM fit (left) and difference smooth (right) of the effect of stød type (regular vs tonal) on the f0 trajectories for the Copenhagen dialect with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth).

The two trajectories are similar in shape with minor height differences in f0 in the mean smooth, the regular stød being higher, but with a lot of overlap in the confidence intervals. This fits well with the hypothesis that the Copenhagen dialect speakers do not produce a separate tonal type of stød, only the Aarhus speakers are expected to have a tonal stød realisation which would result in different f0 trajectories for this contrast. The movement of the trajectories is similar, a steady decline throughout the syllable, which looks different from the f0 trajectory of the regular stød in the GAMM in Figure 19. The differences could be due to differences in stød basis for the regular stød categories – in the non vs regular stød contrast the stød basis is a long vowel whereas the stød basis for the regular stød in the regular vs tonal stød contrast is a short vowel and a sonorant consonant.

4.1.1.2. Intensity

The effects of stød type on the intensity trajectory are reported in Table 5 below. Similarly to the analysis of f0, the results are first reported for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this difference is significant, effects are reported for a model comparison between the full model and another nested model excluding only the smooth term for stød type by normalised time to specifically test for the effects of stød type on intensity trajectory shape.

Comparison	χ^2	*df*	$p(\chi^2)$
CPH non vs regular stød			
Overall	201.22	3	<0.001
Shape	199.32	2	<0.001
CPH regular vs tonal stød			
Overall	7.40	3	0.002
Shape	4.56	2	0.010

Table 5. Effects of stød type on intensity trajectory for the Copenhagen dialect. The overall effect represents the difference in height, the shape effect represents the difference in shape.

Looking at the non-stød vs regular stød subset first, there is a significant difference in both height and shape of the intensity trajectory dependent on the presence/absence of the stød. To visualise the nature of these differences, the model prediction trajectories and the difference smooth are plotted below:

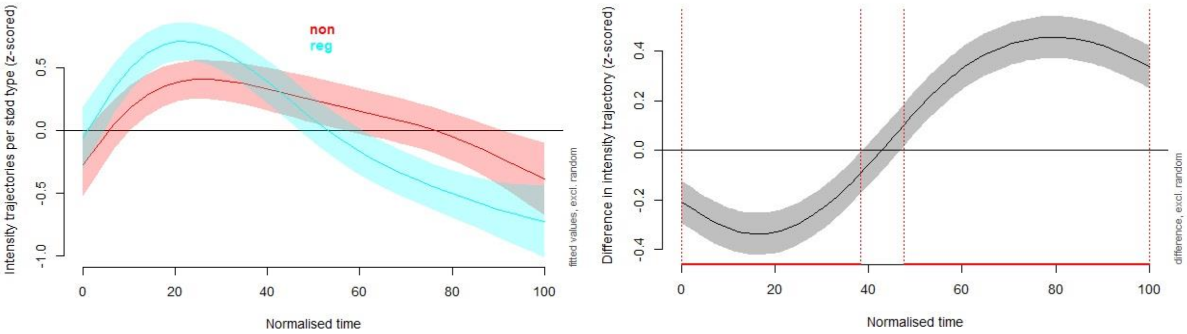

Figure 21. GAMM fit (left) and difference smooth (right) of the effect of stød type (non vs regular) on the intensity trajectories for the Copenhagen speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

The regular stød tokens start with an initial higher intensity spike until around the 40% timepoint after which the two trajectories briefly overlap to then separate in the opposite direction of the syllable-initial difference, i.e. the regular stød tokens are more steeply declining in intensity towards syllable termination, creating a height difference that is significant in the model comparison. For shape, both trajectories follow a curve of an initial rise and a subsequent fall but with the intensity trajectory for the regular stød tokens rising and declining much steeper, thus causing a significant shape difference, particularly between the 0-38% timepoints and 50-100% timepoints as evident from the difference smooth. As both intensity trajectories end relatively low it is difficult to determine the effect of laryngealisation on intensity on this contrast without including other measures – however, it is evident that the regular stød tokens fall steeper and end lower which suggests that a difference in voice quality might be present on these tokens causing this more rapid change.

Looking at the height and shape effects in Table 5 above for the regular vs tonal stød tokens for the Copenhagen speakers' intensity these trajectories are also significantly different from each other. To visualise how they differ the model prediction and difference smooth are plotted below:

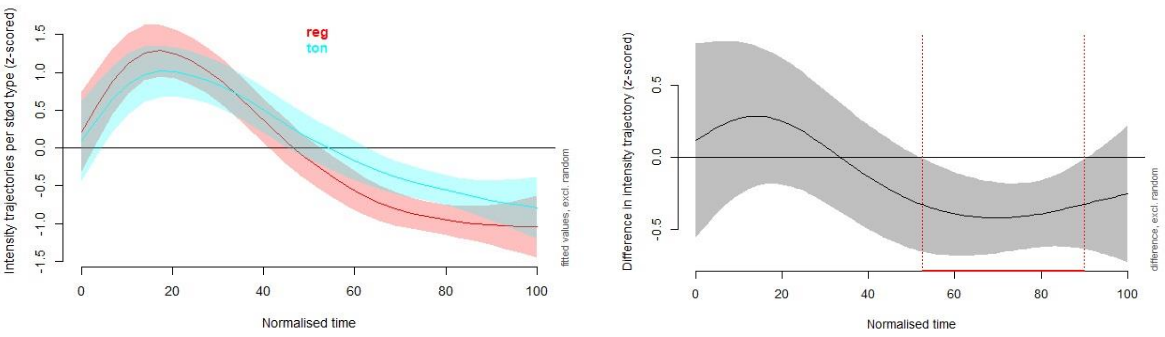

Figure 22. GAMM fit (left) and difference smooth (right) of the effect of stød type (regular vs tonal) on the intensity trajectories for the Copenhagen speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

The GAMM fit shows a similar pattern for the regular stød tokens as the previous contrast in that the intensity trajectory is higher with a steep rise until around the 40% timepoint after which the intensity declines toward syllable termination, the regular stød tokens losing more height than the tonal stød tokens. In terms of shape both trajectories start with a rise proceeded by an intensity decline but the shape differences occur due to the much steeper rise and fall of the intensity trajectory for the regular stød tokens. The difference smooth indicates that the largest differences occur between the 55-90% timepoints where the regular stød trajectory declines steeper, separating from the tonal stød trajectory. It also indicates large variability as evident in the wide confidence intervals indicated by the grey shaded area around the mean smooth. Both trajectories in this condition fall steeper and lower than the non-stød tokens in the previous contrast, suggesting there might be laryngealisation on both the regular and tonal stød types for this contrast. That said, the regular vs tonal stød types were not expected to be different to a significant degree in the model comparison.

No type of laryngealisation is characterised purely by a drop in intensity and it does not appear as a correlate in Keating et al's (2015) classification of types of creaky voice but the intensity parameter was still included as it has been found to be one of the most prominent correlates of the stød (Fischer-Jørgensen 1989). The models fitted here suggest that intensity does play a role in distinguishing stød tokens from non-stød tokens overall in Copenhagen speakers.

4.1.1.3. Cepstral Peak Prominence

As with the other measures, the results are first reported for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this difference is significant, effects are reported for a model comparison between the full model and another nested model excluding only the smooth term for stød type by normalised time to specifically test for the effects of stød type on CPP trajectory shape.

Comparison	χ^2	*df*	$p(\chi^2)$
CPH non vs regular stød			
Overall	413.99	3	< 0.001
Shape	267.28	2	< 0.001
CPH regular vs tonal stød			
Overall	22.87	3	< 0.001
Shape	20.74	2	< 0.001

Table 6. Effects of stød type on CPP trajectory for the Copenhagen dialect. The overall effect represents the difference in height, the shape effect represents the difference in shape.

Table 6 shows a significant difference between the CPP trajectories for the non-stød tokens vs the regular stød tokens in both height and shape of the trajectories. These differences are visualised below:

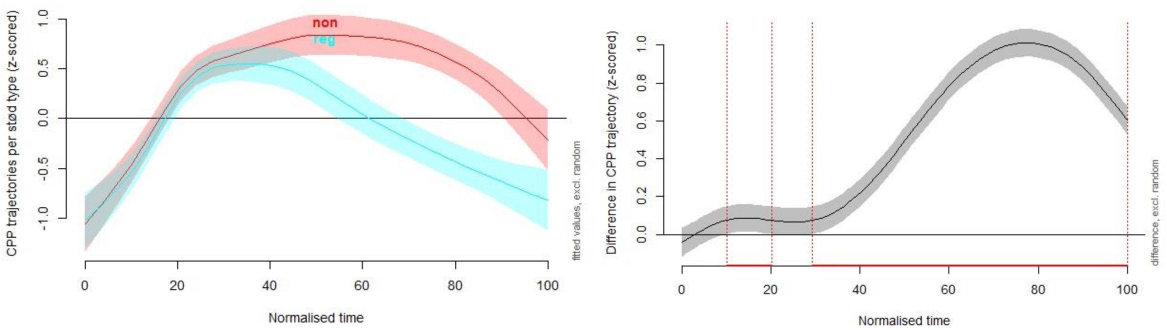

Figure 23. GAMM fit (left) and difference smooth (right) of the effect of stød type (non vs regular) on the CPP trajectories for the Copenhagen speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

As is evident in the GAMM fit and the difference smooth, the largest differences occur between the 30-100% timepoints. The trajectories are very similar initially, both having a steep rise up until they start to diverge just before the 40% timepoint. Both trajectories end with a decline in CPP but differ in the timing and steepness of this decline. The regular stød tokens fall steadily from around the 40% timepoint, whereas the non-stød tokens rise and stay high until the 70% timepoint, only declining in the last 70-100% timepoint interval. A decline in CPP indicates less periodicity and correlates with laryngealisation, suggesting that the regular stød tokens are significantly less periodic than the non-stød tokens, particularly from around timepoint 50% and onwards.

For the regular vs tonal stød tokens there was also a significant difference in both height and shape of the CPP trajectory between the stød types as reported in Table 6. The GAMM visualisations are plotted below:

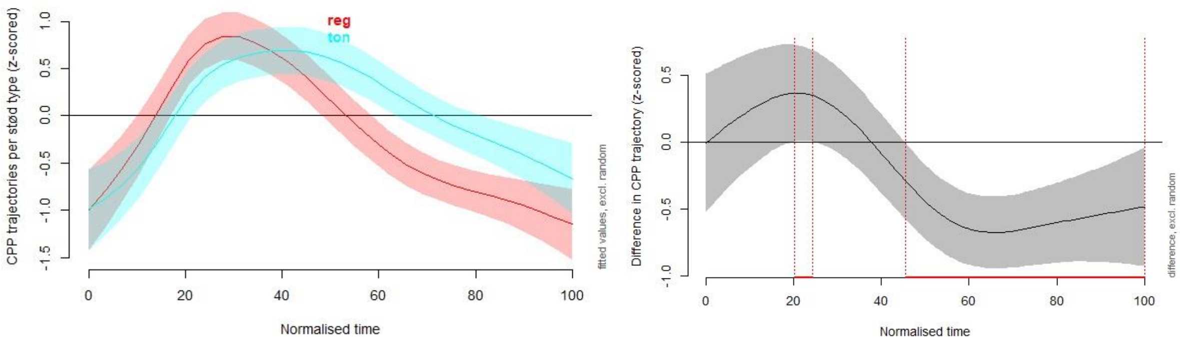

Figure 24. GAMM fit (left) and difference smooth (right) of the effect of stød type (regular vs tonal) on the CPP trajectories for the Copenhagen speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

Both trajectories start with an initial rise, the regular stød tokens slightly higher than the tonal stød tokens, followed by a decline towards syllable termination. The timing of this decline differs such that the regular stød token trajectory starts to decline around the 36% timepoint, whereas the tonal stød token trajectory declines around the 50% timepoint. The trajectory height difference is primarily caused by the regular stød token trajectory having a steeper as well as an earlier decline, ending much lower than the tonal stød tokens. This lower CPP suggests less periodicity and thus more laryngealisation of the regular tokens. As with previous measures, finding a difference in laryngealisation between the regular and tonal stød tokens for the Copenhagen speakers is unexpected as only the Aarhus dialect has been speculated to differ in types of stød.

4.1.1.4. H1-H2

As with previous measures the results are first reported for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this difference is significant, effects are reported for a model comparison between the full model and another nested model excluding only the smooth term for stød type by normalised time to specifically test for the effects of stød type on H1-H2 trajectory shape.

Comparison	χ^2	*df*	$p(\chi^2)$
CPH non vs regular stød			
Overall	118.90	3	< 0.001
Shape	54.60	2	< 0.001
CPH regular vs tonal stød			
Overall	0.62	3	0.743
Shape	N/A	N/A	N/A

Table 7. Effects of stød type on H1-H2 trajectory for the Copenhagen dialect. The overall effect represents the difference in height, the shape effect represents the difference in shape.

There is a significant difference between the non-stød and regular stød H1-H2 trajectories both in height and shape. The GAMM trajectories and the difference smooth are visualised below:

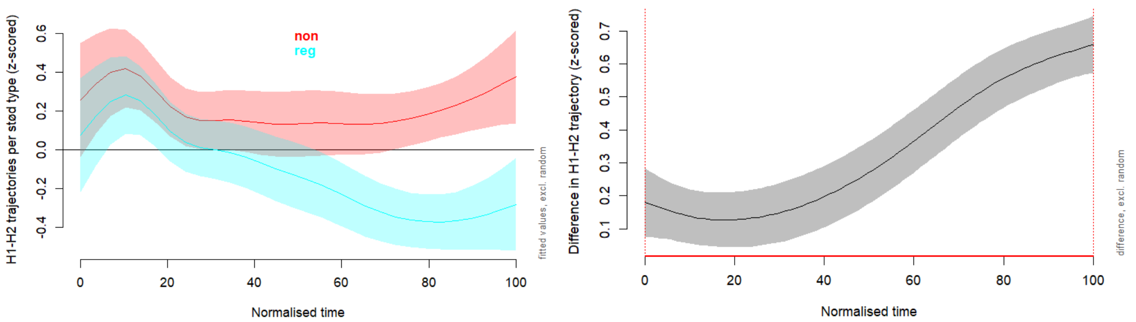

Figure 25. GAMM fit (left) and difference smooth (right) of the effect of stød type (non vs regular) on the H1-H2 trajectories for the Copenhagen speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

The two trajectories start off with similar shapes but with height differences, the non-stød token trajectory being slightly higher. Both trajectories start declining just before the 20% timepoint but where the non-stød trajectory flattens horisontally after the initial decline, the regular stød trajectory continue declining up until the 80% timepoint, creating large differences in height and shape between the conditions. They both, however, end with a slight rise. The declining H1-H2 in the regular stød could indicate a higher degree of laryngealisation on these tokens as H2 shifts to be more dominant than H1 with certain voice quality changes.

As Table 7 shows there was no overall significant difference between the regular stød tokens and the tonal stød tokens in H1-H2 trajectories. The visualised GAMM is plotted below:

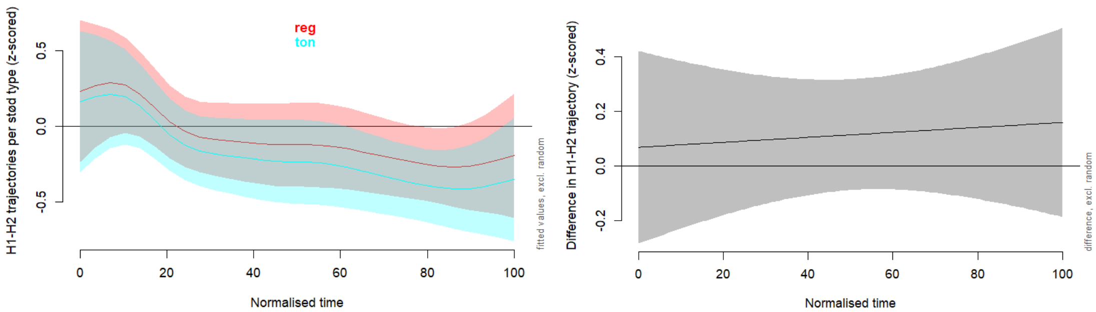

Figure 26. GAMM fit (left) and difference smooth (right) of the effect of stød type (regular vs tonal) on the H1-H2 trajectories for the Copenhagen speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth).

Both the regular and the tonal stød condition have trajectories that have a very minor initial rise and decline steadily throughout the syllable, ending with a very small rise in the mean smooths after the 90% timepoint. The mean smooth of the regular stød is slightly higher throughout but with a lot of overlap in the confidence intervals between types. H1-H2 was expected to decline in both stød types in Copenhagen and indeed, this is what happens in the GAMM until the last 10%. The rise is fairly small and might be driven by anticipation of the following segment or might be utilised to enhance perception of the contrast by not producing it on the entirety of the syllable it is contrastive on. As the previous measures did not all exhibit this behaviour of timing for perceptual contrast, it seems most likely that the small rise is anticipatory rather than contrast-enhancing in nature.

4.1.1.5. Harmonic-to-Noise Ratio

The effects of stød type on the HNR trajectory are reported in Table 8 below. Again, the results are first reported for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this difference is significant, effects are reported for a model comparison between the full model and another nested model excluding only the smooth term for stød type by normalised time to specifically test for the effects of stød type on HNR trajectory shape.

Comparison	χ^2	*df*	$p(\chi^2)$
CPH non vs regular stød			
Overall	715.82	3	< 0.001
Shape	406.07	2	< 0.001
CPH regular vs tonal stød			
Overall	14.84	3	< 0.001
Shape	14.10	2	< 0.001

Table 8. Effects of stød type on HNR trajectory for the Copenhagen dialect. The overall effect represents the difference in height, the shape effect represents the difference in shape.

The presence of the regular stød has a significant effect on the HNR trajectory compared to the non-stød tokens in both height and shape. Compare the GAMM fits and the difference smooth between the non-stød and regular stød below:

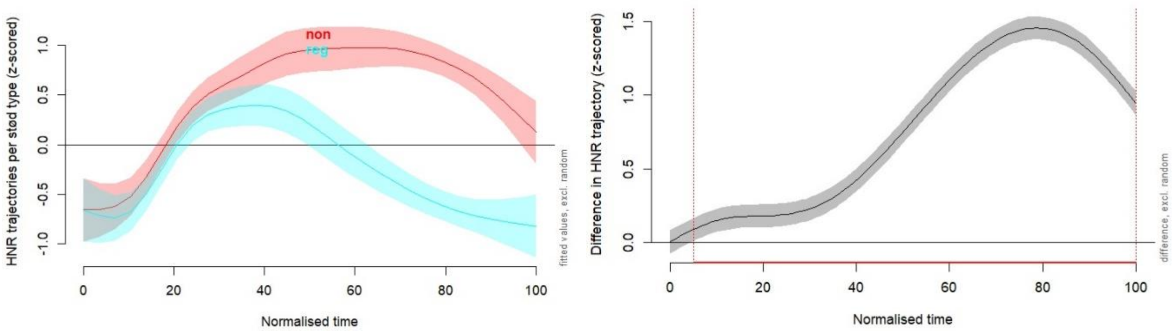

Figure 27. GAMM fit (left) and difference smooth (right) of the effect of stød type (non vs regular) on the HNR trajectories for the Copenhagen dialect with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

As evident in Figure 27 the non-stød and regular stød HNR trajectories start with similar shapes but with a height difference until right before the 30% timepoint where the non-stød trajectory rises until a final minor fall towards syllable termination. In contrast, the regular stød trajectory exhibits a steep fall from around the 50% timepoint ending very low at syllable termination. The difference smooth also shows peak differences around the 80% timepoint where the height between the two trajectories is evident in the GAMM smooth. A low HNR indicates less periodicity relative to noise meaning the difference between the non-stød and regular stød observed could be due to laryngealisation on the regular stød tokens.

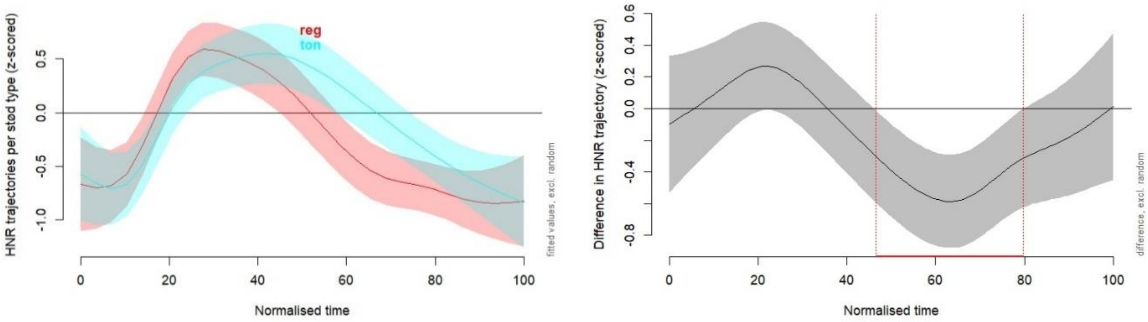

Figure 28. GAMM fit (left) and difference smooth (right) of the effect of stød type (regular vs tonal) on the HNR trajectories for the Copenhagen dialect with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

Looking at Figure 28 showing the regular vs tonal stød trajectories these are also significantly different. Their height difference is less apparent overall but observing the difference smooth there is a 50-80% timepoint interval where they diverge the most. Looking at the 40% timepoint there is also a shape difference where the regular tokens see a steep fall until syllable termination whereas the tonal tokens' trajectory stay higher for longer and fall less steeply. The regular tokens generally have a lower HNR after the 50% timepoint, indicating more laryngealisation for these tokens, at least until the overlap at the end of the syllable. For this dialect a significant difference between the regular and tonal tokens is not expected as the stød should be realised by similar laryngealisations in both conditions but was indicated in the GAMM smooths due to timing of the peak vs decline of the trajectories with the regular tokens peaking earlier and thus declining steeper.

4.1.1.6. Subharmonic-to-Harmonic Ratio

The effects of stød type on the SHR trajectory are reported in Table 9 below. The results are first reported for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this difference is significant, effects are reported for a model comparison between the full model and another nested model excluding only the smooth term for stød type by normalised time to specifically test for the effects of stød type on SHR trajectory shape. It should be noted, however, that this measurement was by far the most poorly tracked and thus the GAMM fit will be less reliable than the previous measures.

Comparison	χ^2	*df*	$p(\chi^2)$
CPH non vs regular stød			
Overall	143.15	3	< 0.001
Shape	45.68	2	< 0.001
CPH regular vs tonal stød			
Overall	1.82	3	0.304
Shape	N/A	N/A	N/A

Table 9. Effects of stød type on SHR trajectory for the Copenhagen dialect. The overall effect represents the difference in height, the shape effect represents the difference in shape.

The non-stød tokens for Copenhagen were particularly poorly tracked compared to the regular and tonal stød tokens so from the model comparisons, it is difficult to determine whether a statistically significant difference is due to a discrepancy in data points included at different time points (e.g. more data points earlier in the syllable vs later in the syllable across conditions) or due to differences in voice quality. As mentioned earlier, the SHR parameter characterises sound with alternating pulse cycles, defined as when the estimated f0 value is 20% higher or lower than the reference f0 value, calculated as an SHR percentage. For this reason, it is unexpected that the two conditions with stød have more tokens with a higher percentage of data points tracked compared to the non-stød condition as the opposite trend was apparent when analysing the f0 – less data points were tracked in the stød conditions compared to the non-stød conditions. Bearing all of this in mind, the GAMM visualisation of the non vs regular stød is plotted below:

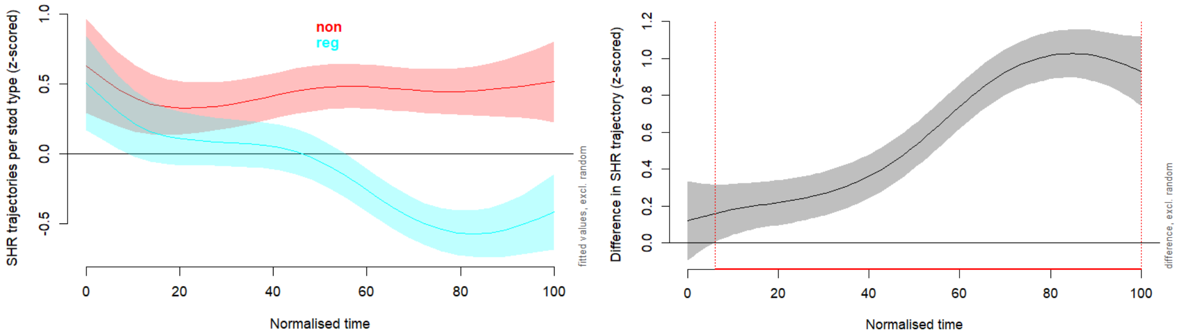

Figure 29. GAMM fit (left) and difference smooth (right) of the effect of stød type (non vs regular) on the SHR trajectories for the Copenhagen dialect with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

The non-stød tokens have a higher SHR trajectory than the regular stød tokens. If period doubling occurs, the SHR percentage is expected to be higher, meaning the model fit shows an unexpected reversal – if diplophonia was present, the regular stød tokens would be expected to have a higher SHR. This finding is particularly surprising given that the regular stød tokens were tracked better overall compared to the non-stød tokens, meaning that if period-doubling was present, the GAMM fit should be biased towards a high SHR for the regular stød more than the non-stød due to less omitted NAs in the model fit and thus more data points. As such, the GAMM trajectory shapes here are likely caused by the relative distribution of NAs over each token rather than the absolute number of NAs in a token.

For the regular vs tonal stød there was no statistical significance in the difference between trajectories overall. The GAMM visualisation is plotted below:

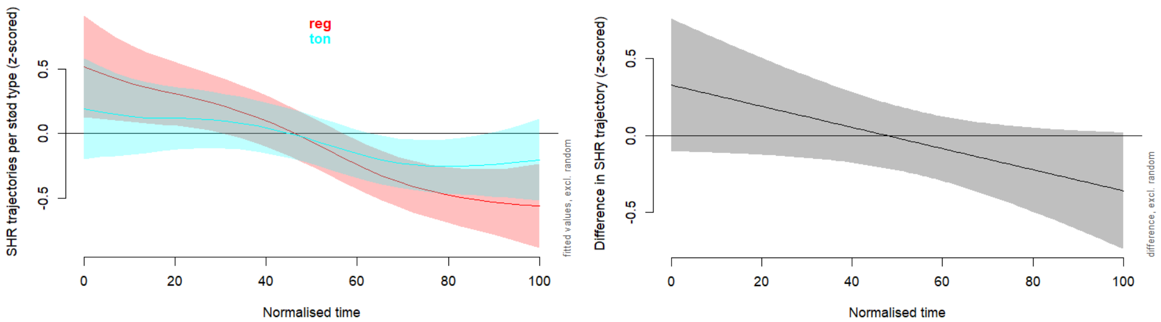

Figure 30. GAMM fit (left) and difference smooth (right) of the effect of stød type (regular vs tonal) on the SHR trajectories for the Copenhagen dialect with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth).

The GAMM fit reveals the regular stød tokens declining throughout the syllable which would suggest a lack of period doubling as in the previous contrast. The tonal stød condition exhibits a more stable trajectory with a small decline followed by levelling at around the 70% timepoint. The two trajectories overlap considerably throughout both in the mean smooth and the confidence intervals, meaning the difference between them is too small to be statistically significant despite the apparent difference in height and shape caused by the decline in the regular stød tokens.

For SHR, the current data does not allow any solid conclusions and the GAMMs presented in this section are too unreliable to meaningfully include in the classification of the phonetic stød types in this research. The patterns observed are unexpected but it cannot be concluded whether that is due to the patterns in the data being genuinely different than hypothesised or due to the lack of tracking, or more specifically, the distribution of the lack of tracking in each trajectory. As such, SHR will not be considered in the classification of the stød in the remaining analysis.

4.1.1.7. Classification of Copenhagen laryngealisation subtypes

In the previous sections, GAMMs were fitted to the token trajectories for each minimal pair contrast to compare how the respective stød conditions differed from each other. The purpose was to compare each stød type to the classification of laryngealisations in Keating et al (2015) as a reference guide for how much phonetic variation occurs during this phonological contrast often described as a kind of creaky voice. To this aid, comparisons are made between the main acoustic measures of the different types of laryngealisation and the results of the statistical model predictions for the Copenhagen speakers. The comparison also serves to summarise the findings presented in the GAMM models above as they will be assessed against the classifications of subtypes in the following.

The first two types of laryngealisation are very similar in that they both correlate with a low f0 and a low H1-H2. Their descriptions from Keating et al (2015) are reposted here for convenience:

(i) Prototypical creaky voice has a low f0, irregular vibrations correlating with high noise and glottal constriction correlating with a low H1-H2.

(ii) Vocal fry has a low f0, glottal constriction correlating with a low H1-H2 and damped pulses, correlating with low noise, i.e. a high HNR.

What separates them is high noise for the prototypical creaky voice in contrast to low noise for the vocal fry. As a low CPP indicates more noise in the signal this is a good candidate for distinguishing two types that are otherwise similar on f0 and H1-H2. Another relevant measure in HNR which differentiates between the irregular vibrations present in prototypical creaky voice (low HNR) as opposed to the damped pulses during vocal fry (high HNR). First, however, it must be established whether the Copenhagen speakers had stød realisations that would qualify for either subtype, i.e. stød realisations produced with both a low f0 and a low H1-H2.

The GAMM smooth for f0 was higher rather than lower for the regular stød compared to the non-stød tokens overall. The trajectory did, however, have a peak followed by a decline, meaning a relative lowering of f0. Despite this, it was not reaching absolute f0 values below those of non-stød tokens, the criteria of a low f0 cannot said to be fulfilled for the regular stød in Copenhagen. The regular vs tonal stød contrast both had a steadily declining f0 trajectory, the mean smooth being slightly higher on average for the regular tokens but not to a degree that was statistically significant. This means that these tokens could potentially have the criterion of a low f0 fulfilled. For the statistical modelling the measures were normalised across speakers which means the absolute values were not analysed. This makes it difficult to decide when a declining f0 can be said to be a low f0 in absolute terms. It has been found that instances of prototypical creaky voice will usually have an absolute f0 value below 100 Hz (Laver 1980: 122, Davidson 2019: 238-239) and thus it is worth considering the raw, non-normalised data when correlating a change in f0 to the types of laryngealisation occurring in the data because there is an expected absolute threshold from the literature for this measure, unlike other measures, allowing some quantification of what a 'low' f0 is in this context. To explore whether the f0 in the current data was below this absolute threshold, a static timepoint first needed to be selected as the GAMMs were fitted dynamically over time, whereas the f0 value threshold would not be relevant across the entire syllable. This is because previous findings (Fischer-Jørgensen 1989) suggest that the stød 'proper' is articulated in the latter part of the syllable in Copenhagen. As such, the raw data f0 means for the non-NA values for the Copenhagen speakers were calculated based on the 50-100% timepoints in each token as the latter part of the syllable is where the lowering effects on f0 are expected. The mean f0s are shown in Figure 31 below, segregated by speaker sex:

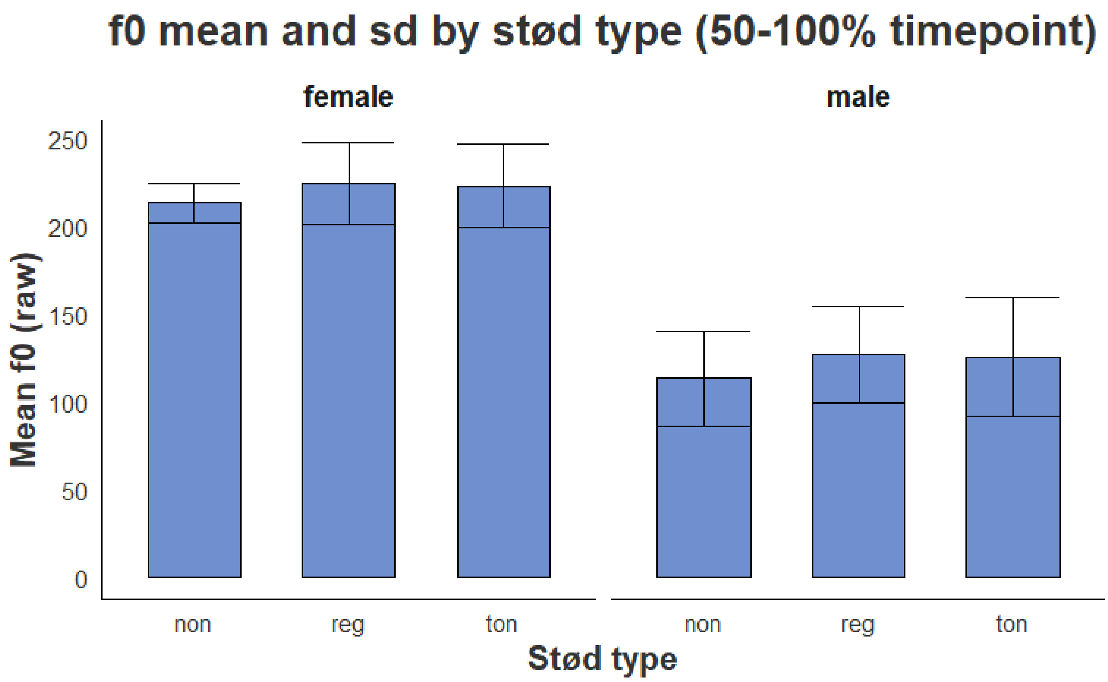

Figure 31. Raw mean f0 at the 50-100% timepoints for the Copenhagen speakers' non-NA values segregated by stød type and speaker sex with the error bar indicating the standard deviation. 'Non' = no stød, 'reg' = regular stød and 'ton' = tonal stød.

The f0 values for the non-stød tokens are generally the lowest which is contrary to the hypothesised effects but consistent with the GAMM trajectories. The female mean f0 for non-stød tokens is 215.4 Hz (*n* = 3663) and for the males it is 115.2 Hz (*n* = 3663). The regular and the tonal stød tokens for the female Copenhagen speakers have very similar f0 values at 229.2 Hz for the regular stød (*n* = 4928) and 228.5 Hz for the tonal stød (*n* = 1243), meaning the regular and tonal stød types, despite declining in the GAMM, do not have an absolute low f0 characteristic of prototypical creaky voice. For the male Copenhagen speakers the regular stød tokens' mean f0 is 129.0 Hz (*n* = 4895) and the tonal tokens' mean f0 is 129.9 Hz (*n* = 1232). While these are noticeably lower than the female means, they do not cross the threshold of being below 100 Hz. It can be observed, however, that the standard deviation for the regular stød tokens is more variable than the tonal tokens and some tokens do have an f0 below 100 Hz. This is also true for the male non-stød tokens, however, meaning this cannot be an effect of the stød only. Just why the non-stød tokens have a lower f0 than both stød tokens is not clear but will be a point returned to in the discussion.

For now, the H1-H2 values will be evaluated to compare to the first to subtypes. For H1-H2 the values for the non-stød vs the regular stød contrast both had a decline but the non-stød token trajectory was small whereas the regular stød token trajectory declined sharply and ended lower, creating a statistically significant difference in the contrast. The regular vs tonal contrast both had a steady decline over their trajectory in H1-H2, the tonal mean smooth being slightly lower throughout but not to a degree that was statistically significant. As the trajectory for both the regular and the tonal stød declined, towards syllable termination it could be that a low H1-H2 was achieved. For H1-H2 there has not been any absolute threshold identified although when voice qualities on the hypercompressed continuum are produced, it is generally expected to result in a negative value due to the dominance of the second harmonic relative to the first harmonic. To explore the raw, non-normalised values, the means and standard deviations were calculated for the latter half of the syllables and plotted for each stød condition, segregated by speaker sex:

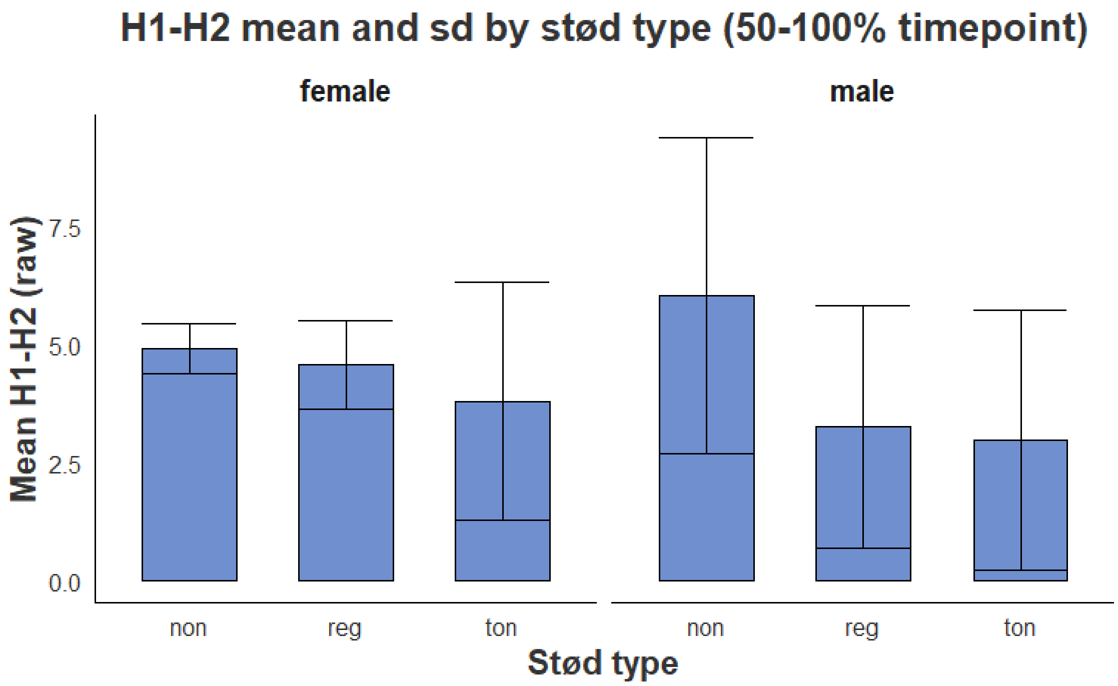

Figure 32. Raw mean H1-H2 at the 50-100% timepoints for the Copenhagen speakers' non-NA values segregated by stød type and speaker sex with the error bar indicating the standard deviation. 'Non' = no stød, 'reg' = regular stød and 'ton' = tonal stød

As Figure 32 shows the H1-H2 values for the Copenhagen speakers are all positive means, indicating a lack of glottal constriction on all tokens. The standard deviations are largest for the male speakers whose lowest values occur for the tonal stød tokens. Even these, however, have a mean of 3.6 dB (*n* = 1232) and do not reach negative values. It should be noted, however, that H1-H2 were not tracked as consistently as other measures, particularly for the male speakers.

In summary, both prototypical creaky voice and vocal fry are defined by a low f0 and a low H1-H2. Despite some GAMMs showing a general decline of these measures throughout the stød basis, plotting the absolute values did not indicate fulfilment of these two criteria. The two types are further distinguished by the presence of high vs low noise, the prototypical creaky voice having high noise and vocal fry having low noise. The relevant measure for this was HNR but since f0 and H1-H2 are not indicating either of these two subtypes, looking at HNR at this stage is not relevant.

That leaves three subtypes left to comparatively explore. The third type is described by Keating et al (2015) as follows:

(iii) Multiply pulsed voice has irregular vibrations correlating with high noise, glottal constriction correlating with a low H1-H2 and subharmonics correlating with a high subharmonic to harmonic ratio (SHR).

This subtype is not indicated in the data as the H1-H2 is not low for any condition. The SHR measurements were generally very poor but the tracked data suggested that the SHR was higher on the non-stød tokens compared to the stød tokens, the opposite of the expected outcome. In terms of signal noise, the CPP and HNR measurements both had declining trajectories for the regular and tonal stød and were lower than the non-stød tokens, indicating increased signal noise caused by irregularities in the vocal fold vibratory patterns. This is, however, only one out of three criteria fulfilled for subtype three which is thus not indicated based on the available data.

The description of the fourth subtype were as follows:

(iv) Aperiodic creak lacks a low f0 but has irregular vibrations correlating with high noise and glottal constriction correlating with a low H1-H2.

The first criterion for this subtype, a lack of low f0, can said to comply with the data if the same threshold applies as to the first two subtypes, i.e. f0 needs to be below 100 Hz to be considered low in the context of creaky phonation. Further, the regular stød was unexpectedly found to have an f0 that were relatively higher than the non-stød tokens, meaning the non-low f0 criterion applies in a relative sense too. The second criterion of high noise also applies as CPP and HNR were both declining for the regular and tonal stød tokens as discussed for subtype three just before. The third criterion, a low H1-H2, is not fulfilled despite a relative decline of the regular and tonal stød tokens in the GAMMs as the values were positive, suggesting a lack of glottal constriction. In summary, subtype four is not compliant with the data.

The fifth and last subtype in Keating et al's (2015) classification was described as:

(v) Non-constricted creak which has a low and irregular f0 correlating with high noise but no indication of glottal constriction, i.e. non-low H1-H2.

This subtype complies with the data in having high(er) noise for the regular and tonal stød tokens as indicated by the declining CPP and HNR and having a non-low H1-H2. The subtype is described as having a low and irregular f0 *correlating* with high noise which is a complex criterion to evaluate for the data because it suggests that a low f0 must result in high noise. For the present data, f0 is not low but noise is still high(er), meaning these two descriptors are not always correlated. It can be argued that the criterion is fulfilled if one interprets the acoustic consequence of low and irregular f0, i.e. high noise, as the main classifier, but it could also be argued that f0 is not low, despite the high noise, and the criterion thus not fulfilled. If the latter interpretation is adopted, the stød types in Copenhagen do not fit neatly into any of the categories of subtypes described in Keating et al (2015), despite the stød being described as a kind of creaky voice. The final acoustic parameter measured, intensity, was not part of the subtype descriptors but was included in the current study because it had been found to be the most robust correlate of the stød in previous research (Riber Petersen 1973, Fischer-Jørgensen 1989) and thus thought relevant to capture phonetic variation, at least in the Copenhagen dialect. The intensity trajectories in the GAMMs did differentiate both stød contrasts to a statistically significant degree, the non vs regular stød having a larger intensity peak and a lower decline for the regular stød. The regular vs tonal stød contrast had the tonal stød intensity trajectory end slightly higher but both had a peak and a steep decline too. The decline in intensity is thought to correlate with irregular vocal fold vibrations but it would usually be expected to see these effects in f0 as well, making the results a bit inconclusive. However, since noise was also increased on the stød tokens, it might be that there are effects that affect noise and intensity more than f0, pertaining to the non-constricted creak type but with a slight variation compared to the strict classification criteria. What can be concluded is that intensity does seem to differentiate the stød conditions in the Copenhagen dialect, supporting previous findings.

To summarise the acoustic analysis for the Copenhagen dialect thus far, the purpose was to answer research question 1, how much phonetic variation occurs during the phonologically contrastive stød. To compare dialectal variation, both dialects needed to be analysed and characterised, starting here with the Copenhagen speakers. It was found that some measures behaved as expected, e.g. a decrease in intensity during the regular stød and an increase in noise evident in lower CPP and HNR measures for both stød types. Other measures behaved unexpectedly, e.g. a higher f0 for the regular stød compared to the non-stød and a lack of glottal constriction evident in a higher than expected H1-H2 for both stød types. Also unexpected was the statistically significant differences between the regular and tonal stød types for the GAMMs for intensity, CPP and HNR. This minimal pair was hypothesised to only differ in the Aarhus dialect due to a tonal realisation of the stød in closed monosyllables with a long vowel whereas for Copenhagen, it was expected to be realised as a voice quality change somewhere on the hypercompressed continuum.

In terms of characterising the Keating et al (2015) subtype most similar to the Copenhagen stød realisation, the measures suggested highest compatibility with subtype five, the non-constricted creak, but with the caveat that f0 was not low in absolute terms, despite a relative decline on the stød tokens.

In order to investigate phonetic variation between dialects, the following sections present the same acoustic analysis as above but for the Aarhus dialect to compare the stød realisations between the two.

4.1.2. Aarhus

4.1.2.1 f0

Effects for f0 are reported in Table below; first for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this effect is significant, effects are reported for a model comparison excluding the smooth term for stød type by normalised time to specifically test for the effects of stød type on f0 trajectory shape.

Comparison	χ^2	*df*	$p(\chi^2)$
AAR non vs regular stød			
Overall	699.00	3	<0.001
Shape	467.13	2	<0.001
AAR regular vs tonal stød			
Overall	0.92	3	0.606

Shape	N/A	N/A	N/A

Table 10. Effects of stød type on f0 trajectory for the Aarhus dialect. The overall effect represents the difference in height, the shape effect represents the difference in shape.

Table 10 shows a significant effect of regular stød compared to no stød on both f0 trajectories' height and shape. To analyse where they differ according to stød type, the mean smooth and difference smooth are plotted below:

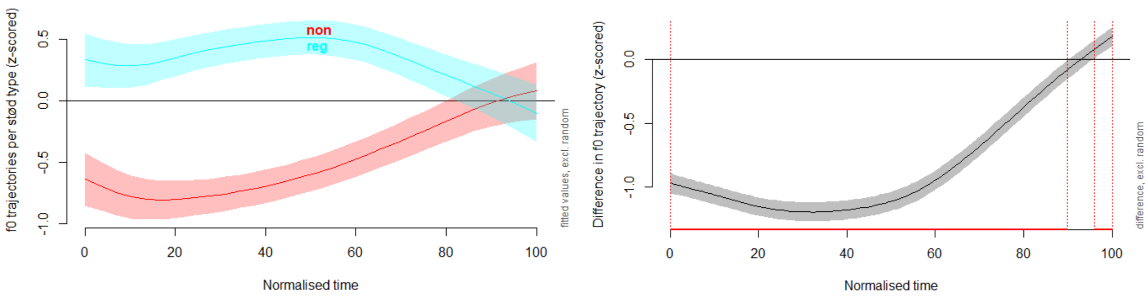

Figure 33. GAMM fit (left) and difference smooth (right) of the effect of stød type (non vs regular) on the f0 trajectories for the Aarhus dialect with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

The mean f0 trajectory for the tokens with no stød starts with a small dip, then a steady increase throughout the syllable. The f0 is comparatively low until after the 70% timepoint where the trajectory starts approaching the regular token trajectory. In contrast, the mean f0 trajectory for the regular stød tokens is very high and more level until a decrease throughout the final portion of the syllable starting after the 60% timepoint. Looking at the difference smooth, the f0 trajectories for this contrast differ significantly throughout most of the syllable with only a small gap of overlap just after the 90% timepoint. Despite this overlap, the trajectories have very different shapes – the non-stød token f0 trajectory is rising towards syllable termination and the regular stød token f0 is falling towards syllable termination. The fall of the regular tokens' f0 is expected with laryngealisation. However, the non-stød trajectory starts much lower than the regular stød token trajectory, despite these tokens not being hypothesised to be laryngealised, which suggests that the final f0 dip in the regular tokens might not be due to laryngealisation as it would be expected to be lower if this was the case. As such, these tokens seem to differ more in the former part of the syllable, an unexpected finding. It is not clear whether part of the stød realisation and/or perceptual cues for the stød is present early in the syllable before the stød proper which means the stød is realised through the relative movement of the f0 rather than an absolute value of a low target. This is suggested in the fall of the regular stød vs the rise of the non-stød token trajectories in the GAMMs despite surprising absolute differences in f0 height. This is only speculation, however.

The regular vs tonal stød contrast in the Aarhus dialect had no significant difference in the overall f0 trajectories. The GAMM predictions are visualised below:

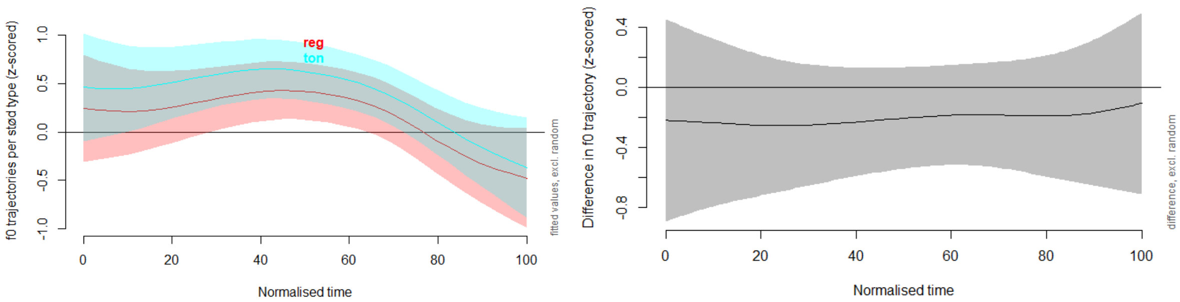

Figure 34. GAMM fit (left) and difference smooth (right) of the effect of stød type (regular vs tonal) on the f0 trajectories for the Aarhus dialect with the mean smooth indicated by the thick line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth).

The regular and tonal stød trajectories have a small difference in height with the tonal tokens having a slightly higher mean smooth for f0. Their shapes are very similar, both starting level, rising slightly and ending in a decline from around the 60% timepoint towards syllable termination. Based only on f0 it is difficult to determine whether the syllable-final decline is a correlation of a subtype of laryngealisation or a fall in tone. Since the trajectories are very similar, it might be that both the regular and tonal stød in Aarhus are laryngealised like in Copenhagen or it might be that both are somewhat tonal in their production. My own auditory impression of the regular tokens suggests that the latter might be the case for many tokens but the remaining measures will help determine what causes these trajectories to be similar in Aarhus – if they are both tonal they should not differ much in e.g. noise. The tonal stød syllable has been noted to sound ‘two-toned’ (Kyst 2004; 2007) which is argued to occur because the height of the syllabic tone reaches the same height as the other high tones in the phrase they occur in and the same lowest level as the lowest tones of the phrase. This makes the stød sound particularly long and falling and almost disyllabic to Kyst (2004) in closed monosyllables with a long vowel like *larm* [lɑːˀm] ‘noise’, attributed to the auditory impression of two tones because of the range of f0 in a short stretch of speech. The current study looks only at the stød basis as this is the phonological unit making up the contrast so it cannot be determined if the tones reach the same highs and lows as the surrounding phrase for the tonal stød. It can, however be measured whether the regular and tonal stød differ in measures indicating laryngealisation – if the tonal stød is articulated with tone as the realisation of the stød contrast instead of laryngealisation, then this should be observable in the remaining acoustic measures and the EGG measures. For now it can be noted that f0 in the regular vs tonal stød GAMMs do exhibit a pattern of a rise and a decline which would be expected with tonal realisations. The peak, however, looks somewhat moderate and the decline could also reflect laryngealisation. Moving on to the other measures will help inform the interpretation of f0 patterns here.

4.1.2.2. Intensity

The results are again first reported for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this difference is significant, effects are reported for a model comparison between the full model and another nested model excluding only the smooth term for stød type by normalised time to specifically test for the effects of stød type on intensity trajectory shape.

Comparison	χ^2	*df*	$p(\chi^2)$
AAR non vs regular stød			
Overall	143.86	3	< 0.001
Shape	117.15	2	< 0.001
AAR regular vs tonal stød			
Overall	12.93	3	< 0.001
Shape	9.30	2	< 0.001

Table 11. Effects of stød type on intensity trajectory for the Aarhus dialect. The overall effect represents the difference in height, the shape effect represents the difference in shape.

Table 11 shows a significant different between non-stød and regular stød token intensity trajectories in both height and shape. The differences are visualised below:

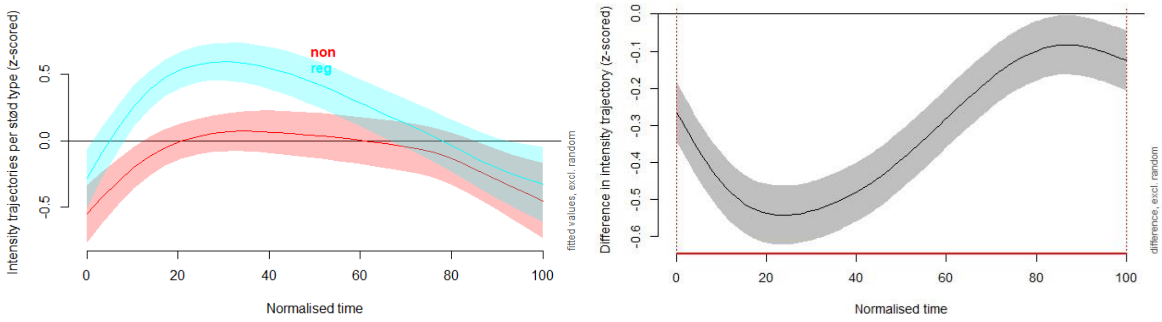

Figure 35. GAMM fit (left) and difference smooth (right) of the effect of stød type (no stød vs regular) on the intensity trajectories for the Aarhus speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

Both trajectories follow a path of a rise and a fall in intensity. However, the regular stød tokens see a much steeper initial spike peaking around the 30% timepoint and is consequently followed by a much steeper fall. The difference smooth shows a significant difference between the trajectories throughout, also evident in the GAMM fit where no overlap occurs in the mean trajectory lines, although some overlap occurs the confidence intervals in the latter part of the syllable. For intensity the Aarhus speakers appear to differentiate between these stød conditions more in the initial part of the syllable rather than the later part, utilising a relatively higher peak in the regular stød that allows for a steeper decline of the intensity trajectory, despite the absolute intensity never going below that of the non-stød tokens which have a lower mean smooth overall.

Turning to Table 11 again, there is also a significant difference between the intensity trajectory of the regular stød tokens and the tonal stød tokens. These trajectories are visualised below:

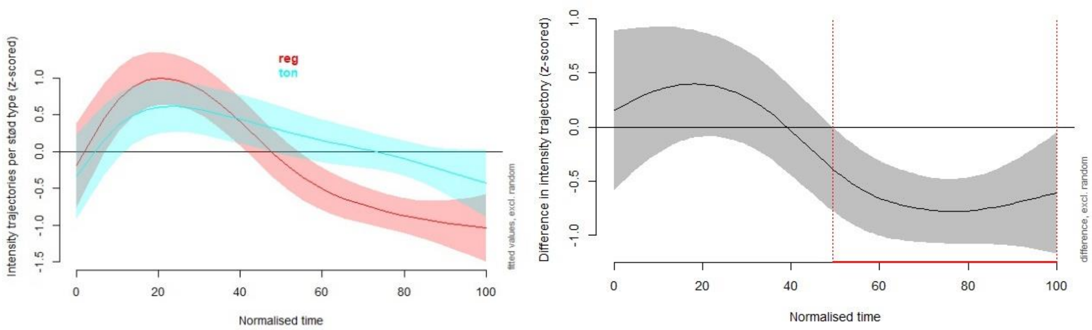

Figure 36. GAMM fit (left) and difference smooth (right) of the effect of stød type (regular vs tonal) on the intensity trajectories for the Aarhus speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

Both trajectories start with an initial rise with the regular stød tokens having a higher mean rise peaking around the 20% timepoint. The mean smooths overlap at the 40% timepoint and then split; as evident in the difference smooth the trajectory paths differ most significantly after this split from the 50-100% timepoints. For the tonal tokens there is a relatively linear and modest intensity fall, whereas the regular stød tokens fall more steeply and end low in intensity. This could indicate a higher degree of laryngealisation on the regular tokens as presence of irregular vibrations would be expected to lower the intensity.

4.1.2.3. Cepstral Peak Prominence

As with the previous measures, the results are first reported for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this difference is significant, effects are reported for a model comparison between the full model and another nested model excluding only the smooth term for stød type by normalised time to specifically test for the effects of stød type on CPP trajectory shape.

Comparison	χ^2	*df*	$p(\chi^2)$
AAR non vs regular stød			
Overall	133.76	3	< 0.001
Shape	100.24	2	< 0.001
AAR regular vs tonal stød			
Overall	8.65	3	< 0.001
Shape	5.66	2	0.004

Table 12. Effects of stød type on CPP trajectory for the Aarhus dialect. The overall effect represents the difference in height, the shape effect represents the difference in shape.

As evident from Table 12 there is a significant effect of stød type on the CPP trajectory for both height and shape for the non-stød vs regular stød tokens, visualised below:

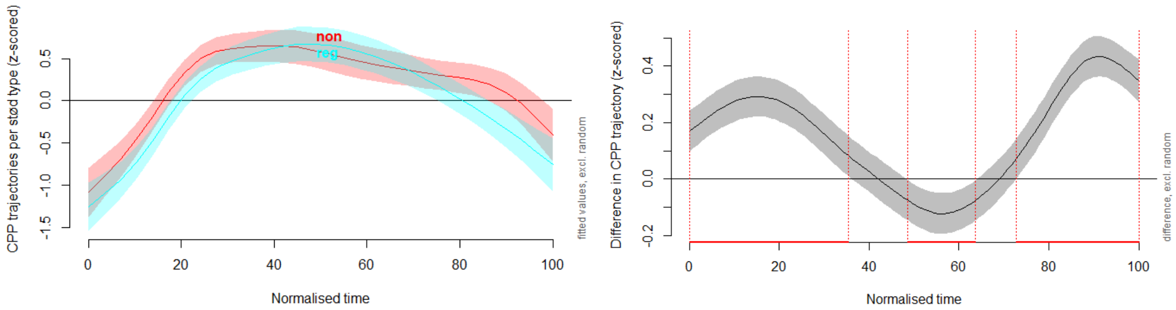

Figure 37. GAMM fit (left) and difference smooth (right) of the effect of stød type (non vs regular) on the CPP trajectories for the Aarhus speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

Between the initial 0-35% timepoint the two trajectories follow a similar shape but with a height difference, the regular stød tokens having a lower mean CPP smooth that the non-stød tokens. After the 40% timepoint the trajectories overlap whereafter the regular stød token mean trajectory continues to rise whereas the non-stød token trajectory begins slightly declining, creating a difference in height as well as shape. The regular stød tokens also decline in CPP but only after the 50% timepoint. As evident from both the GAMM fit and the difference smooth the most notable difference occurs towards syllable termination where the regular stød tokens decline more rapidly and steeply in CPP than the non-stød tokens. Their CPP ends lower, possibly indicating laryngealisation as a lower CPP correlates with less periodicity. The absolute difference in height after the mid-point of the syllable seems less prominent than the relative difference in height and shape caused by the staggering and steepness of the respective CPP declines.

For the regular vs tonal stød contrast there is also a significant height and shape difference between the two stød type mean CPP trajectories. These are visualised below:

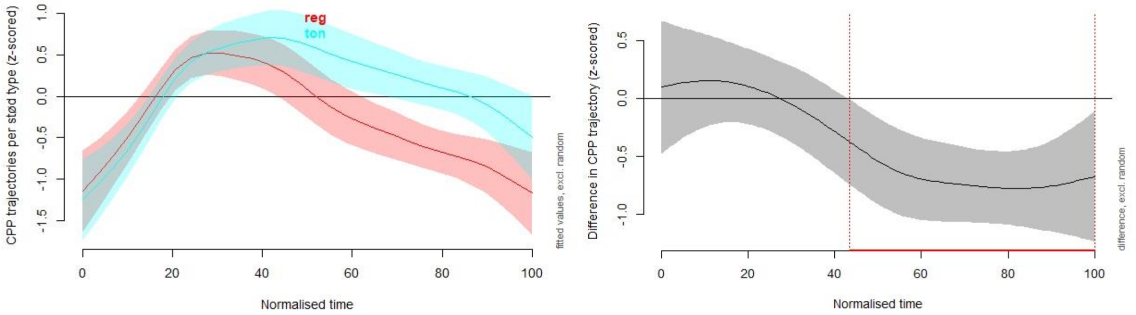

Figure 38. GAMM fit (left) and difference smooth (right) of the effect of stød type (regular vs tonal) on the CPP trajectories for the Aarhus speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

The trajectories start similarly with an initial steady rise after which divergence begins around the 30% timepoint. The tonal stød token trajectory reaches a higher peak but declines less steeply compared to the regular stød tokens, creating especially height but also shape differences between the 42-100% timepoints. The regular tokens end considerably lower than the tonal tokens, a difference possibly caused by laryngealisation on the former, although the tonal tokens also see a fall in CPP which might indicate a degree of laryngealisation on these tokens too, but lesser so.

4.1.2.4. H1-H2

As with the previous measures, the results are first reported for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this difference is significant, effects are reported for a model comparison between the full model and another nested model excluding only the smooth term for stød type by normalised time to specifically test for the effects of stød type on H1-H2 trajectory shape.

Comparison	χ^2	*df*	$p(\chi^2)$
AAR non vs regular stød			
Overall	165.18	3	< 0.001
Shape	160.98	2	< 0.001
AAR regular vs tonal stød			
Overall	1.07	3	0.544

Shape	N/A	N/A	N/A

Table 13. Effects of stød type on H1-H2 trajectory for the Aarhus dialect. The overall effect represents the difference in height, the shape effect represents the difference in shape.

Table 13 shows a significant difference between the non-stød and regular stød tokens both in height and shape as visualised below:

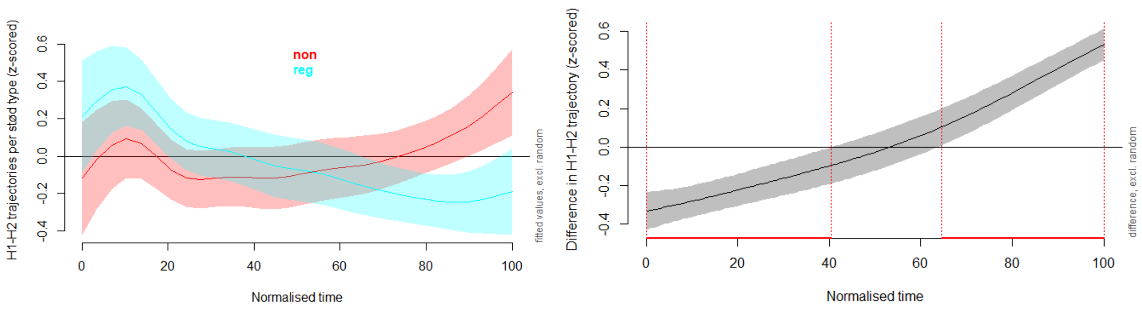

Figure 39. GAMM fit (left) and difference smooth (right) of the effect of stød type (non vs regular) on the H1-H2 trajectories for the Aarhus speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

The two trajectories both start with an initial rise and a peak around the 18% timepoint after which they both decline. The regular stød tokens initially have a higher H1-H2 than the non-stød tokens. The non-stød token trajectory rises again after the 50% timepoint, whereas the regular stød token trajectory continues to decline until the 95% timepoint where the mean smooth has a very small rise towards syllable termination. This creates large differences between the non-stød and regular stød H1-H2 in the last 40% of the syllable where the regular tokens have a lower H1-H2, indicating a shift from the H1 being dominant to the H2 being dominant, suggesting a degree of laryngealisation. This contrast seems to be attributed more to the relative steepness of the regular tokens' decline rather than reaching any absolute value.

The regular vs tonal stød tokens were not significantly different in the model comparison, suggesting that they are produced more similar than the non-stød vs regular stød contrast. The model visualisation is plotted below:

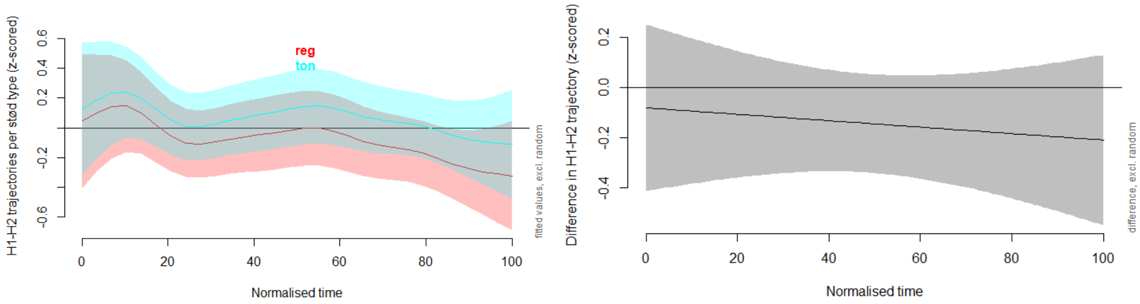

Figure 40. GAMM fit (left) and difference smooth (right) of the effect of stød type (regular vs tonal) on the H1-H2 trajectories for the Aarhus speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth).

Both trajectories start with an initial rise followed by a decline around the 15% timepoint. There is a small rise again peaking just before the 60% timepoint, followed by syllable-final declines for both stød types. The tonal stød tokens have higher mean H1-H2 throughout the entire syllable, indicating a slightly more dominant H1. However, the confidence intervals overlap quite significantly and it is not clear if the higher mean smooth for the tonal tokens would be a relevant perceptual difference or not. As H1-H2 was one of two measurements more dependent on f0, it is difficult to interpret whether the decline in H1-H2 here is due to laryngealisation or tone, given that both stød types have a small decline after the 50% timepoint.

4.1.2.5. Harmonic-to-Noise Ratio

The effects of stød type on the HNR trajectory are reported in Table 14 below. Again, the results are first reported for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this difference is significant, effects are reported for a model comparison between the full model and another nested model excluding only the smooth term for stød type by normalised time to specifically test for the effects of stød type on HNR trajectory shape.

Comparison	χ^2	*df*	$p(\chi^2)$
AAR non vs regular stød			
Overall	291.15	3	< 0.001

Shape	275.58	2	< 0.001
AAR regular vs tonal stød			
Overall	7.60	3	0.002
Shape	9.23	2	< 0.001

Table 14. Effects of stød type on HNR trajectory for the Aarhus dialect. The overall effect represents the difference in height, the shape effect represents the difference in shape.

The model comparison for the non-stød vs regular stød came out significant in terms of both height and shape of the respective trajectories. The difference is visualised below:

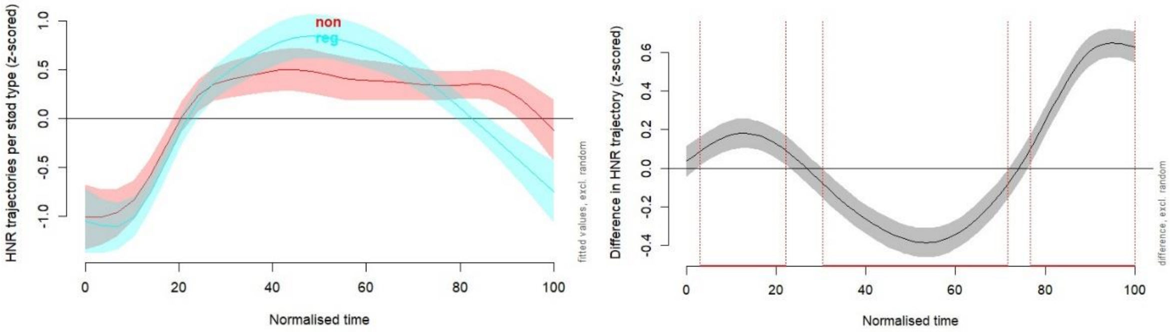

Figure 41. GAMM fit (left) and difference smooth (right) of the effect of stød type (non vs regular) on the HNR trajectories for the Aarhus dialect with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

Both trajectories start with a similar shape but the regular token trajectory has a lower HNR in the mean smooth. The shape difference is obvious after around the 30% timepoint where the regular stød tokens see an increase in HNR and then a steep fall from around timepoint 70% and onwards, suggesting an increase in noise which could be due to laryngealisation on the regular tokens. The non-stød token trajectory starts with an initial rise but then remains relatively flat up until a minor fall in the last 10% of the total duration. The difference smooth confirms that the trajectories differ most in the last 20% of the duration but with a large difference around the middle of the trajectory too, caused by a higher HNR peak in the regular token trajectory. Both trajectories end on considerably higher HNR values than they have at their onset, indicating that the syllable-final fall in HNR may not conclusively be due to laryngealisation and the syllable onset is not laryngealised. It might be, however, that laryngealisation on the regular tokens is evident in the relative steepness of the HNR fall towards syllable termination rather than the absolute HNR value reached.

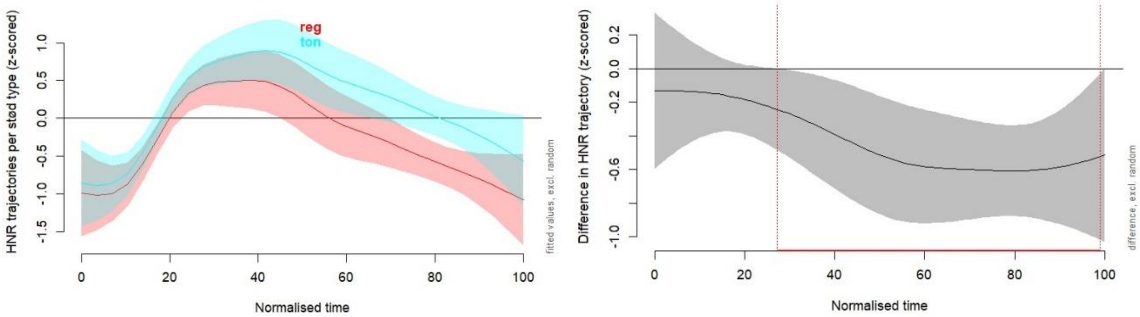

Figure 42. GAMM fit (left) and difference smooth (right) of the effect of stød type (regular vs tonal) on the HNR trajectories for the Aarhus dialect with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

Figure 42 shows the trajectories for the regular vs tonal stød contrast for the Aarhus speakers which start off similar but diverge after the 20% timepoint where both the GAMM fit and the difference smooth indicate that most variation occurs. Both height and shape differences were found to be significant in the model comparison, caused by the tonal contrast rising higher initially and staying higher through the following decline in HNR towards syllable termination. The regular tokens also see an initial rise but peak lower and stay lower throughout the decline towards the end of the syllable, suggesting more laryngeal noise in the signal for the regular tokens compared to the tonal tokens, although the differences in peak height could mean the relative post-peak fall in HNR is similar between the stød types but with a difference in the absolute height at which it happens. As the shape term model comparison suggested a significant difference, however, this appears not to be the case as shape difference is clearly a factor too.

4.1.2.6. Subharmonic-to-Harmonic Ratio

Table 15 reports effects of stød type on the SHR trajectory, first for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this difference is significant, effects are reported for a model comparison between the full model and another nested model excluding only the smooth term for stød type by normalised time to specifically test for the effects of stød type on SHR trajectory shape.

Comparison	χ^2	*df*	$p(\chi^2)$
AAR non vs regular stød			
Overall	12.87	3	< 0.001
Shape	9.20	2	< 0.001
AAR regular vs tonal stød			
Overall	10.02	3	< 0.001
Shape	6.06	2	0.002

Table 15. Effects of stød type on SHR trajectory for the Aarhus dialect. The overall effect represents the difference in height, the shape effect represents the difference in shape.

The effect of stød type on the SHR trajectory is significant for the Aarhus dialect in the model comparison. Both overall height and shape differences were found for the non-stød vs regular stød contrast. The GAMM smooths are visualised below:

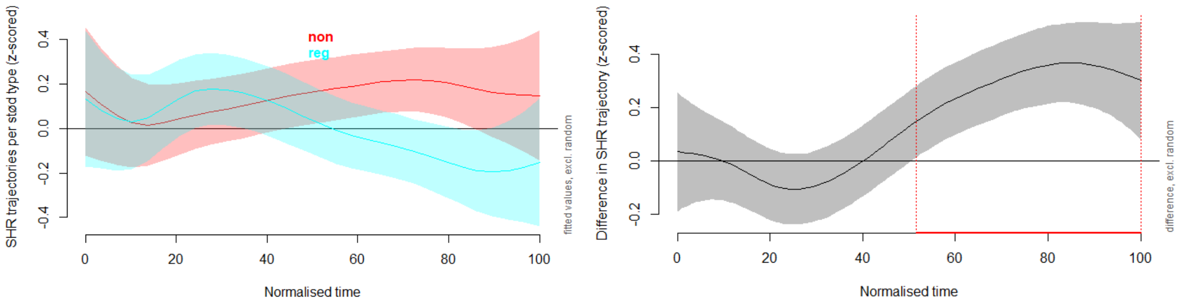

Figure 43. GAMM fit (left) and difference smooth (right) of the effect of stød type (non vs regular) on the SHR trajectories for the Aarhus speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

In interpreting the SHR outcome, it should again be kept in mind that this parameter was poorly tracked and thus less reliable. The data that *is* tracked, however, reveal a difference in the non-stød vs regular stød minimal pair contrast. This difference is mainly due to the trajectory height and shapes after the 50% timepoint as evident in the difference smooth. The non-stød mean smooth stays relatively high whereas the regular stød mean smooth declines until around the 95% timepoint. This creates a discrepancy between the SHR values, but in the opposite direction of what would be expected – the non-stød tokens have a higher SHR than the regular tokens, indicating period-doubling on the tokens expected to be produced with modal phonation. It can be speculated that this may be an artefact of the distribution of tracked data but this is difficult to confirm. The confidence intervals for both smooths are very large and this variability is more than likely influenced the variability in tracking. The Aarhus speakers had lower density peaks over 70% of tracked data points per token compared to the Copenhagen data, meaning the SHR tracking was even poorer for the Aarhus dialect. As such, interpreting the trends in the data is difficult. Nonetheless, GAMMs were also fitted to the regular vs tonal stød contrast and these are visualised below:

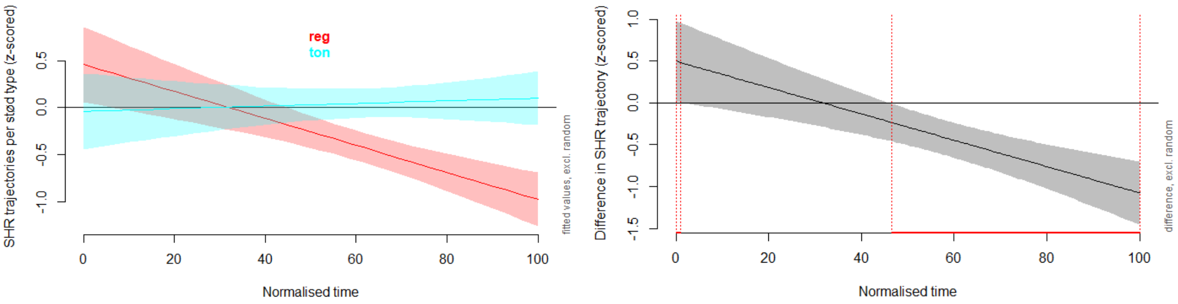

Figure 44. GAMM fit (left) and difference smooth (right) of the effect of stød type (regular vs tonal) on the SHR trajectories for the Aarhus speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

The model comparison found significant differences between the regular and tonal stød contrast which are due to the trajectories diverging just before the 50% timepoint where the regular stød token mean smooth steeply declines whereas the tonal stød token mean smooth stays level and higher. If the data is useful at all, this indicates more subharmonics for the tonal stød condition compared to the regular stød condition which is the opposite of what is hypothesised.

4.1.2.7. Classification of Aarhus laryngealisation subtypes

Like with the Copenhagen speakers, this section seeks to establish the correlation between the main acoustic measures of the different types of laryngealisation outlined in Keating et al (2015) and the results of the statistical model predictions presented in the previous sub-sections. The purpose is to explore the phonetic variation in stød production. The first two types of laryngealisation were described as follows, reposted again for convenience:

(i) Prototypical creaky voice has a low f0, irregular vibrations correlating with high noise and glottal constriction correlating with a low H1-H2.

(ii) Vocal fry has a low f0, glottal constriction correlating with a low H1-H2 and damped pulses, correlating with low noise, i.e. a high HNR.

First it should be established whether the Aarhus speakers had stød realisations produced with both a low f0 and a low H1-H2, and only if so is it relevant to further distinguish the two subtypes by correlates of noise. For f0 there were unexpected observations for the non vs regular stød contrast where the regular stød f0 trajectory was very high in the initial part of the syllable compared to the non-stød tokens. In the latter part of the regular stød syllable it did decline moderately, however, indicating lowering of f0. The regular vs tonal contrast both had moderately declining f0 trajectories in the latter syllable part too with no significant difference between the two trajectories. As such, f0 can be said to be relatively lowered during the production of the stød. As prototypical creaky voice can be expected to have an f0 value below 100 Hz (Laver 1980: 122, Davidson 2019: 238-239), it is necessary to look at the raw, non-normalised values to confirm whether the relative decline in the GAMM trajectories can be considered low in absolute terms. The raw means and standard deviations were calculated and are plotted below, segregated by stød type and speaker sex and filtered at the 50-100% timepoint:

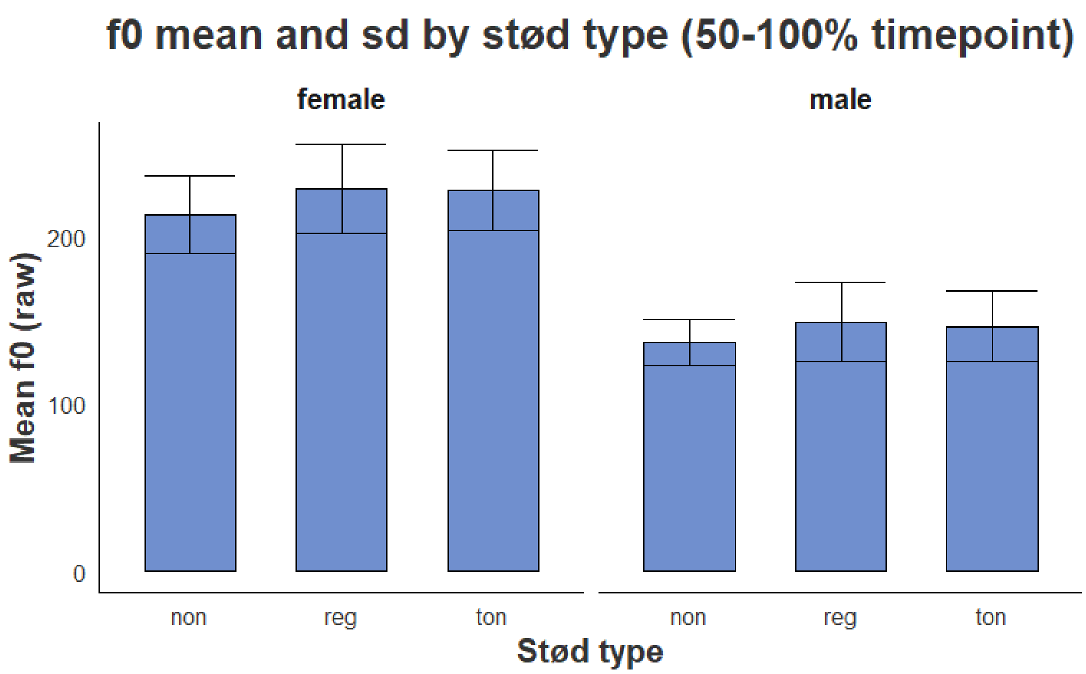

Figure 45. Raw mean f0 values at the 50-100% timepoint for the Aarhus speakers segregated by stød type and sex with the error bars indicating the standard deviation. 'Non' = no stød, 'reg' = regular stød and 'ton' = tonal stød.

The same trend is observed for the Aarhus speakers as was observed for the Copenhagen speakers where the non-stød tokens are slightly lower in mean f0 which is unexpected. The non-stød f0 mean for the female speakers is 208.3 Hz (*n* = 3773) and for the male speakers it is 131.9 Hz (*n* = 4499). For both the regular and the tonal stød, the mean f0 is above 100 Hz, the average being 230.9 Hz (*n* = 5038) for the female regular stød tokens and 231.0 Hz (*n* = 1254) for the female tonal stød tokens. For the male regular stød the mean f0 is 149.5 Hz (*n* = 5995) and for the tonal stød tokens the male f0 mean is 150.2 Hz (*n* = 1485). As these numbers make clear, f0 can not be claimed to be low in absolute terms for this dialect, neither for the regular nor the tonal stød condition.

The second parameter that should be low for the data to comply with the first two subtypes is H1-H2. The raw H1-H2 numbers are expected to be negative if glottal constriction occurs as this causes the second harmonic to become more dominant than the first harmonic. To explore the raw values for this parameter, the non-normalised data was summarised for mean H1-H2 and standard deviation, plotted in the figure below and segregated by speaker sex and stød type:

Figure 46. Raw mean H1-H2 values at the 50-100% timepoint for the Aarhus speakers segregated by stød type and sex with the error bars indicating the standard deviation. 'Non' = no stød, 'reg' = regular stød and 'ton' = tonal stød.

For the female speakers the regular stød tokens are lowest with a mean of 5.21 dB (n = 5038) compared to a mean of 5.83 dB (n = 3773) for the non-stød tokens and 5.68 dB (n = 1254) for the tonal stød tokens. The male speakers have the lowest mean for the tonal stød tokens at 5.67 dB (n = 1485) followed very closely by the regular stød tokens at 5.90 dB (n = 5995) and the non-stød tokens were highest at a 5.92 dB mean (n = 4499). None of the value are close to being negative despite some large standard deviations for the tonal female stød tokens and the male non-stød and tonal stød tokens, respectively. In conclusion, H1-H2 cannot be said to be low for this dialect in any of the stød conditions. Due to both a lack of a low f0 and a lack of a low H1-H2, the two first subtypes can be rejected as a good fit for the stød realisations in Aarhus and inspecting noise levels to differentiate the types is thus irrelevant at this stage.

That leaves three subtypes left to explore. The third one was described as follows:

(iii) Multiply pulsed voice has irregular vibrations correlating with high noise, glottal constriction correlating with a low H1-H2 and subharmonics correlating with a high subharmonic to harmonic ratio (SHR).

The correlates of noise were CPP and HNR. CPP for the Aarhus speakers saw a sharp decline for the regular stød towards syllable termination, ending lower than the non-stød trajectory. This suggests more noise on the regular stød tokens and the differences in trajectories in the GAMMs were statistically significant. For the regular vs tonal contrast, the regular stød also had a steeper decline ending lower than the tonal stød tokens, indicating that the regular stød condition is produced with more noise and to a statistically significant degree. The GAMM fits for HNR showed a steeper decline towards syllable termination for the regular stød tokens compared to the non-stød tokens, causing the regular stød trajectory to end lower, indicating higher degree of noise on these which was statistically significant in the model comparison. The regular vs tonal stød pair also both had a syllable-final decline in HNR indicating increased noise but with the tonal stød tokens being higher than the regular tokens. This suggested more noise on the regular tokens and the difference was statistically significant. As such, the first criterion for subtype three, multiply pulsed voice, is fulfilled as all stød conditions had increased noise. The criterion of a low H1-H2, however, is not fulfilled, cf. the previous two subtypes, and the SHR measurements were higher for the non-stød tokens compared to the regular stød tokens, indicating more period doubling in the absence of stød. This was an unexpected finding as it is difficult to make strong conclusions on the basis of the current data as SHR was generally tracked poorly. It can be concluded, however, that the stød tokens in Aarhus are not compatible with the description of subtype three.

Moving on to the last two subtypes, the description of the fourth subtype was as follows:

(iv) Aperiodic creak lacks a low f0 but has irregular vibrations correlating with high noise and glottal constriction correlating with a low H1-H2.

The lack of low f0 applies to the stød data in Aarhus as the regular stød tokens had an unexpectedly high f0 compared to the non-stød tokens. Further, for both the regular and tonal stød, the absolute f0 values showed means that were too high to be classified as low. As such, the first criterion is fulfilled. For the second criterion, high noise should yield a low CPP and HNR, both of which were relatively lower when the stød was present, and with the regular stød tokens having a lower CPP and HNR in both minimal pair contrasts. However, the H1-H2 was not low for the Aarhus dialect during either stød type when inspecting the raw mean values, suggesting a lack of glottal constriction. Thus, based purely on the acoustic analysis, aperiodic creak does not seem to be prevalent in the Aarhus stød data despite a non-low f0 and high(er) noise.

The last subtype of creaky voice defined by Keating et al (2015) was the following:

(v) Non-constricted creak which has a low and irregular f0 correlating with high noise but no indication of glottal constriction, i.e. non-low H1-H2.

This subtype complies with the Aarhus data in having high(er) noise for the regular and tonal stød tokens as indicated by the declining CPP and HNR and having a non-low H1-H2. As discussed for the Copenhagen dialect earlier, the subtype is described as having a low and irregular f0 *correlating* with high noise which suggests that a low f0 must result in high noise. For the Aarhus data, f0 is not low but noise is still high(er). Again, it can be argued that the criterion is fulfilled if one interprets the main classifier as high noise despite not being the consequence of a low f0 but it could also be argued that the criterion is not fulfilled if a low f0 is weighted higher than the proposed consequence of high noise. If the latter interpretation is adopted, the stød types in Aarhus do not fit neatly into any of the categories of subtypes described in Keating et al (2015).

Intensity was also measured for the Aarhus dialect and has been found to correlate with the stød in previous research on Modern Standard Copenhagen (Riber Petersen 1973, Fischer-Jørgensen 1989), making it relevant to explore whether this is true for Aarhus too. The non-stød trajectory and the regular stød trajectory both ended low and at similar level but the regular stød trajectory had a much higher initial peak before the onset of the decline, creating large differences in intensity early in the syllable. The regular vs tonal stød contrast had a high peak and steep decline for the regular stød and a small peak and small, moderate decline for the tonal stød. Both minimal pair contrasts were significantly different in the model comparisons. The decline in intensity is thought to correlate with irregular vocal fold vibrations but it is noticeable that these effects are not reflected in the f0 trajectories. However, since noise was also increased on the stød tokens, it might be that there are effects that affect noise and intensity more than f0, just like was seen for the Copenhagen dialect earlier.

The final point of interest was to explore whether there is evidence of laryngealisation on the tonal stød in Aarhus, or if this is realised purely through tone as suggested by Kyst (2004). The relevant acoustic correlate of tone is f0 which was expected to decline throughout the syllable if the stød is realised as a falling tone. However, since laryngealisation is also likely to correlate with a decline in f0 this must be supported by measures that indicate a lack of laryngealisation. This can be difficult as there are many different types of laryngealisation and they do not all share the same acoustic correlates. In the context of this study, however, it was expected that if Kyst's (2004) observations apply to the data, the tonal stød tokens in Aarhus would have less noise indicated via a higher CPP, H1-H2 and HNR compared to the regular stød tokens and a falling f0. The GAMM smooths for these measures for the regular vs tonal contrast are reprinted below for convenience:

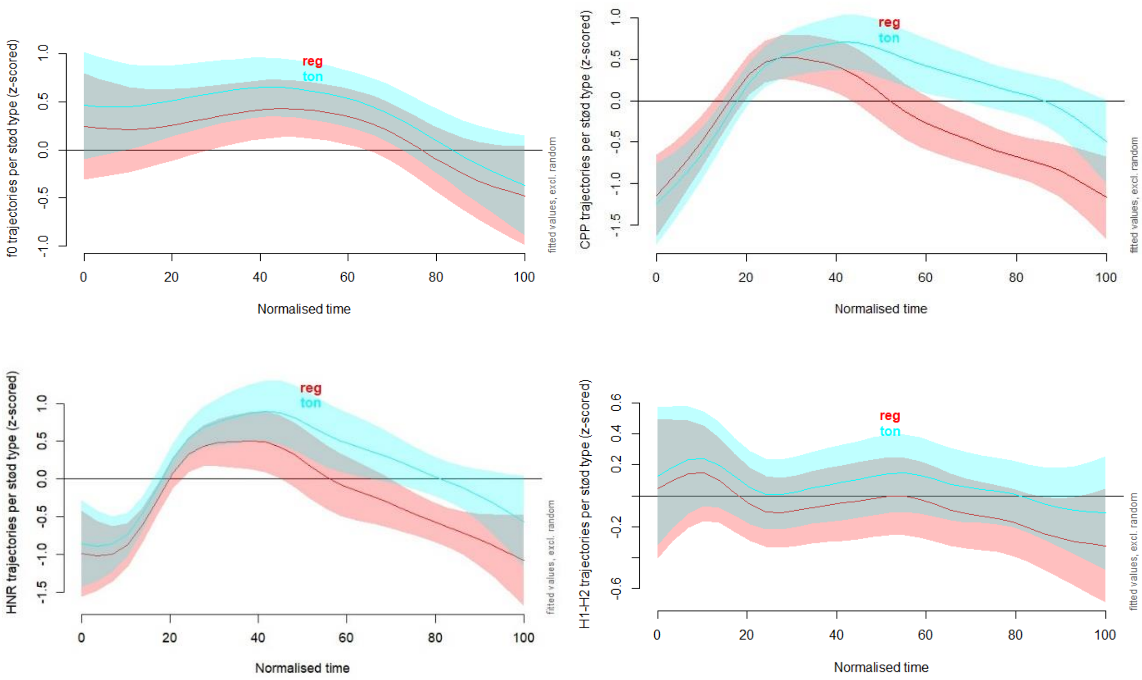

Figure 47. GAMM smooths for the Aarhus regular (red) vs tonal (blue) stød token trajectories for acoustic measures f0 (top left), CPP (top right), HNR (bottom left) and H1-H2 (bottom right) with mean values indicated by the full lines and confidence intervals indicated by the shaded margins surrounding the lines.

The f0 trajectories for both stød types do exhibit a small peak and a subsequent decline but in a moderate range considering the expectation of a two-toned auditory impression. The two stød types look similar in shape but with a small height difference, the tonal stød being higher but the difference was not statistically significant in the model comparison. This can mean that either the regular stød is tonal or the tonal stød is regular. For CPP the GAMM smooths shows how both have an initial rise, but the regular stød declines sooner and steeper just before the 40% timepoint and the tonal stød declines later and less steeply from around 50%. The later onset in decline means that the tonal trajectory is generally higher. This suggests more noise on the regular stød tokens as CPP is lower. The other measure of noise, HNR, indicated a similar pattern as the CPP – the regular stød tokens are lower and declines sooner and steeper than the tonal stød tokens. Differences in CPP and HNR were statistically significant in the model comparisons, suggesting that the regular vs tonal stød contrast in Aarhus have relevant differences in noise in their production. They do not, however, look as expected for a tonal stød realisation as the HNR and CPP would be expected to decline relatively less due to a lack of increased noise. It can be noted, however, that the non-stød tokens for HNR were somewhat level – for CPP, they also exhibited a decline, making the observations for the tonal stød a bit less conclusive. For H1-H2 the regular and tonal stød realisations were very similar, the tonal stød being slightly higher but with lots of trajectory overlap and no statistically significant difference in the model comparison. This indicates similar levels of glottal constriction on the two stød types. Looking at the absolute values they were positive, meaning a lack of glottal constriction. What causes the introduction of noise on these stød tokens is thus not entirely clear.

To summarise, the tonal stød in Aarhus did not have the expected outcomes in the GAMM trajectories due to moderate f0 range and an increase in noise throughout the syllable. The H1-H2 measure did, however, indicate a lack of laryngealisation and the noise measures for the tonal stød were higher than the regular stød tokens in the GAMM smooth trajectories, indicating less noise in the signal for the tonal tokens despite a decline. In the following sections, the articulatory data is analysed, adding to the acoustic results presented here and adding detail to the differences between the stød types which will aid in answering the question of how much phonetic variation occurs during the stød in the two dialects.

To summarise the acoustic analysis for Aarhus so far, it was found that some measures in Aarhus behaved as expected, e.g. a decrease in intensity during the regular stød and an increase in noise evident in lower CPP and HNR measures for the regular stød relatively to the non-stød tokens. Unexpected findings included a much higher f0 for the regular stød compared to the non-stød tokens, despite both stød types having moderate, relative f0 declines. Other findings were that the tonal stød in Aarhus was not significantly different to the regular stød in terms of f0 but that it did have less noise despite a relative decline in CPP. The H1-H2 measurement indicated similar levels of glottal constriction on the regular and tonal tokens and the raw numbers indicated no glottal constriction on either.

4.2. EGG measurements

4.2.1. Copenhagen

Having presented the results of the acoustic analysis the attention is now turned to the EGG measurements. Recall that all EGG data points are included in the model but that the number of cycles tracked differed between speakers. With this in mind, effects for the Contact Quotient (CQ) values are reported in Table 16 below; first for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this effect is significant, effects are reported for a model comparison excluding the smooth term for stød type by relative time to specifically test for the effects of stød type on CQ trajectory shape differences.

Comparison	χ^2	*df*	$p(\chi^2)$
CPH non vs regular stød			
Overall	41.994	3	< 0.001
Shape	22.028	2	< 0.001
CPH regular vs tonal stød			
Overall	0.939	3	0.598
Shape	N/A	N/A	N/A

Table 16. Effects of stød type on CQ trajectory for the Copenhagen dialect. The overall effect represents the difference in height, the shape effect represents the difference in shape.

There is a significant difference between CQ values for the non-stød vs regular stød tokens for both height and shape of the CQ measurement trajectories. The nature of the difference is visualised below:

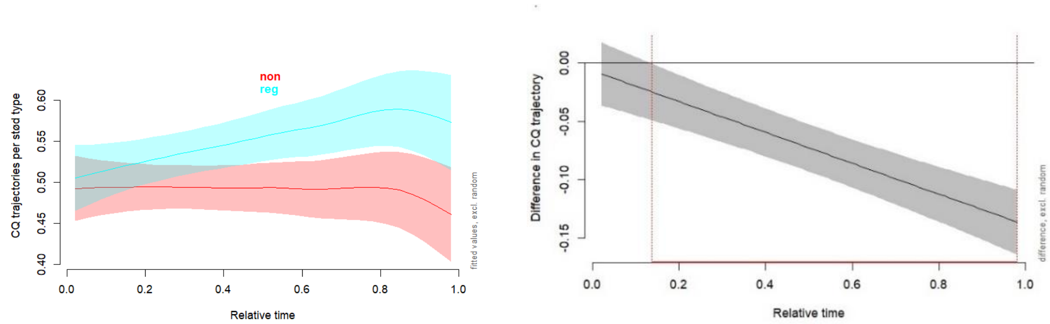

Figure 48. GAMM fit (left) and difference smooth (right) of the effect of stød type (non vs regular) on the EGG-derived CQ trajectories for the Copenhagen dialect with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths (coloured for the GAMM fits, grey for the difference smooth). The red lines on the difference smooth plot indicate the area where the two trajectories differ most.

As Figure 48 shows, the non-stød tokens have a relatively flat CQ trajectory with a minor decline in the last 10% of the progression over time. The non-stød CQ values hover just below 0.50 indicating modal phonation and similar lengths of the closing and opening phases. The CQ values dip slightly from the 80% timepoint, ending around 0.45, possibly indicating the introduction of a slightly breathier voice quality. In contrast, the regular stød tokens see the CQ values gradually rise throughout the duration of the syllable ending on a high just below 0.60, indicating that each glottal cycle has a longer contact phase relative to the de-contacting phase. This could correspond to increased laryngealisation on the regular tokens but with the caveat that prototypical creaky voice would be expected to have a higher CQ up until around 0.90 for each glottal cycle. This may be the result of lumping together all tokens for all speakers where some individual speakers may have longer contact phases than others, i.e. laryngealisation realised more like a prototypical creaky voice whereas others may have laryngealisation on the lower end towards the middle of the non-modal, hyperfunctional continuum. It may also be, however, that the laryngealisation during the stød is simply different from prototypical creaky voice, a very likely explanation based on the acoustic analysis above, meaning values cannot be expected to be similar to prototypical values of a long closing phase. Overall, however, there is an observed significant effect of the presence of the regular stød on the CQ values indicating that this stød type affects vocal fold patterns in the dialect of Copenhagen.

For the regular vs tonal stød contrast there was no significant difference in CQ values, indicating that these types are produced with similar vocal fold contact patterns. Inspecting the GAMM smooth trajectories in a plot it does appear that the tonal tokens have a slightly higher mean CQ:

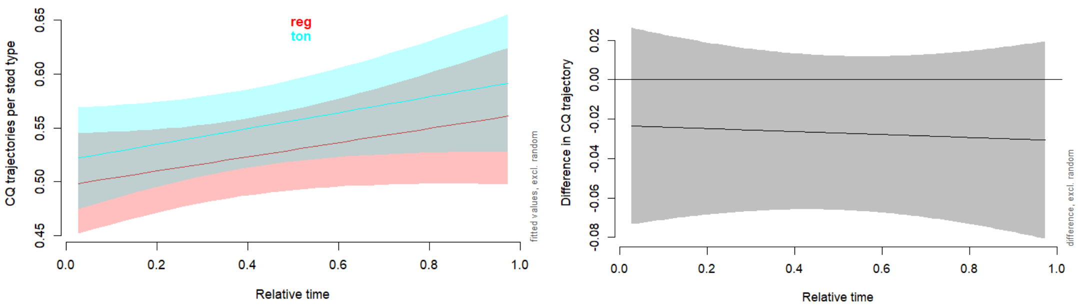

Figure 49. GAMM fit (left) of the effect of stød type (regular vs tonal) on the EGG-derived CQ trajectories for the Copenhagen dialect with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the coloured smooths. The difference smooth (right) shows no red lines as no significant difference between trajectories was detected.

As the confidence intervals show in Figure 49 there is considerable overlap between the CQ values for these stød types so although the tonal stød tokens have a higher CQ this cannot be statistically distinguished from the CQ values of the regular tokens. Both types have CQ values that start around 0.50 and rise throughout the syllable, indicating continuously increased laryngealisation with CQ values ending just below 0.60 for the tonal tokens. As just noted, however, these values are not particularly high despite indicating a slightly longer closing phase. An interesting point here is just how much of a relative increase in CQ values create a relevant perceptual difference and how these values compare to what can be expected for phonologically contrastive laryngealisation in other languages. The discussion chapter will return to this issue later; for now, the EGG analysis for the Aarhus dialect in presented.

4.2.2. Aarhus

Effects for CQ values for the Aarhus speakers are reported in Table 17 below; first for a model comparison of overall effect of stød type (full model with a parametric term and a difference smooth vs a nested model excluding the parametric term and the difference smooth) and if this effect is significant, effects are reported for a model comparison excluding the smooth term for stød type by normalised time to specifically test for the effects of stød type on CQ trajectory shape.

Comparison	χ^2	*df*	$p(\chi^2)$
AAR non vs regular stød			
Overall	2.842	3	0.128
Shape	N/A	N/A	N/A
AAR regular vs tonal stød			
Overall	1.551	3	0.376
Shape	N/A	N/A	N/A

Table 17. Effects of stød type on CQ trajectory for the Aarhus dialect. The overall effect represents the difference in height, the shape effect represents the difference in shape.

The difference between the CQ values for both minimal stød pairs is not significant, indicating that the phonation type with which they are produced is relatively similar. However, the statistical model does not reveal whether the phonation type is modal or non-modal so it is useful to inspect the GAMM fits by plotting the value trajectories:

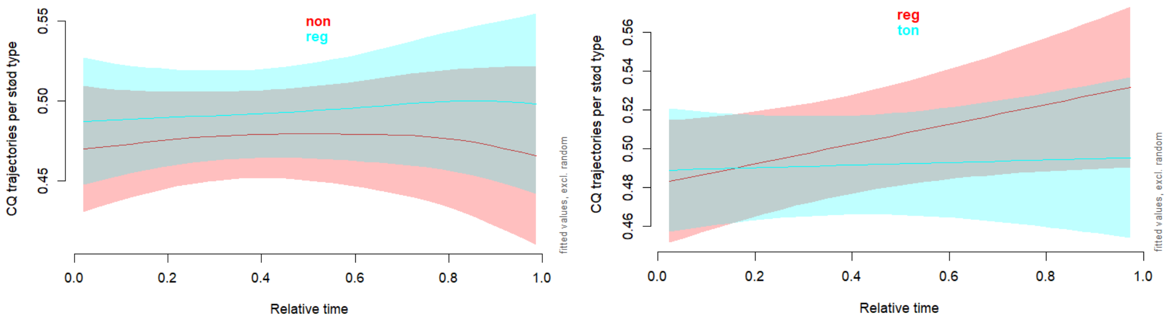

Figure 50. GAMM fits of the effect of stød type, non vs regular (left), regular vs tonal (right) on the CQ trajectories for the Aarhus speakers with the mean smooth indicated by the full line and the 95% confidence intervals indicated by the shaded areas around the smooths.

As the GAMM smooths show the trajectories do differ in some ways despite the lack of significance in the difference model. For the non-stød vs regular stød contrast the CQ trajectories are similar in shape but the regular tokens are slightly higher throughout. Laryngealisation is not indicated, however, as they hover between a CQ of 0.45-0.50 associated with modal phonation where the vocal fold open and closing phase are roughly equal in length. The non-stød tokens are slightly lower on average, indicating a trend towards a slightly breathier phonation on these tokens, but to a very small degree that is unlikely to be phonologically relevant for contrasts in laryngealisation generally correlating with higher rather than lower CQ values.

For the regular vs tonal intervals, the trajectories are increasingly different throughout the syllable. The tonal tokens stay flat, hovering just below 0.50 whereas the regular tokens start slightly lower but then see a steady, linear increase in CQ until ending at around 0.53, slightly higher than modal phonation. As indicated in Table x.x.1. this difference is not statistically significant and a CQ of 0.53 is very moderately non-modal compared to a vocal fold contact pattern of up to 0.90 that might be present for a prototypical creaky voice (Esling et al 2019: 63-64). Again, the issue becomes what degree of increase in CQ is perceptually and articulatorily relevant for the presence of laryngealisation, a question that will be returned to in the discussion chapter. Most importantly for answering whether the tonal stød in Aarhus is laryngealised to any degree, the CQ values show that the tonal stød stays flat around 0.49. This suggests an absence of laryngealisation on these tokens which supports the auditory impression of a purely tonal realisation but raises questions about what articulatory mechanisms, if not constriction at the glottis, contributed to the CPP and HNR declining in the acoustic analysis, correlated with more noise in the signal. Further, the CQ values indicate that neither stød type, not even the regular stød tokens, are laryngealised as the trajectories all hover around values for modal phonation except for the syllable-final values for the regular tokens in the regular vs tonal contrast. In conclusion, the EGG data analysis for the dialect of Aarhus suggests that the tonal stød is produced with modal phonation but that this is also true for the regular stød and it questions the source of noise suggested in the acoustic analysis.

4.2.3. Interim summary

The aim of this chapter was to answer research question 1 of how much phonetic variation occurred in the production of the phonologically contrastive stød. The chapter presented an analysis of the phonetic variation using acoustic and articulatory data from speakers originating in Copenhagen and Aarhus to provide a novel perspective on intra-language phonetic differences in laryngealisation. The reference point for characterising the variation was the subtypes of laryngealisation outlined in Keating et al (2015) and the trajectories of the acoustic measures were modelled dynamically throughout the phonological stød basis using GAMMs. Both dialects had unexpected results for f0 as this trajectory was higher for the regular stød tokens compared to the non-stød condition, thus rejecting the hypothesis that f0 would fall during the stød due to laryngealisation on the stød. Raw measures indicated a lack of glottal constriction for both dialects as H1-H2 values were positive but both dialects did have an increase in signal noise correlating with a decreased HNR and CPP for stød tokens. Increased signal noise is suggestive of laryngealisation. Both dialects had stød types that were most similar to Keating et al's (2015) subtype five, non-constricted creak, but were lacking fulfilment of the criterion of a low f0, meaning none of the subtype characterisations were completely compatible with the phonetic manifestation of the stød, despite the stød being described as a kind of creaky voice. SHR measures did not indicate any period doubling during the stød in any dialect, unlike Hansen (2015) who found diplophonia on most of their stød tokens, but the measures were fairly poorly tracked, so these results are not highly weighted in the interpretation of the analysis. Both dialects also saw a decline in intensity during the stød, a finding compatible with previous studies of the stød in Copenhagen (Riber Petersen 1973, Fischer-Jørgensen 1989, Hansen 2015) that found this measure the most consistent correlate of the production of the stød. For both dialects, the regular stød had the lowest intensity. This parameter was not part of the subtype classification, however.

The dialects were expected to be dissimilar in the regular vs tonal minimal pair – this pair was hypothesised to only differ noticeably in the Aarhus dialect due to a tonal realisation of the stød in closed monosyllables with a long vowel whereas for Copenhagen, the stød types were expected to be realised as the same kind of voice quality change on the hypercompressed continuum. Contrary to this hypothesis, the GAMMs showed statistically significant differences between the regular and tonal stød types in Copenhagen in the intensity, CPP and HNR model fits. The tonal stød in Aarhus was not significantly different to the regular stød in terms of f0 and H1-H2 but it was produced with less noise according to differences in CPP and HNR which were statistically significant in the model comparisons with the tonal stød having relatively less noise – the CPP did, however, decline relatively throughout the tonal stød trajectory.

For the EGG measurements, the Copenhagen dialect data confirmed the hypothesis that the contact quotient during the stød would be higher. The non-stød tokens hovered just below 0.50 whereas the regular stød tokens gradually increased during the stød basis ending just below 0.60. For the regular vs tonal contrast both trajectories also increased in CQ values during the stød basis, ending just below 0.60 again. These values indicate increased vocal fold contact area in each glottal cycle for the Copenhagen stød. For the Aarhus speakers, the differences in CQ values were not significant in the model comparisons for neither minimal pair contrast. The non-stød and regular stød trajectories hovered between 0.45-0.50 with the non-stød tokens having a slightly lower mean smooth. The regular vs tonal tokens were also slightly different, the tonal token smooth staying flat around 0.50 but the regular stød tokens increasing to around 0.53. The CQ measures thus indicate longer closing phases during the Copenhagen stød compared to the Aarhus stød. The tonal stød in Aarhus did not have CQ values indicating any laryngealisation, indicating its realisation is achieved via tone only, as suggested by Kyst (2004) – however, the acoustic measures were less conclusive in this regard as tone should not increase noise and HNR and CPP were declining.

An interim conclusion of research question 1 is that the dialects had relatively little phonetic variation between them on the acoustic parameters during the regular stød; both had a high(er) f0, lower intensity, increased signal noise and a lack of glottal constriction. There were, however, evidence of phonetic variation within the dialects as the regular vs tonal stød pair were significantly different in the model comparisons for intensity, CPP and HNR in Copenhagen and intensity, CPP, HNR and SHR for the Aarhus dialect. This indicates that the acoustic correlates of the stød changes with the stød type which creates phonetic variation within each dialect. It should, however, also be noted that this is likely to be influenced by the differences in stød basis, the regular stød having a short vowel and a consonant but the tonal stød having a long vowel as the stød basis. The analysis of the EGG measures showed more variation between dialects than the acoustic measures – both stød types in Copenhagen had increased CQ values reaching 0.60 whereas the stød tokens in Aarhus did not reach more than 0.53 at their highest CQ value. This indicates increased vocal fold contact in each glottal cycle for Copenhagen more than for Aarhus and suggests different laryngeal behaviours for each dialect during the stød, and thus inter-language phonetic variation in the contrast. The phonetic variation is not only related to acoustic values as was the focus of the analysis for research question 1, but also to the relative importance of each measure and in the timing of the variation. Research question 2 and 3 asked more specifically about these aspects and are answered in the following chapters.

5. Which measurements correlate with (subtypes of) the stød?

This chapter is focused on research question 2, namely determining which acoustic measurements correlate strongest with the stød in each dialect along with determining which acoustic measurements correlate strongest with the vocal fold contact patterns derived from the EGG signal. This is analysed to offer a perspective on intra-language variation on acoustic correlates of laryngealisation and more generally to explore how the correlation between laryngealisation and acoustic variables compare cross-linguistically in languages containing different types of hyper-compressed voice qualities used for phonological distinction. There are also more methodological interests for answering this research question which are motivated by the LAM framework. If the production of laryngealisation involves the entire larynx rather than just the vocal folds, the question remains how this translates to articulatory measures such as EGG which is widely used as a proxy for glottal activity but might fail to capture epiglottal activity. As such, in correlating the acoustic parameters with the EGG data, this might be a step on the way to explore what the relationship is between laryngeal, not just glottal, activity in different types of contrastive laryngealisation, EGG readings and acoustic correlates, both within and across languages.

As the GAMMs fitted above were focused on the differences in measures that characterise the minimal pair contrasts, the methods used to test the variable correlations here will do the same. This means that what is tested is which variable correlate strongest with the differences observed between no stød vs regular stød and regular stød vs tonal stød, respectively.

A simple and commonly used method to compute variable importance is to fit a logistic regression model to each acoustic measurement and then compare models to find the lowest Akaike Information Criterion score which would indicate the best fitting model and thereby the most important acoustic correlate to capture stød type differences in laryngealisation. However, since the acoustic correlates are all intended to measure the same thing, laryngealisation, they are not independent which violates the assumption of a regression model and there is likely to be a high degree of collinearity between variables (for more on the issue of collinearity and regression models, see Winter 2019: 112-115). For this reason a better suited model is needed where the predictor variables are assumed to be collinear, not independent. A good option for this are random forest models. Random forest models can, like regression models, be used to predict which variables are most probable or important based on a set of predictors but are, unlike regression models, able to handle complex interactions and high levels of collinearity between predictor variables (Strobl et al 2008). Fitting random forests is done via an algorithm that constructs so-called conditional inference trees that estimate the likelihood of a given outcome variable (and its levels) on the basis of binary questions about the predictor variables that are asked recursively throughout a process of sub-sampling the data into inference trees based on a null hypothesis of independence until an additional subset is no longer justified. A further subset is justified, and a predictor thus useful, as long as the null hypothesis of independence is rejected. For more details on random forests and a demonstration of how they can be used for linguistic data specifically, see Tagliamonte & Bayeen (2012).

Random forests were fitted using the *party* package in R (R Core Team 2022). The *party* package contains the *Cforest* function which is an upgrade of a previously used method by Breiman (2001) which has since been shown to overestimate the importance of predictor variables with collinearity (Strobl et al 2008). The *Cforest* function improves this via a conditional permutation scheme more robust against inflation of the variable importance which is important for the current data due to expectations of relatively high collinearity of predictors. The following sections describe the methods for fitting the models to the acoustic and articulatory data. The outcomes of the models are then analysed and presented in the context of research question 2.

5.1. Acoustic variable importance

The acoustic data was split into four different sets, two for each dialect containing (i) the non-stød and regular stød contrast tokens and (ii) the regular vs tonal stød contrast tokens. This split enabled the model to find the best predictor variable fit for the difference between the respective stød types as was previously done for the GAMMs. The NAs were filtered out from each data set to enable the model to converge. A random forest model was fitted to each subset with stød as the outcome variable and the normalised values for acoustic parameters f0, intensity, CPP, HNR and H1-H2 as predictor variables for the conditional variable importance. SHR was not included in the models because the data was poorly tracked and thus rather inconclusive. To visualise the relative importance of each acoustic variable, the model outcomes are plotted below for each minimal pair contrast:

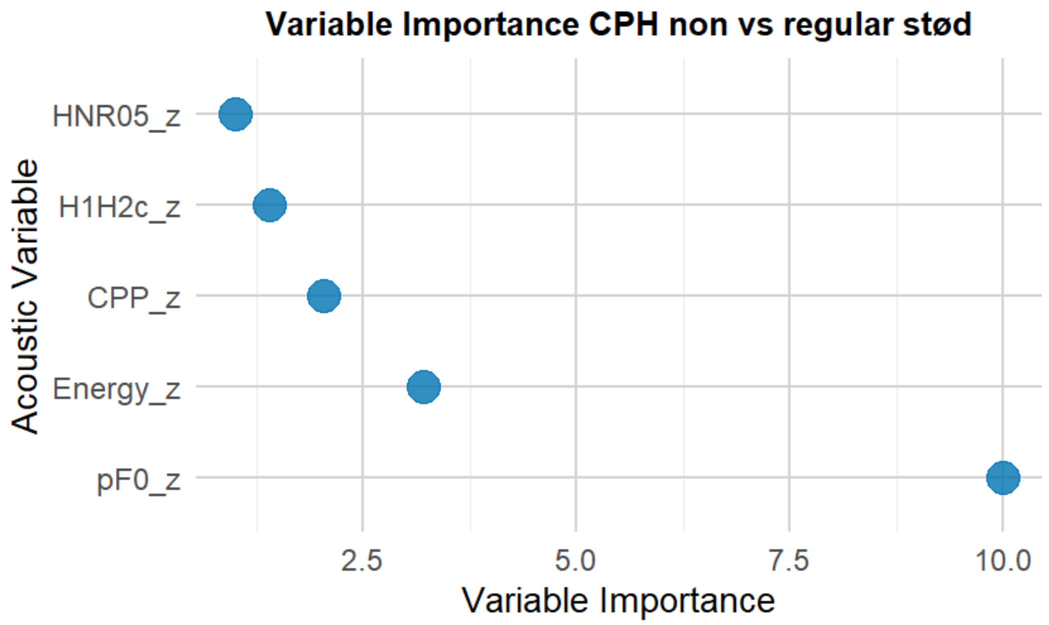

Figure 51. Acoustic variable importance for the Copenhagen non vs regular stød contrast with scores expressed on a scale of 1-10 where 1 = least important and 10 = most important.

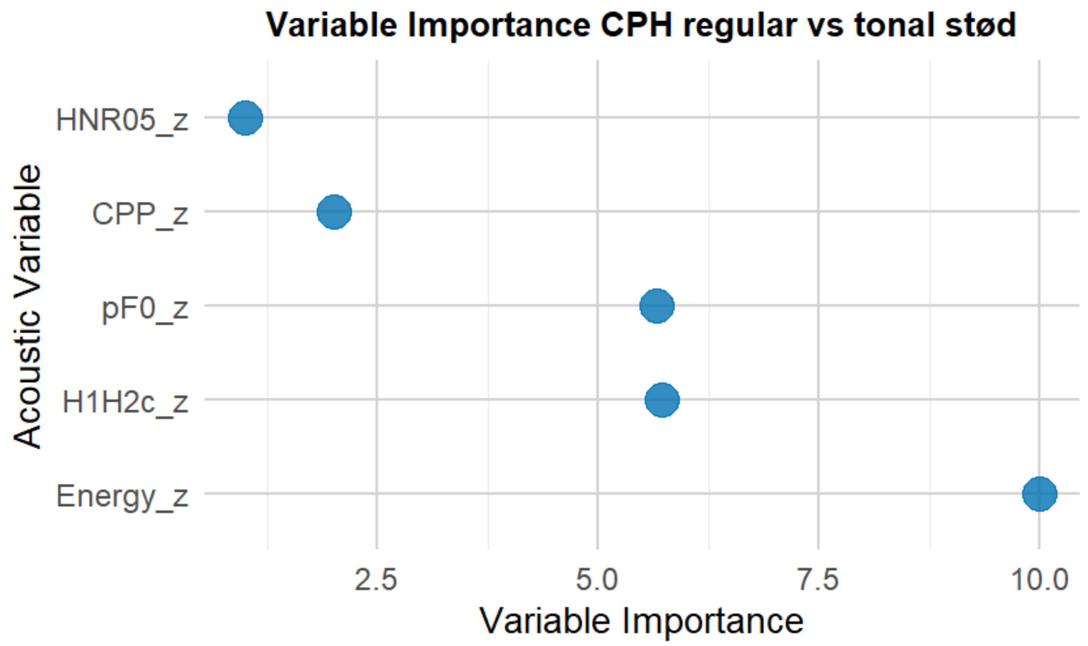

Figure 52. Acoustic variable importance for the Copenhagen regular vs tonal stød contrast with scores expressed on a scale of 1-10 where 1 = least important and 10 = most important.

For the Copenhagen dialect the variable importance changes depending on the stød contrast, indicating some phonetic variation in the production. For the non vs regular stød contrast, the main predictor variable for the difference is f0. Intensity (coded in the plot as Energy_z) is the second most important predictor variable at just over 2.5 out of 10 on the variable importance continuum. The remaining measures are all below the 2.5 mark and seem to contribute little to differentiate the two stød types relative to f0 and to a smaller degree intensity. HNR is the lowest ranked predictor variable. It should be noted that the scores do not say anything about the absolute prediction power of the variables, only about their relative importance for the contrast when compared against each other. In the acoustic GAMMs all measures were significant for this contrast which indicated that the presence of the stød impacts all of these acoustic measures compared to no production of the stød. However, the outcome of the random forest model here suggest that all parameters are not equally important despite being significant.

The regular vs tonal stød contrast for Copenhagen has intensity as the highest ranked predictor variable at 10, followed by closely placed predictors H1-H2 and f0 at just over 5 on the x-axis. The three highest ranked variables are somewhat closer to each other on the x-axis for this contrast compared to the non-stød vs regular stød, meaning less prediction strength for just one parameter. HNR is ranked as the lowest predictor variable. In the GAMMs for this contrast, the only two parameters that were not significantly different between stød trajectories in the model comparisons were f0 and H1-H2 which makes it interesting that they rank higher than CPP and HNR in the variable importance as these latter two measures *were* significant in the GAMMs. To understand this, it is helpful to think about how a random forest model determines variable importance. A predictor is valued based on its ability to help make accurate model predictions across its interaction with the different decision trees in the model which means it is not valued based on its independent ability to predict an outcome variable. Essentially, predictors are considered valuable as long as they help reject the null hypothesis of independence as explained above, indicating that they have a meaningful relationship with the outcome variable in the context of the sub-sampled data. This makes is plausible that a parameter that is non-significant in the GAMMs can be ranked higher than one that *is* significant because the GAMMs were fitted for each parameter only and did not include interactions with the other acoustic variables as the purpose was to isolate the effects of each measurement on the stød. For research question 2, however, the approach must compare measures relative to each other as the question is which measurements are primary correlates with each type of stød contrast which is difficult to decide only via isolated effects that are not ranked.

In summary, the two stød types in Copenhagen differ on the variable importance for the two minimal pair contrasts, the non-stød vs regular stød correlating most with f0 and the regular vs tonal stød correlating most with intensity. The following two plots show the outcomes of the random forest models fitted for the same contrasts in the Aarhus dialect speakers:

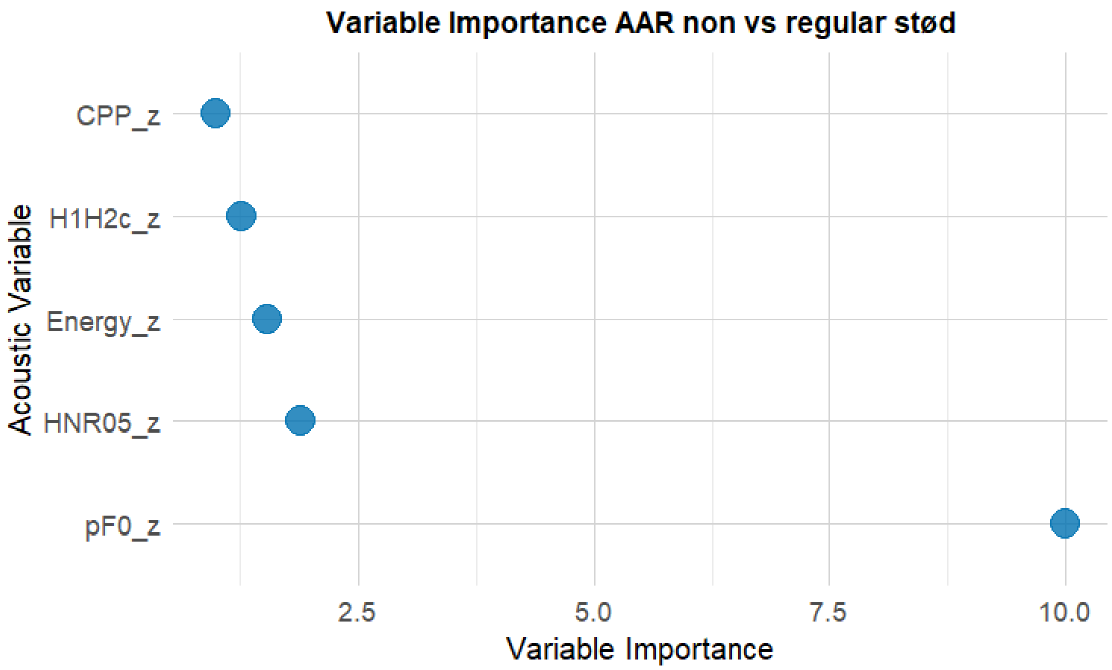

Figure 53. Acoustic variable importance for the Aarhus non vs regular stød contrast with scores expressed on a scale of 1-10 where 1 = least important and 10 = most important.

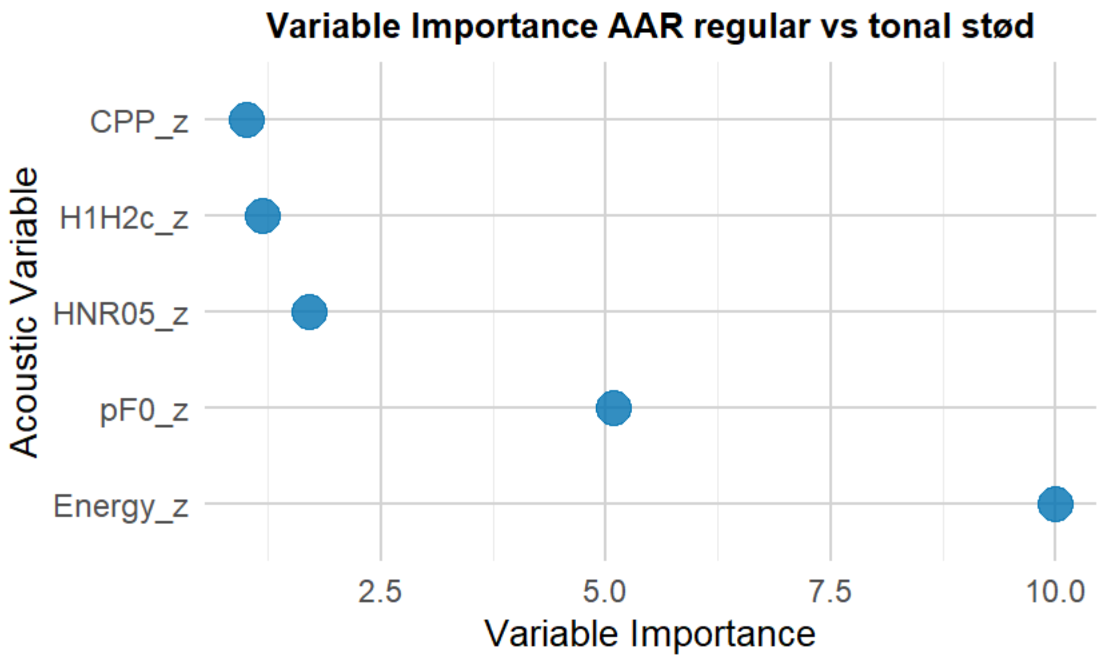

Figure 54. Acoustic variable importance for the Aarhus regular vs tonal stød contrast with scores expressed on a scale of 1-10 where 1 = least important and 10 = most important.

For the non-stød vs regular stød contrast the highest ranked variable is f0 at 10. All the other variables are ranked below 2.5, meaning f0 alone most accurately predicts the difference between stød and no stød here. In the GAMMs f0 did have significant difference in the model comparison but mainly because the regular stød trajectory was higher rather than lower than the non-stød f0. CPP is the lowest ranked variable. The regular vs tonal stød contrast has intensity as the highest ranked variable at 10 with f0 as the second highest ranked variable at just over 5 on the x-axis. The remaining parameters are all ranked below 2.5 with CPP being the lowest. In the GAMMs, the only two non-significant acoustic measures for this contrast were f0 and H1-H2, yet f0 seems to have some prediction accuracy here in the context of the other variables. For a tonal stød realisation one would expect f0 to be the highest ranked parameter, so it is somewhat surprising that it is only ranked second highest. It can be noted, however, that what is being modelled is not the tonal behaviour but the parameters' ability to predict the difference between the two types of stød.

It can also be relevant to consider the differences in stød basis for this contrast in terms of intensity being a good predictor. The non-stød vs regular stød contrast distinguishes words such as *låner* [ˈlɔːnɐ] '(a) borrower' and *låner* [ˈlɔːˀnɐ] 'borrows'. The regular vs tonal stød contrast distinguishes words such as *lund* [lɔnˀ] ‘grove’ and *lån* [lɔːˀn] ‘(a) loan’. As such, the non-stød vs regular stød contrast is the difference between a long vowel and a long vowel with stød. The regular vs tonal stød contrast is one between stød on a sonorant consonant preceded by a short vowel vs stød on a long vowel proceeded by a sonorant consonant. Intensity was overwhelmingly the most important variable for this contrast. This makes it tempting to explain the surprising finding that these two types of stød differ in the dialects by the possible implications of the variable importance of intensity found here: despite the stød being classified as a syllabic, not a segmental, prosody, there is an inherent difference in what segment ‘receives’ the stød in the regular vs tonal stød contrast. The fact that intensity best captures this difference could be explained, at least partly, by the general difference in intensity between vowels and consonantal sonorants overall, irrespective of the presence of the stød. Sonorant consonants generally have weaker amplitude peaks and some dampening of the sound wave (e.g. Johnson 2003). This in itself could influence the variable importance.

To summarise the acoustic variable importance, both dialects had f0 ranked highest for the non-stød vs regular stød contrast and intensity ranked highest for the regular vs tonal stød contrast. Intensity has been shown to be a relatively consistent correlate of the stød (e.g. Fischer-Jørgensen 1989) but these results suggest that f0 is also an important variable in the distinction of presence vs absence of the stød. There was some dialectal variation in the second highest ranked predictors where H1-H2 and f0 were both ranked at a 5 out of 10 for the Copenhagen regular vs tonal stød and the same was found for f0 for the Aarhus regular vs tonal stød. The variables related to signal noise, HNR and CPP, were consistently ranked low in the models.

Having analysed the acoustic variables in relation to each other, the following section analyses the acoustic parameters’ variable importance for the EGG data to explore the relationship between vocal fold contact patterns and acoustic correlates for the stød contrast(s).

5.2. EGG vs acoustic variable importance

To analyse which acoustic measurement correlates the strongest with the vocal fold contact patterns derived from the CQ values the two data sets had to be merged. To do this, the acoustic and EGG data had to be temporally aligned as the EGG data was tracked in cycles per token in relative time whereas the acoustic data was time normalised to 11 dynamic timepoints over the trajectory for each token. The relative time of the CQ values was normalised to 11 timepoints to match the time points in the acoustic data. Time normalisation is only possible if there are at least two values to interpolate between so any tokens with less than two CQ values tracked were filtered from the data. This filtering removed 44 tokens out the 3788 included in the analysis. Time normalising the CQ values does smooth out some details such as a brief change in CQ values for just one or two cycles. However, this should not be detrimental to the observation of general trends in vocal fold patterns during the stød and was deemed an acceptable consequence of the method. A weakness of this method of comparison is that the acoustic data was tracked over the time of the entire stød basis annotation and any non-tracked data was written as 0s. For the EGG data, the tracking was done on the same annotation of the stød basis but what was written to the resulting file was each glottal cycle tracked per token and the relative time of the minimum and maximum EGG and dEGG values within each glottal cycle. This means that the data does not contain any 0s if a glottal cycle was not tracked at any given time point on the annotated interval, the data simply contains the relative time of opening and closing of the vocal folds in each *detectable* cycle. This means that the temporal alignment with the acoustic data is less accurate for tokens where there are a lot of glottal cycles missing. Speakers MCPH2, MPCH3 and MCPH5 had fairly few CQ observations compared to other speakers, meaning the data is somewhat biased towards the influence of the female data for the Copenhagen dialect. With this methodological drawback in mind, the method used to correlate the CQ values with the acoustic parameters was fitting random forest models to test which acoustic variable was the most valuable predictor for the trends in CQ values. This involved subsetting the data into the two respective minimal pair contrasts for each dialect and modelling the outcome variable, i.e. the CQ values, as a function of the five acoustic predictor variables to determine their relative variable importance. The resulting model outputs are plotted for each dialect below:

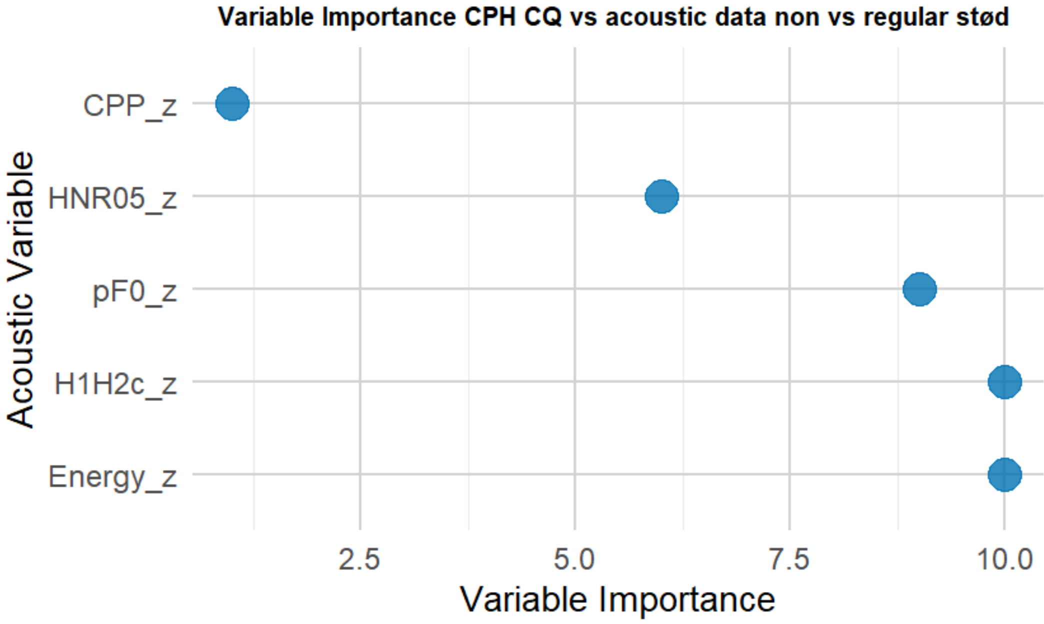

Figure 55. Acoustic variable importance for the Copenhagen CQ values for the non vs regular stød contrast with scores expressed on a scale of 1-10 where 1 = least important and 10 = most important.

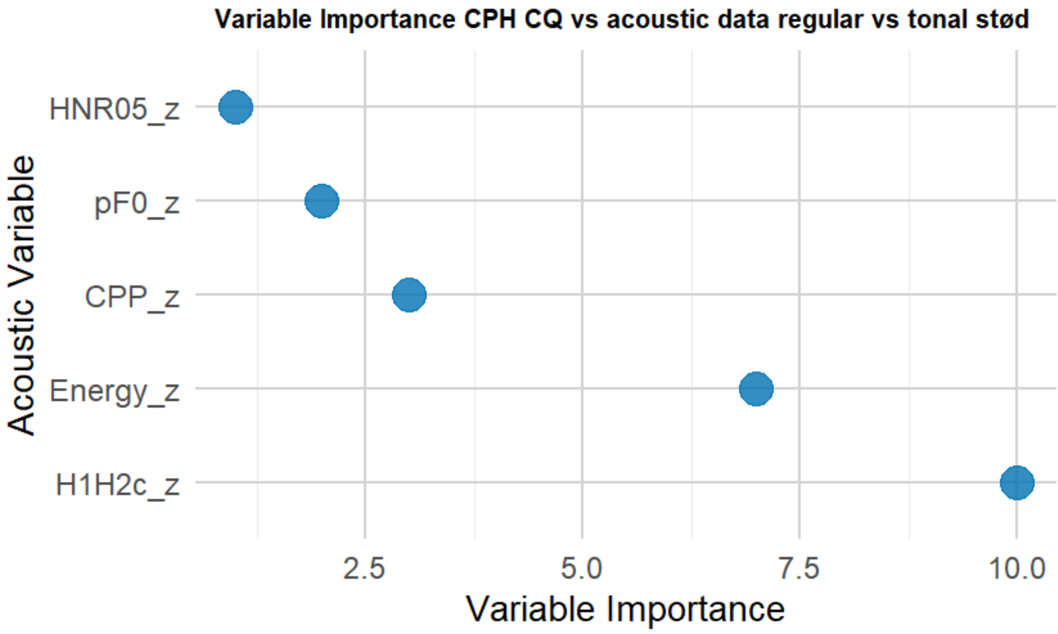

Figure 56. Acoustic variable importance for the Copenhagen CQ values for the regular vs tonal stød contrast with scores expressed on a scale of 1-10 where 1 = least important and 10 = most important.

For the Copenhagen non-stød vs regular stød contrast the highest ranked variables are intensity and H1-H2, equally important at 10 on the x-axis. Two other variables were ranked relatively high, f0 at over 7.5 and HNR over 5. CPP was the lowest ranked variable. For the regular vs tonal CQ values, the highest ranked variable was H1-H2 at 10, followed by intensity just below 7.5 and CPP just over 2.5. HNR was the lowest ranked variable. These results suggest that H1-H2 is a relatively important predictor variable for the differences in vocal fold contact area between both stød types for the Copenhagen dialect. Intensity also seems to contribute well to accurate model predictions.

For the Aarhus dialect the random forest model predictions ranked the variables as follows:

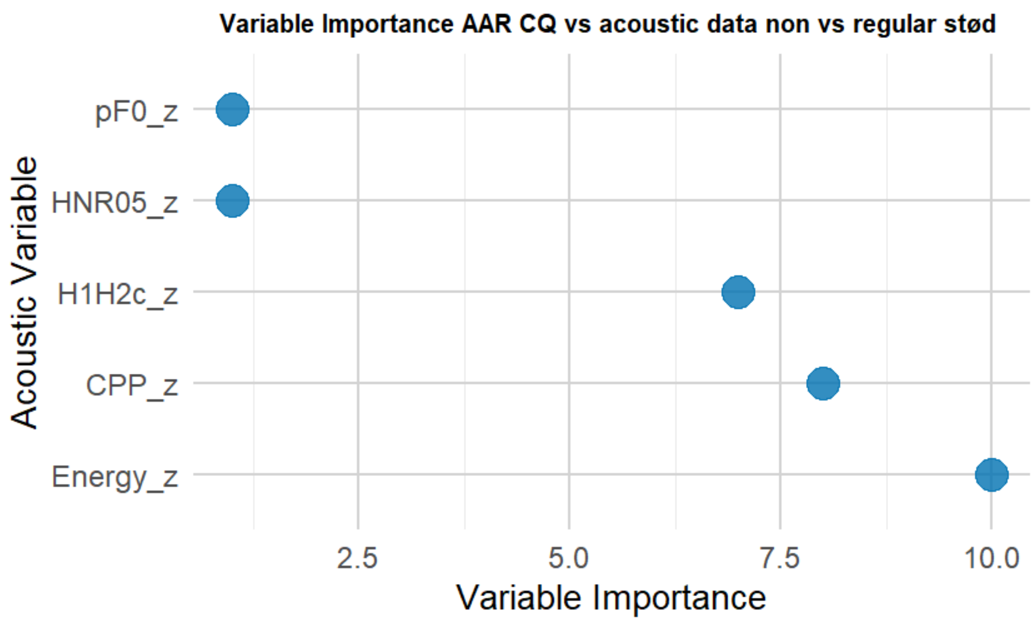

Figure 57. Acoustic variable importance for the Aarhus CQ values for the non vs regular stød contrast with scores expressed on a scale of 1-10 where 1 = least important and 10 = most important.

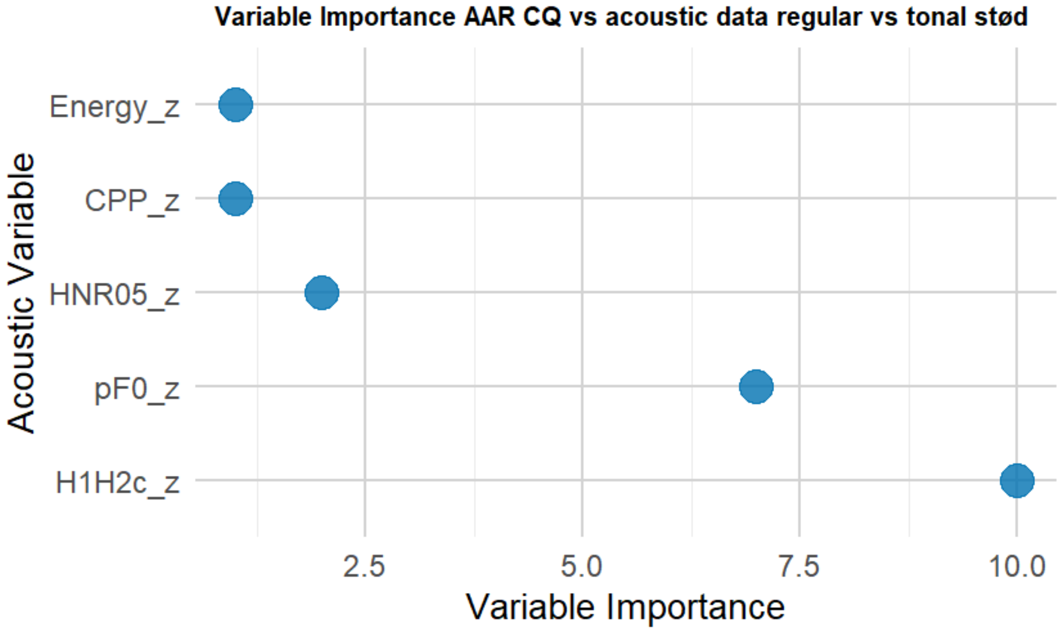

Figure 58. Acoustic variable importance for the Aarhus CQ values for the regular vs tonal stød contrast with scores expressed on a scale of 1-10 where 1 = least important and 10 = most important.

For the Aarhus non-stød vs regular stød contrast, the best predictor of difference in CQ values is both H1-H2 and intensity, equally ranked at 10 on the x-axis. This is followed relatively closely by f0 at over 7.5 and HNR over 5. CPP is the lowest ranked variable.

For the Aarhus regular vs tonal stød contrast, the highest ranked variable is H1-H2 at 10. This is followed by f0 just below 7.5 with quite a large gap down to the third highest ranked variable which was HNR at just below 2.5. The lowest ranked variables are both CPP and intensity.

These results are interesting because one would expect H1-H2 to be the highest ranked variable for all conditions given the proposed relationship between H1-H2 and glottal constriction, e.g. in the classification in Keating et al (2015). This is only found for three out of four minimal pair contrasts whereas the Aarhus non vs regular stød had intensity ranked highest with H1-H2 further down just below 7.5. Intensity was generally highly ranked for all the contrasts' differences in CQ values. This would be tempting to explain by the potential reduction in glottal airflow during longer vocal fold contact patterns leading to a decrease in intensity. This explanation is difficult to reconcile against the finding that the raw H1-H2 values were quite high for both dialects. What is also interesting is the quite similar results between the dialects in these random forest predictions for the variables despite the differences found in the EGG GAMM analysis for the previous research question which indicated more glottal constriction in the Copenhagen dialect than in the Aarhus dialect.

While the random forest models are useful for variable importance testing they do not indicate the exact nature of the relationship between the outcome variable and the predictors nor whether the relationship is statistically significant. It might be that even the highest ranked acoustic variable is a weak predictor. However, the GAMM model outcomes from earlier do give some indication in this regard, not on strength, but at least on statistical significance. Another common method for correlating two continuous variables statistically is Pearson's *r,* or Pearson's correlation coefficient, which gives either a positive correlation (as one variable increases, the other increases) or negative correlation (as one variable increases, the other decreases) and this is expressed in a range from -1 to +1. The closer Pearson's *r* value is to either extreme of this range, the stronger the correlation between variables (Winter 2019: 89). An issue with the method for this data is an assumption that the relationship between the variables is linear. As evident in the GAMM visualisations this is not the case throughout the entire syllable trajectory. Another potential issue is that of outliers. Pearson's correlation coefficient is usually not considered suitable for data with considerable outliers. Simple correlation was not pursued further for the current data as the need for more advanced statistical methods was indicated if one was to test for significance in correlation based on the random forest models alone.

To summarise the analysis in this chapter, the purpose was to answer research question 2, i.e. which measurements correlate with the stød. This was done to investigate intra-language variation on acoustic correlates of laryngealisation and more generally to explore how the correlation between laryngealisation and acoustic variables compare cross-linguistically. It was also an initial methodological step to explore what the relationship might be between laryngeal activity in different types of contrastive laryngealisation, EGG readings and acoustic correlates, both within and across languages. The acoustic variable importance showed relatively little intra-language/dialectal variation in that both dialects had f0 ranked highest for the non-stød vs regular stød contrast and intensity ranked highest for the regular vs tonal stød contrast. There was some dialectal variation in the second highest ranked predictors where H1-H2 and f0 were both ranked at a 5 out of 10 for the Copenhagen regular vs tonal stød and the same was found for f0 for the Aarhus regular vs tonal stød. The variables related to signal noise, HNR and CPP, were consistently ranked low in the models.

For the correlation between CQ values and acoustic variables there was also relatively little dialectal variation, particularly considering the differences in CQ values from the GAMMs that *did* suggest dialectal differences. Three contrasts had H1-H2 as the best predictor, perhaps unsurprising, but the Aarhus non vs regular stød contrast had intensity as the best predictor instead. This suggests some lenience in the correlation between glottal activity and acoustic outcomes for the stød, as otherwise all contrasts should have had H1-H2 ranked highest. When characterising phonemic laryngealisation in other languages intensity is not a parameter that is looked at very much, making the results for Danish more difficult to compare cross-linguistically. The results do suggest, however, that intensity is potentially a helpful acoustic measure, at least for laryngealisation types similar to the stød. The consistently low ranking of CPP and HNR in the EGG variable importance could indicate that during the stød, the glottis itself is not the focal source of signal noise but rather, other structures could contribute to the decline observed in HNR and CPP during some of the GAMM smooths. This finding is entirely compatible, and even predictable, from the LAM framework that puts the aryepiglottic constrictor mechanism as the source of much of what is labelled on the hyper-compressed phonation continuum.

6. How are the voice quality changes timed?

The previous sections of analysis focused on answering research question 1 and 2. This leaves research question 3 unexplored which is concerned with when in the stød syllable the onset of laryngealisation occurs, as measured by a change in acoustic measures and a change in vocal fold contact patterns derived from the EGG signal. This question was motivated by the large amount of cross-linguistic variation in contrastive laryngealisation previously outlined which found that laryngealisation rarely occurs throughout the entire syllable or segment in phonological voice quality contrasts, but instead spans various proportions of the segment. Timing is thus integral to characterising the phonetic variation in phonological voice quality contrasts because it co-occurs with the phonological domain or scope of the contrast. The scope of contrastive voice quality can reasonably be assumed to be more restricted in its scope compared to sociolinguistic uses of voice quality due to the former belonging to a phonological domain in a language and the latter being free to span entire utterances, unrestricted by phonological structure.

6.1. Timing

Determining the timing of the occurrence of stød can be a rather intricate matter because it necessitates first having identified the relevant acoustic and articulatory correlates of the stød which, as the analyses here as well as previous research (e.g. Riber Petersen 1973, Fischer-Jørgensen 1989) has attested to be somewhat variable and sometimes even undetectable despite auditory salience. A common acoustic approach has been to identify the onset of the stød 'proper' by declines in f0 and intensity. In the current study the acoustic analysis was primarily conducted via dynamically fitting GAMMs to the stød trajectories and comparing these between minimal pairs. To visualise the difference, each GAMM was plotted alongside a difference smooth from the model that indicated where along the trajectories the correlates diverged the most. This divergence might not necessarily correlate with the onset of the stød exactly as the effects would presumably need a at least few milliseconds to develop over time to reach significant divergence and are only shown in the difference smooth in cases where the trajectories do diverge, meaning there might be effects of the stød in the trajectories that are visible but are not shown in the difference smooth because the effects in the minimal pairs are too similar. Further, it does not measure non-modal voice quality onset in the individual stød trajectory per sé but rather measures the onset of the differences in the minimal pair contrasts on average, smoothing out much variation between tokens and speakers. With these limitations in mind the difference smooths were used as a reference point for the timing differences between stød trajectories, summarised in timepoint onsets and offsets in a table below for each dialect and each stød type contrast to give an overview of the timing of difference intervals:

Contrast	**Measurement**	**Timepoint interval of most significance (determined by difference smooth)**
CPH non vs reg	f0	0-100%
CPH reg vs tonal	f0	No sig. difference
AAR non vs reg	f0	0-90%
AAR reg vs tonal	f0	No sig. difference
CPH non vs reg	Intensity	0-40%, 50-100%
CPH reg vs tonal	Intensity	55-90%
AAR non vs reg	Intensity	0-100%
AAR reg vs tonal	Intensity	50-100%
CPH non vs reg	CPP	30-100%
CPH reg vs tonal	CPP	43-100%
AAR non vs reg	CPP	0-37%, 70-100%
AAR reg vs tonal	CPP	42-100%
CPH non vs reg	H1-H2	0-100%
CPH reg vs tonal	H1-H2	No sig. difference
AAR non vs reg	H1-H2	0-40%, 65-100%
AAR reg vs tonal	H1-H2	No sig. difference
CPH non vs reg	HNR	5-100%
CPH reg vs tonal	HNR	50-80%
AAR non vs reg	HNR	30-72%, 78-100%
AAR reg vs tonal	HNR	28-100%

Table 18. Timepoint intervals with most significance between acoustic correlates of the levels of each stød minimal pair contrast as determined by the difference smooth from the fitted GAMMs. Black indicates measures for the Copenhagen dialect, red indicates measures from the Aarhus dialect. SHR was excluded due to the poor tracking of the data.

Table 18 shows large variability in the onset of differences, both according to dialect, stød type and acoustic measure. This is surprising given that the measures are supposed to capture the same phenomenon, laryngealisation. It is perhaps not surprising, however, that the stød is highly variable, and the aspect of timing appears to be no exception. Based on previous research, the 'stød proper' occurs in the latter 50% of the syllable (Fischer-Jørgensen). The timing in this table, however, suggests a more complex relationship than that. Some measures also have more than one timepoint interval of significance within the same trajectory.

Looking at the numbers, however, it seems that timings can roughly be divided into two large groups:

1) Measures that are significant throughout (almost) the entire syllable, e.g. f0 and HNR for the non vs regular contrast in both dialects, and
2) Measures that are overwhelmingly significant in the latter part of the syllable, e.g. intensity for the regular vs tonal contrast for both dialects.

There is a tendency in both dialects for measures in the non vs regular stød contrast to belong to group 1 (f0, intensity, H1-H2, HNR) and a tendency for measures in the regular vs tonal stød contrast to belonging to group 2 (intensity, CPP, HNR). This suggests timing differences not so much between dialects but between stød types, meaning phonetic variation depending on the stød basis. To investigate this further, the timing of the CQ trajectory differences are also part of the picture. These are summarised in the table below:

Contrast	**Measure**	**Timepoint interval of most significance (determined by difference smooth)**
CPH non vs reg	CQ	18-100%
CPH reg vs tonal	CQ	No sig. difference
AAR non vs reg	CQ	No sig. difference
AAR reg vs tonal	CQ	No sig. difference

Table 19. Timepoint intervals with most significance between the CQ levels of each stød minimal pair contrast as determined by the difference smooth from the fitted GAMMs. Black indicates measures from Copenhagen, red indicates measures from Aarhus.

Due to the lack of significance for three of the four stød pairs, the CQ values are difficult to conclude anything about the timing from. The one contrast that was significant suggests quite an early onset timing for the increase in vocal fold contact in each glottal cycle starting as early as the 18% timepoint. This correlates nicely with the H1-H2 timing for the same contrast which spanned the full trajectory but does not explain the discrepancy between H1-H2 indicating no glottal constriction and the CQ values indicating increased vocal fold contact throughout the regular stød.

Some predictions for the timing patterns of contrastive phonation were outlined in section 2.4. where Silverman (199571997) proposed that timing should ensure optimal perception and that this could manifest itself in a few ways: (i) placing the phonation contrast on the first portion of the vowel for early auditory detection for the listener, (ii) having a transient portion of non-model phonation in between two stretches of modal phonation for maximum contrast, or (iii) non-modal phonation types are in themselves salient enough that they do not require perceptual enhancement and therefore may last the entire duration of a vowel. Looking at the trends in the current data, acoustic correlates in group 1 were significant throughout (almost) the entire syllable, suggesting Silverman's proposal number (iii). However, measures in group 2 that are overwhelmingly significant in the latter part of the syllable could suggest a modified Silverman's proposal (i) where the salient portion is on the latter part of the vowel, or in this case, stød basis.

Some tokens, however, had laryngealisation that expanded beyond the phonological scope of the stød basis which was a surprising finding described in the following section.

6.2. Post-stød creak

This section covers a different aspect of timing that became apparent during the preparation of the data and seemed prevalent enough to include, despite not being part of the initial research questions. When annotating the data for the acoustic analysis, an unexpected pattern emerged where a remarkably high number of tokens had what could be called 'post-stød creak', a visually and auditorily distinct presence of laryngealisation *after* the segment or syllable where the stød was expected to occur based on the phonological stød basis. This laryngealisation is labelled as post-stød creak because its manifestation looks overwhelmingly like prototypical creaky voice as described below. As mentioned in the previous section, it was generally assumed that the phonetic realisation of a phonological voice quality contrast is restricted to the phonological unit it is contrastive on due to its core semantic function in the phonology of the language. Therefore it seemed relevant to provide at least some account of the prevalent observation in the data that the stød is not restricted to the stød basis, i.e. its phonological unit, in all cases. This phenomenon has not previously received much attention in Danish, although it has been noted that '[..] stød in a long vowel may extend into a succeeding sonorant consonant. Stød in a sonorant consonant may already begin during the last part of the preceding short vowel and may continue into a succeeding voiced sound [..]' (Grønnum & Basbøll 2001: 234) and that the stød has a very variable duration which may continue into the following syllable (Grønnum & Basbøll 2007: 199). The present study is concerned with dialectal differences in phonetic variation and it became obvious going through the data for annotation that the dialects differed noticeably in the number of tokens with post-stød creak. Thus it seemed important to outline these differences in more systematic detail. As the presence of post-stød creak was not part of the original hypotheses about the phonetic variation of the stød, this analysis is not intended as an in-depth and exhaustive description of the phenomenon. The materials and experiment were not designed with this analytic focus in mind, so rather, this is intended as an initial broad account of the trends discovered in the data, but has to be interpreted with this *aber dabei*. It is, however, intended to encourage more deliberate research into the phenomenon, particularly from a cross-linguistic perspective, but also from a Danish dialectal perspective.

The procedure for this account was as follows. Post-stød creak, henceforth shortened PSC, was marked manually at the initial stage of annotation where the stød basis was also annotated for the main analysis in the thesis in Praat (Boersma & Weenink 2021). As it took going through a bit of data to discover the PSC trend, the first 2-3 annotated speakers were revisited for PSC marking post-annotation of the stød basis. PSC was marked on a separate segment tier on all tokens where the following criteria were fulfilled immediately following the stretch of phonological stød basis: 1) clearly audible prototypical creaky voice, 2) visually detectable irregular vibrations in the waveform and 3) vertical, spaced striations in the spectrogram. An example of a token that fulfilled these criteria is given below:

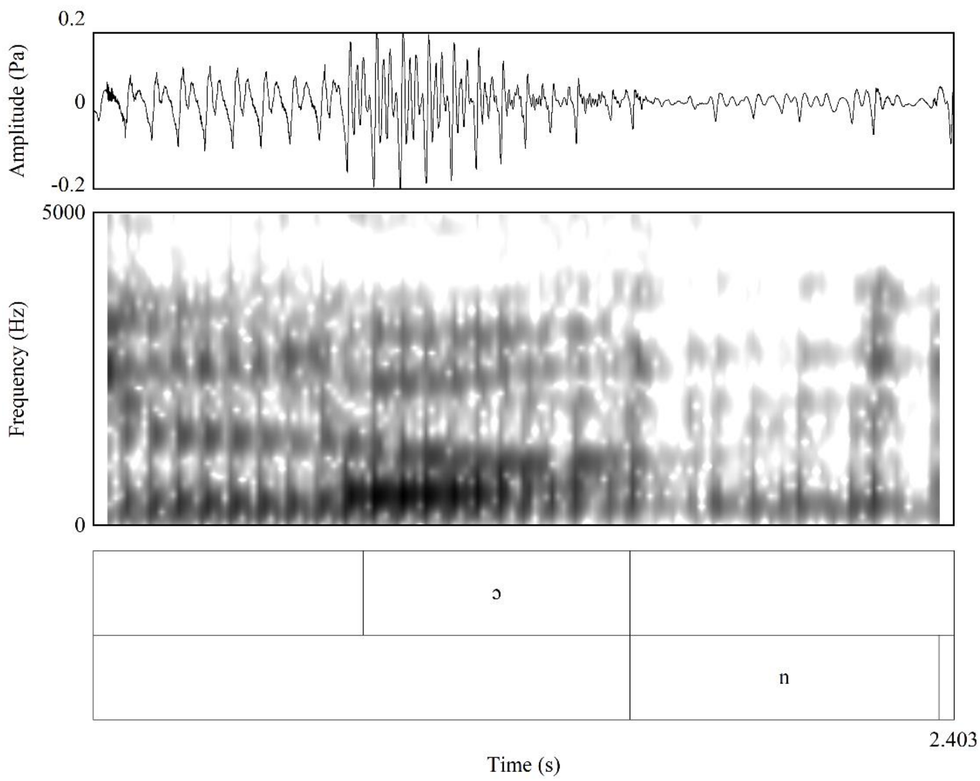

Figure 59. Praat picture showing an example of the monosyllabic word lån 'loan' uttered by a Copenhagen speaker with PSC on the nasal segment (Tier 2) adjacent to the long vowel that is the stød basis (Tier 1). The waveform is shown on top, the spectrogram in the middle and the annotation tiers on the bottom.

For an initial account of PSC only tokens with the clear signs of creaky voice described above were included as this was easy and unambiguous to mark in the annotation process without conducting any acoustic analysis as PSC was not part of the research questions but rather an unexpected bi-discovery. Going on the visual and auditory cue of unambiguous creaky voice is also why the phenomenon is labelled post-stød *creak* rather than laryngealisation in this section, although the stød is, as demonstrated in the sections above, not prototypical creaky voice. Applying this criterion means that it is entirely possible the extent of PSC is much larger than accounted for here if tokens with all types of laryngealisation were to be included.

It should be noted that the audio-visual procedure of identifying creak in this section is different to the procedure used in the main analysis of phonetic variation during the stød. There are a couple of reasons for this. Firstly, unlike the stød basis, PSC does not have a phonological basis to dynamically analyse its occurrence on, meaning fitting GAMMs as was done in the main analysis would be less useful because it would be impossible to establish minimal pairs and thus impossible to isolate the voice quality effects of PSC from segmental effects. The main analysis in the previous sections was conducted on annotations of the phonological stød unit, i.e. the stød basis, whereas for this smaller analysis, PSC does not have a defined phonological basis but, rather, is defined as creaky voice occurring *immediately after* the phonological unit consisting of the stød basis. Secondly, as PSC only occurred on a handful of tokens for some speakers, it was less time-consuming to manually inspect using audio-visual identification. For this reason it was only done for the analysis in this section and not in previous chapters, as well as needing previous chapters to potentially identify many types of stød beyond this more prototypical kind. Thirdly, as PSC was an unexpected discovery related to timing, the purpose of the analysis was to account quantitively, rather than qualitatively, for its presence, meaning a simple identification of whether it was present on a token or not was sufficient for this account which does not claim anything about the phonetic details of the phenomenon beyond its timing. In the methods section for the main study it was described how in two cases there was non-stød creaky voice on the entire elicited phrase and the tokens was therefore not included in the analysis. This was justified on the grounds that the focus of the study was contrastive use of non-modal voice quality and sociophonetic phrase-spanding creaky voice cannot be considered phonologically contrastive. In a similar vein, the effects labelled PSC here differ from random instances of creaky voice in that they are defined as creaky voice occurring *immediately after* the phonological unit stød. As such, the difference between PSC and random occurrence of creaky voice on e.g. non-stød tokens is not (necessarily) phonetic, but rather, the difference is phonological because the former is defined by adjacency to a phonological voice quality contrast lexically relevant to the speakers of the language and the latter is not. As such, it would not be appropriate to analyse instances of creak not adjacent to the phonological stød basis in an analysis that has the sole purpose of accounting for instances of post-*stød* creak, not general occurrence of creaky voice. That said, it is certainly an interesting question whether there are any consistent phonetic differences between what is defined here as PSC and occurrences of non-phonological laryngealisations. In the data, some non-phonological laryngealisations do occur, mostly as pre-vocalic irregularities at the onset of a word-initial vowel. An example of a token containing both PSC and pre-vocalic non-phonological laryngealisation is given below:

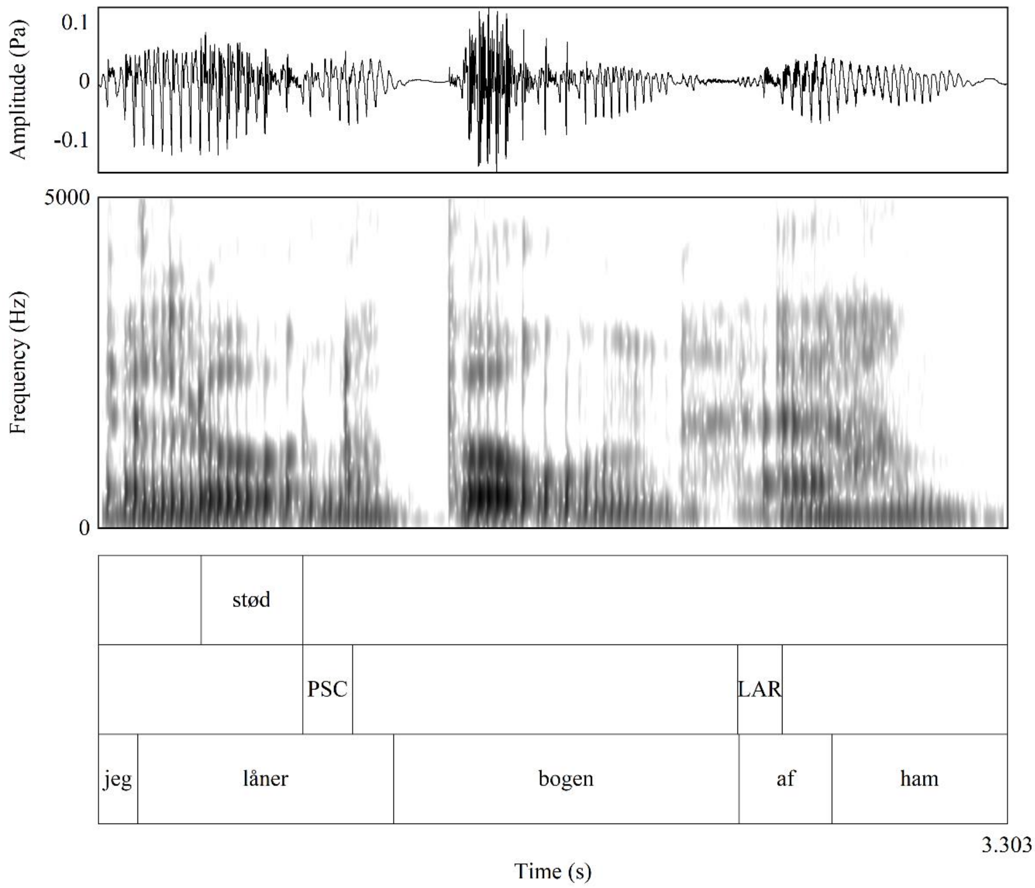

Figure 60. Waveform (top) and spectrogram (middle) of a male Copenhagen speaker saying the prompt jeg låner bogen af ham 'I am borrowing the book from him' with the stød basis marked on Tier 1, PSC and non-phonological laryngealisation ('LAR') marked on Tier 2 and the words marked on Tier 3 in standard orthography. The prompt has regular stød.

Note that in the example there is also stød production in the first syllable of the word *bogen* ‘the book’ but this was not annotated as this was not a stød token constituting a minimal pair from the word list. Whether the two laryngealisations marked on Tier 2 are phonetically different is outside the scope of an analysis of phonological voice contrast, particularly as the study was not designed for this analytic purpose and as a consequence, many factors that would effect the occurrence of these laryngealisations are not controlled for. This is an interesting topic for future research, however, and therefore it is worth posting the example here. For the purposes of this section, the important difference in the two laryngealisations on Tier 2 is position. PSC occurs immediately after the stød basis, LAR does not.

Having made the scope of this section clear, the overall prevalence of PSC is shown in Table 20 below, calculated for each dialect:

Dialect	**PSC, regular stød**	**PSC, tonal stød**
Copenhagen	194/893 = **21.7%**	132/225 = **58.7%**
Aarhus	41/1003 = **4.1%**	13/249 = **5.2%**

Table 20. Percentage of tokens with PSC out of the total number of tokens with stød, stratified by dialect and stød type.

As is evident from the table, PSC is much more common in the Copenhagen dialect compared to the Aarhus dialect. This makes it interesting from a comparative dialectal perspective as taken on in this thesis. Further, PSC occurs more frequently on tonal stød tokens compared to regular stød tokens, meaning an interesting correlation between PSC and both dialect and stød type. To test whether the correlation is statistically significant a binomial logistic regression model (Generalised Linear Mixed Model, GLMM) was fitted in R (R Core Team 2022). PSC occurrence was coded as a binary outcome, yes or no, on each token and the model was fitted for two different conditions: (i) the effect of dialect on PSC occurrence and (ii) the effect of stød type on PSC occurrence. For each condition a full model was fitted with PSC as the outcome variable and stød type, dialect and speaker sex as respective predictor variables, with speaker and word included as random effects and an interaction term between dialect and speaker sex. The full model was compared to a reduced model excluding the fixed effect of interest, i.e. dialect for the first condition and stød type for the second condition. The model summaries showed that the effect of dialect on PSC occurrence is significant ($\chi 2[2]=31.047$, $p=<0.001$) which was expected based on the relatively large differences in prevalence observed in Table 8.2.1. The effect of stød type on PSC occurrence was also found to be statistically significant ($\chi 2$[9]=27.717, p=0.001). This effect appears to be driven by the relatively large differences in PSC prevalence between the two stød types in the Copenhagen dialect as differences between regular and tonal stød PSC are much smaller in Aarhus.

The next interesting finding is the segmental span of PSC. As mentioned above, Grønnum & Basbøll (2007: 199) noted that the acoustic correlates of the stød may continue into the following syllable. They did not, however, quantify exactly how long into the proceeding syllable the stød did, or indeed *could*, occur. Whether the stød continues into the following syllable, or just expands to the adjacent segment within the syllable depends on which segment 'receives' the stød – if the stød basis does not constitute the entire syllable (i.e. is a tonal stød type), the creakiness may be evident on the proceeding segment rather than the following syllable. Indeed, inspecting the PSC tokens here, the overwhelming trend for both dialects is that PSC occurs on the segment immediately adjacent to the one receiving the stød. Interestingly, however, it may span two segments, three segments, and in one instance, even four segments after the stød, a fairly large segmental span compared to the cross-linguistic trends in contrastive phonation outlined in section 2.4. The percentage of PSC on the respective post-stød segments are given in the table below:

Dialect	**PSC1**	**PSC2**	**PSC3**	**PSC4**
Copenhagen	331/1118 = **29.6%**	33/1118 = **3.0%**	3/1118 = **0.3%**	1/1118 = **0.09%**
Aarhus	54//1252 = **4.3%**	22/1252 = **1.8%**	2/1252 = **0.2%**	0/1252 = **0.0%**

Table 21. Overview of tokens with PSC on segments 1-4 after the stød calculated as a percentage of the total number of tokens with stød stratified by dialect.

As the table shows, having PSC on the fourth segment is a rare single case occurrence in this data but it is worth noting that this is possible. The dialectal differences in PSC prevalence decrease the further away from the stød segment PSC occurs, meaning the dialectal differences pertain most prominently to patterns in PSC1 occurrence.

The third interesting finding when analysing PSC was a very large degree of inter-speaker variation in its prevalence. Figure 61 and 62 show the PSC occurrence for each individual speaker, aggregated by stød type, regular vs tonal:

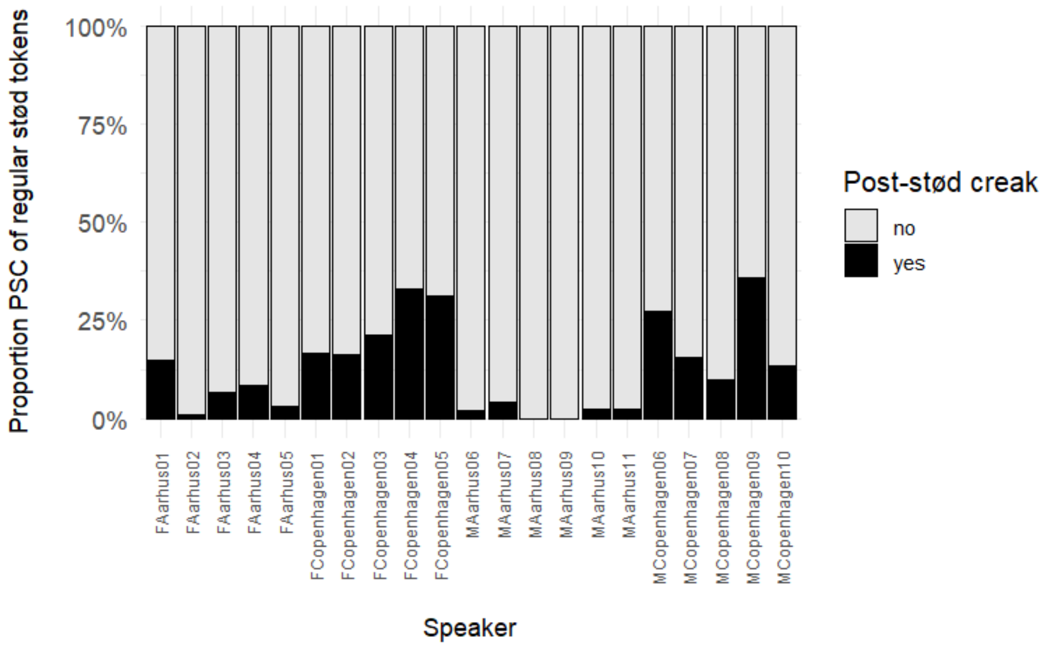

Figure 61. Bar graph showing the proportion of PSC tokens for each individual dialect speaker as a percentage of the total number of regular stød tokens. Speaker code starts with 'F' for female speakers and 'M' for male speakers.

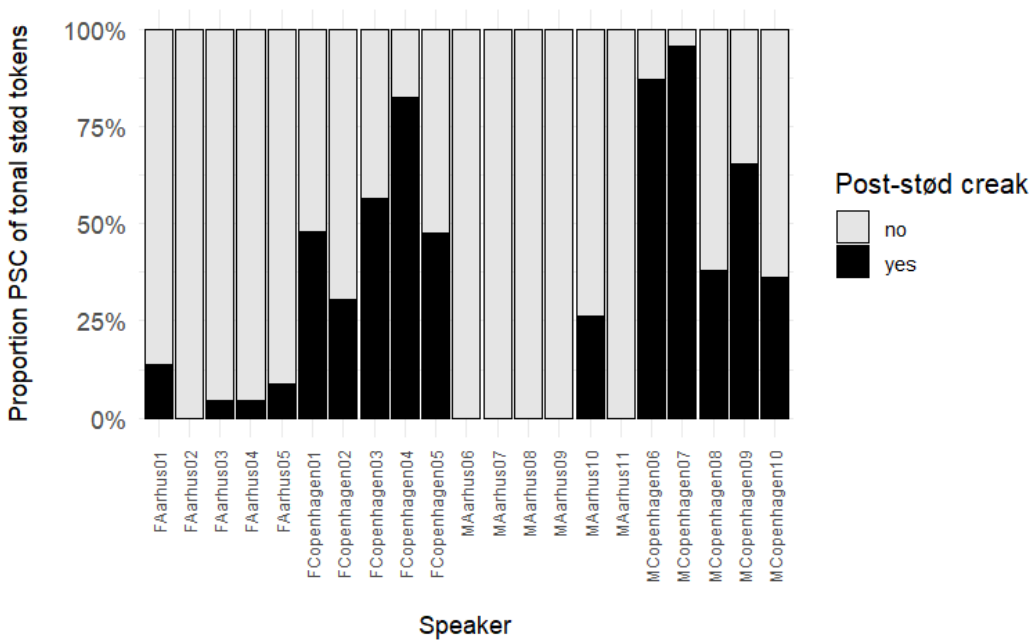

Figure 62. Bar graph showing the proportion of PSC tokens for each individual dialect speaker as a percentage of the total number of tonal stød tokens. Speaker code starts with 'F' for female speakers and 'M' for male speakers.

The graphs show that for some speakers PSC occurs on almost all tokens of a stød type, e.g. speaker MCopenhagen07 in Figure 62 showing tonal stød PSC percentages. In contrast, a handful of speakers from Aarhus have no PSC occurrence on any tokens of any type. This suggests that PSC prevalence has a fairly large degree of idiosyncrasy and is not merely the result of e.g. coarticulatory factors of the stød. It is somewhat surprising to find PSC on the tonal stød tokens in Aarhus, and perhaps also surprising on the regular stød tokens, as these were shown to lack laryngealisation in the EGG analysis for this dialect. When no laryngealisation is present in the manifestation of the voice quality contrast it is not clear why creaky voice would be found on the adjacent segment(s). It could be a consequence of a falling tone where the vocal fold vibrations will become irregular if the f0 falls below a certain threshold where modal phonation can no longer be sustained. However, the GAMMs for f0 did not suggest that this was the case. To explore this further, one would need to conduct a more qualitative tone analysis of the tonal stød tokens to confirm what a low tonal threshold is and also confirm that the tokens which have PSC are the same tokens as the ones reaching below this threshold as PSC clearly does not occur on all tonal tokens. There might also be sociophonetic factors determining the inter-speaker variation. Sociophonetic information was not collected on the participants as this was not the focus of the study but this would be an interesting subject for future research, as would uncovering the extent of the overall idiosyncrasy of PSC as this is highly relevant to forensic phonetic applications of PSC as a potential identification contributor to the voice profile of a Danish individual.

Finally, what conditions PSC is difficult to determine as the study was not designed to elicit this phenomenon specifically and thus was not controlling for factors that may influence its presence such as intonation and phonetic environment. Further, when pooling all speakers, no word in the data has PSC consistently which suggests influences beyond phonetic environment or word-internal factors, so even if these were controlled for, this might not result in any consistency in occurrence.

As a phonetic phenomenon it is perhaps unremarkable that the adjacent stød segment has some effects of laryngealisation which has also been noted in previous work (e.g. Grønnum & Basbøll 2007: 199). It is, however, unusual from a cross-linguistic perspective for the phonological voice contrast to span even the full segment, nevermind the adjacent segments to the contrastive segmental unit. In cases where the vowel receives the stød, e.g. in the word *lån* 'loan' shown in Figure 8.2.1. above, and the adjacent nasal is creaky, this could partly be explained by the domain of the stød, given that it is a syllabic rather than segmental property and the nasal belongs to the monosyllable despite not constituting the stød basis because the vowel is long. However, there are also cases where contrastive pairs differ, at least partly, in segmental stød occurrence rather than syllabic stød occurrence, e.g. *lån* [ˈlɔːˀn] ‘loan’ vs *lund* [ˈlɔnˀ] ‘grove’, conditioned by the length of the vowel. These monosyllabic long vowel+sonorant consonant combinations seem to be what triggers the tonal stød condition in Aarhus. Given that the prevalence of PSC was statistically significant as a function of stød type, it could be that the usual cross-linguistically restrictive segmental scope of non-modal phonation observed in other languages is less restricted on these tokens because the vowel length itself serves as a salient cue to the contrast. This in turn lessens the perceptual load from the laryngealisation to cue the stød. This explanation does not, however, account for why PSC also occurs on regular stød tokens where the minimal pairs differ only in stød with no vowel length difference, e.g. *låner* [ˈlɔːˌnɐ] ‘(a) borrower’ vs *låner* [ˈlɔːˀˌnɐ] ‘borrows’.

In summary, more work needs to be conducted to give the phenomenon of PSC more attention from a sociophonetic, forensic and typological perspective. For now, the analysis of the timing of the stød has revealed systematic differences in the post-segmental span between dialects and between speakers that defy previously documented trends of segmental scope of contrastive voice quality in other languages, assuming PSC is part of the phonological contrast due to its adjacency to the stød basis. This was an unexpected finding when segmenting the data but worth documenting, particularly in a comparative dialectal study as it seems to differentiate dialects, supported by fitting GLMMs showing that it did so to a statistically significant degree.

7. Discussion

The discussion starts with a brief overview of the results presented in the previous sections as a reminder and summary. It is then divided into sections discussing different parts of the methodology used to obtain the results and how the findings relate to existing the literature on both the phonetics of the Danish stød specifically and the phonetic variation in phonological contrasts more generally. As with any piece of research, every attempt at an answer to a research question has spawned more new questions and these are summarised as well in a section about future perspectives on the current research findings.

The aim of this thesis was to explore phonetic variation in phonologically contrastive laryngealisation through the comparative lens of dialectal differences within a language, an aspect of phonological voice quality variation not previously well-studied or documented. This was done via researching the Danish stød as a case study. Including a dialect comparison between the Copenhagen and Aarhus dialects to study phonetic variation in laryngealisation had the methodological consequence of dividing the stød into two types, regular and tonal, elicited as minimal pairs contrasting with no stød and regular stød, respectively. This division was done to accommodate previous findings by Bodil Kyst (2004, 2007) who suggested the presence of a purely tonal stød realisation under certain conditions in Aarhus speakers. This in turn had motivated the stød as a case study for phonetic variation in phonological contrasts because such variation had already been tentatively attested and thus could be expected to constitute a source of interesting phonetic variation. The division of the stød yielded the finding that the regular vs tonal stød type was statistically significantly distinguished in some of the acoustic measures in not only the Aarhus dialect as hypothesised but also in the Copenhagen dialect, something previously unattested in the literature for Modern Standard Copenhagen which is not traditionally divided into stød types. It has been noted that there is a difference in which segment 'receives' the stød (e.g. Fischer-Jørgensen 1989) but the phonetic consequences of this difference have not been systematically studied. The acoustic findings here suggests that it is useful to consider what type of stød basis the stød occurs on when measuring its articulatory and acoustic correlates. This could also be relevant to other languages with phonemic laryngealisation if they have varying segment or syllable types that this can be contrastive on.

Fitting GAMMs to the stød trajectories to compare their mean smooths dynamically showed that the dialects were similar in the acoustic parameters that were important for distinguishing the stød types, meaning less phonetic variation than expected. These acoustic parameters were compared to the subtypes of laryngealisation in Keating et al (2015) but none of the descriptions fitted precisely to the data, suggesting that there is potentially more variation than the five subtypes included in the paper. The closest subtype for both dialects was non-constricted creak. The dialects differed more in the EGG data where Copenhagen speakers had higher mean CQ values than the Aarhus speakers during the production of both stød types. This suggests some phonetic variation in the vocal fold contact patterns between dialects. This also supported the hypothesis by Kyst (2004) that the tonal stød in Aarhus not laryngealised, at least not via glottalisation traceable via EGG, but it also suggested that the regular stød in Aarhus is perhaps not very laryngealised either. Fitting random forest models to the acoustic data proposed that the dialects were very similar in the acoustic variables that best predicted the difference between stød types, namely intensity and H1-H2. Modelling the acoustic predictors with the CQ values as the outcome variable found that for three out of four contrasts H1-H2 was the best predictor whereas for one, intensity was the best predictor variable.

The analysis of stød timing showed that two larger trends described the data well based on the significance suggested in the difference smooths. Group 1 timing included measures that were significant throughout (almost) the entire syllable whereas group 2 timing included measures that were overwhelmingly significant in the latter part of the syllable. An interesting finding in the timing analysis was that of post-stød creak on a relatively large number of tokens and that its prevalence was significantly correlated with both dialect and stød type, being most common in the Copenhagen tonal stød and exhibiting large inter-speaker variation with some speakers having no post-stød creak on any tokens. This finding could be of particular interest to Danish forensic phonetics and sociolinguistics given how relatively idiosyncratic it appears to be. This is in fact where most phonetic variation was found between dialects outside of differences in CQ values.

Having summarised the major results, the methods, results and perspectives on previous research are now discussed in the proceeding sections.

7.1. Materials and obtaining acoustic measurements

The stød was elicited using a word list with minimal pairs contrasting in two ways, either (i) no stød vs a regular stød or (ii) a regular stød vs a tonal stød. The stød type was defined based on the stød basis. The resulting data is biased by the fact that it was not possible to use a carrier phrase to embed the targets words in due to the homographic nature of some stød pairs. This results in minimal pairs where the stød basis is well controlled but the surrounding environment is not, meaning the stød realisation differences could be partly influenced by differences in overall sentence intonation and stress placement. The material was, however, quite naturalistic compared to other studies on the stød which is a strength for exploring a realistic range of phonetic variability. As such, it is worth noting here that part of the variability observed could be due to sentence structure differences rather than inherent phonetic differences in laryngealisation as well as due to differences in stød basis.

The acoustic measurements used as correlates for voice quality changes in laryngealisation in this study were f0, intensity, H1-H2, HNR, CPP and SHR. These were motivated partly by being used previously to classify subtypes of creaky voice by Keating et al (2015), providing a framework within which to analyse the dialectal variation of the stød, and partly motivated by previous research on the stød suggesting that intensity is a strong correlate in Modern Standard Copenhagen (e.g. Fischer-Jørgensen 1989). CPP was also included as a measure less sensitive to very irregular vocal fold vibrations (Heman-Ackah et al 2003) that would possibly occur during the stød. Further, these measures could all easily be obtained via the software VoiceSauce (Shue 2010; Shue et al 2011). With automatic tracking software the researcher is able to analyse a greater number of tokens than with manual tracking which is favourable for an analysis such as the one conducted here. However, with automatic tracking there is an inherent risk of erroneous measurements. Given the large number of tokens it was not possible to check the tracking for each utterance individually and manually correct them on each measure and this might have lead to some erroneous measurements being included in the final analysis. The tracking performance was plotted to account for any missing acoustic data to make the analysis and results transparent and interpretable in the context of the tracking quality. A strength of having lots of tokens is that minor outliers are unlikely to affect the average result enough to make it completely invalid or unreliable so despite the lack of manual individual checking, the results are not considered invalid. Another issue with tracking acoustic measurements during the stød specifically is that periodicity may become so irregular that voicing is no longer detected (e.g. Hansen 2015), and thus portions of the signal might not be tracked. This is an interesting kind of methodological paradox almost, given that voice quality changes generally require voiced speech to be articulated but that laryngealisation might obscure the vocal fold vibrations so much that voicing ceases and that disturbance in itself constitutes the voice quality change but cannot be tracked by measures presupposing that voicing is present. The paradox is partly solved by acknowledging that despite the loss of (trackable) voiced phonation it cannot be said, however, that such a change is equal to the absence of voicing, as in a whisper, but should merely be characterised as an irregularity in voicing which may lead to poor tracking, an unfortunate methodological consequence. The f0 values did have some missing data points, particularly towards syllable termination, indicating that this effect might have happened. However, the absolute f0 values were not particularly low which could indicate two things in terms of tracking: (i) irregular vibrations were not present to a degree that significantly lowered the f0, or (ii) proportions of a number of tokens might have syllable-final missing intervals due to vocal fold vibration irregularities leading to untracked data and thus artificially inflating the f0 average to appear higher. Based on the findings of PSC on a large handful of tokens particularly in Copenhagen, both explanations seem relevant and the most likely scenario is a combination of the two. In future work (the lack of) tracking could be quantified more accurately with e.g. a Praat script extracting timings for the entire annotated intervals and the detectable voiced portion of the annotations via f0, subtracting the latter form the former to get an idea about any residual portion of the annotation interval not tracked by VoiceSauce. In the current study, this was done more roughly by converting 0s to NAs, revealing missing data.

The study design also suffered somewhat from the fact that covid-19 protocols were to be followed, not allowing the researcher to get too close to the participants. This meant they fitted the lapel microphone and EGG electrodes to their own body. The clip mic is unlikely to have moved too much during the experiment as it was fixed to the participant but it is possible that smaller movements have affected a more distance-sensitive measure like intensity due to this setup. This is more likely to have an effect between speakers rather than within speaker, however, due to limited movement in one session. The GAMM models were fitted with random smooths to inform the models that the data was coming from different speakers and different words, meaning this lack of intensity calibration is less likely to be a problem when interpreting the results and they can be considered reliable, particularly as intensity was a measure that was very well tracked in VoiceSauce. The setup is more likely to have affected the EGG data as the electrodes might have been poorly placed initially or may have moved during the recording session due to the larynx moving with articulation. The electrodes are fixed in place with a Velcro strap but this is not always enough to stop undesired movement during talking. Indeed, some speakers had quite a low amount of glottal cycles tracked compared to others, possibly a consequence of this. This is discussed further in a separate section on obtaining the articulatory measurements.

7.2. Acoustic correlates of laryngealisation

The premise of the studying the stød as a case study for phonetic variation in hyper-compressed voice quality contrasts was that the stød is a laryngealisation akin to a type of creaky voice based on previous descriptions (Fischer-Jørgensen 1989, Grønnum et al 2013). Given that there is more than one type of creaky voice (Keating et al 2015) it was considered relevant to try to characterise which type the stød most closely resembles and in particular if there was more than one type present between stød types and dialects as a means to uncover and describe phonetic variation in phonological contrasts in voice quality. As articulatory measures are indirect correlates of articulatory gestures, there is always the possibility that finer detail is not captured in the acoustic signal with the methods used and this is also the case for the current study. As discussed in the previous section f0 measures might suffer somewhat from a lack of periodicity but suggestions for how to account for this were given as well as for how to deal with the lack of control in intensity. For f0, it might also be that slightly more tokens would have been captured with even lower minimum settings in the algorithm. It has been reported how intentionally sustained creakiness in a lab can be as low as 40 Hz (Catford 1964: 32), and elsewhere reported between 24 Hz and 56 Hz, averaging 35 Hz (Michel and Hollien 1968). The standard VoiceSauce settings of 40 Hz were deemed sufficiently low based on other studies reporting creaky values much higher than these, closer to the boundary for creaky voice at below 100 Hz below 100 Hz (Laver (1980: 122, Davidson 2019: 238-239). It is possible that lowering these settings might have captured more data but overall f0 was tracked well and the raw measures did not indicate that f0 values were anywhere near the lower boundary, suggesting lowering the setting 10 Hz would not have made a large difference. The attention will now be turned to the remaining measures used.

In the acoustic analysis H1-H2 was found to be useful in distinguishing between no stød vs regular stød contrasts for both dialects but did not significantly distinguish regular vs tonal stød trajectories in neither Copenhagen nor Aarhus. As H1-H2 has been correlated more with glottal constriction rather than e.g. noise in the signal more broadly, this confirms that noise during laryngealised phonation is not necessarily caused by glottal constriction directly but can be achieved through other articulatory strategies simultaneously with the presence or absence of glottal constriction. This observation is compliant with the LAM framework that was used here a guideline for how to conceptualise laryngealisation.

Garellek & Keating (2011) cite H1-H2 as the most widely used measure of phonation and as a measure that has been found to distinguish contrastive phonation in a number of languages (Esposito 2006; 2010a). Much of the research on voice quality and H1-H2 has been conducted with the purpose of distinguishing modal from breathy phonation, e.g. Hanson et al (2001) finding that H1-H2, or specifically H1-A1 with A1 being the first formant, seems to reflect the degree of posterior glottal opening. Given that creaky and breathy phonation are two extremes on the same spectrum, it is reasonable to assume that H1-H2 will also correlate with changes in laryngealisation, and Hansen (2015) did find that H1-H2 captured laryngealisation during the stød to a degree. However, since it should be possible to produce some types of laryngealisation whilst also maintaining a posterior glottal gap, utilising the aryepiglottic constrictor mechanism, patterns in H1-H2 may not be as sensitive to laryngealisation as it seems from these studies. Further, Awan et al (2015) found that females tended to produce modal phonation with a small posterior glottal gap, meaning a more breathy baseline which is likely to persist to a degree despite e.g. the CQ increasing in each glottal cycle during laryngealisation. However, since laryngealisation in this study is compared to a modal phonation minimal pair rather than analysed on its own, this issue is less important as the baseline in the individual's glottal gap is controlled for in that way. It could, however, have affected the raw numbers plotted in the acoustic analysis of H1-H2. It should also be noted that H1-H2 has been found useful for creaky voice specifically (e.g. Blankenship 2002, Andruski & Ratliff 2000) and thus is not only a breathy non-modal phonation correlate and can be used despite only considering glottal activity rather than (ary)epiglottal activity.

Another issue prevalent in earlier studies is that H1-H2 is affected by the amplitude and frequency of the formants in the signal. For this reason, H1-H2 is rarely used as an uncorrected measure anymore but is implemented where the influence of formants is corrected for (Chai & Garellek 2022). This was also done in the current study. Even though H1-H2 is widely used there are also findings that even perceptually strong phonation contrasts are not captured well by changes in H1-H2 (Esposito 2012) and the relationship between H1-H2 and articulatory patterns is not fully transparent. This has lead to a discussion of the suitability of H1-H2 as a measure phonation type in a recent paper by Chai & Garellek (2022) who proposes to overcome previous issues with H1-H2 by implementing a new measure called residual H1. Residual H1 relies on normalising the amplitude of H1 against the root mean square of the overall sound energy rather than against H2 and the authors find that residual H1 correlates stronger with the EGG-derived CQ values in their tested data and that residual H1 is better at distinguishing constricted phonation types from modal phonation in particular, making it a potentially more promising measure to use for future studies on laryngealisation. It has yet, however, to implemented into automatic tracking software such as VoiceSauce and to be more widely tested on e.g. different subtypes of laryngealisation and would of course be less successful in distinguishing non-constricted types such as utterance-final, non-constricted creak reported by Slifka (2006). This would be an issue for the present data from Aarhus where glottal constriction is not indicated. Other studies have used a variation of H1-H2 replacing H2 with the amplitude of the second or third formant, labelled H1-A2 and H1-A3, respectively. It is possible that these would have been more suitable options as well as residual H1, at least for the Copenhagen stød types. The choice of using H1-H2 was motivated mainly by its role in characterising laryngeal subtypes in Keating et al (2015) and its previous utility in correlating with compression in the voice during the stød found by Hansen (2015). It is also an open question how well H1-H2 or residual H1 would capture epiglottal activity, a question relevant to the LAM framework.

Moving on to CPP and HNR, both of these measures were significant for all conditions, the CPP and HNR being lower during the stød generally, and during the regular stød in particular, indicating lower periodicity and higher noise. These measures were ranked fairly low in variable importance, however. CPP is generally a less frequently used measure of phonation compared to e.g. H1-H2, but has been found by Esposito (2006; 2010) to be the best out of eight different measures to distinguish breathy from modal phonation. But what about creak? Blankenship (2002) attested it less effective in distinguishing laryngeal from modal phonation but a study by Garellek & Keating (2011) found CPP to also be significant in distinguishing laryngealised vs modal voice. CPP was found by Heman-Akcah et al (2003) to be a more sensitive measure of dysphonia which was thought relevant in the context of laryngealisation and particularly when measuring the stød as there was likely to be irregular vocal fold vibrations that cause unstable voicing. This transient absence of voicing is not unlike patterns seen in dysphonic speech and thus measures which perform better in characterising signals with unstable phonation were thought useful in a study on the stød. A further motivation for including CPP, despite it not being part of the laryngealisation subtype characterisation in Keating et al (2015), was that while the HNR is a correlate of signal noise if it is low, the HNR will remain high during the presence of dampened pulses, even though signal noise might simultaneously be high. This made it relevant to have an additional measure of noise less affected by pulse dampening alongside the HNR to avoid confounding effects of inflation of this measure. CPP was found useful at capturing contrasts in combination with H1-H2 as both the CPP and HNR was found to distinguish between the regular and tonal stød in both dialects whereas H1-H2 did not. Based on Keating et al's (2015) use of H1-H2 as a correlate of glottal constriction, it was speculated that these findings suggest that the stød can be articulated with a noise source independent of glottal constriction, meaning other articulatory mechanisms beyond the shape of the glottis and contact patterns of the vocal folds are utilised to create the contrast. This is congruent with both the proposals of the LAM and findings by Hansen (2015) who found that irregularity during the stød can occur independent of compression in the voice source, evident in the fact that the compression peak and intensity dip (which he interprets as correlating with glottal constriction) can be timed in at least two different ways and thus high compression and glottal constriction cannot be codependently linked because they do not co-occur in time. Fiberoptic recordings of the vocal folds for one vowel on seven speakers in Fischer-Jørgensen (1989) during the stød did show that subjects had signs of contraction in that the distance between the folds was narrower and subjects also had ventricular fold approximation to varying degrees, supporting the impression that vocal fold activity alone is not sufficient for producing and describing rich phonetic variation and thus is potentially not captured fully by EGG.

SHR was included as a measure to investigate the presence of multiply pulsed voice which would result in subharmonic resonances in the signal. This was partly motivated by the multiply pulse subtype in Keating et al (2015) but also by research on the acoustics of the stød by Hansen (2015) who found diplophonia on almost all stød tokens in his material. However, only one speaker was analysed and diplophonia was quantified by measuring H1-H1½, motivated by the idea that a second tone, or subharmonic, would be detectable at half of the f0 frequency. Diplophonia has sometimes been linked to vocal hoarseness and it is possible that the speaker in Hansen's (2015) material had a hoarse voice quality in general, perhaps due to smoking or other lifestyle factors. Whether diplophonia and multiply pulsed creak are the same or different phenomena is difficult to decide based on their descriptions. For the present study, SHR was quite poorly tracked and the question of whether it is a relevant parameter for the stød remains open. Diplophonia has been more qualitatively studied by looking at individual waveforms. This was not feasible in a more quantitative approach taken on in the present study but could perhaps have been done for a smaller subset of tokens. Dejonckere & Lebacq (1983) defined three types of waveform patterns in the EGG signal during diplophonia in pathological voice conditions and it is possible that some of these would be present in the current data if inspected individually, despite the stød being a non-pathological, deliberate use of voice quality and thus might pattern differently. Based on the auditory impression of the data, however, the consistent perception of a double-toned f0 is not thought relevant to pursue further. The tracking of SHR might have improved with different settings. Herbst (2021) writes how preliminary tests in this regard have revealed that the algorithm outcome depends strongly on the parameter settings, particularly the minimum f0, the frame length and the computed upper SHR ceiling. Herbst (2021) even found the that the standard settings tend to be suboptimal. As such, future studies wanting to include this measure should be aware of this potential issue. Unfortunately, this paper only came to my attention recently when adjusting settings was too late.

In summary, the acoustic measures chosen seemed to capture some changes between stød types well and are all commonly used measures of phonation, rendering them previously verified as useful for laryngealisation and the results comparable to the existing body of research. This is perhaps with the exception of intensity which is commonly used in studies of the stød but less commonly used in studies of laryngealisation more generally. One issue with the acoustic analysis was that it suggested differences between stød types that were not congruent with the findings from the EGG data analysis. This issue is discussed further below.

7.3. Obtaining articulatory measures

For this study EGG was chosen to obtain articulatory measures of laryngealisation. This option was deemed appropriate as laryngealisation is thought to, at least in most subtypes, be achieved via a manipulation of the glottal space resulting in changes in the vocal fold contact patterns. It was acknowledged, however, that some aspects of laryngeal manipulation as proposed in the LAM framework may not be captured using this method. That said, EGG is non-invasive and portable, a consideration important for the current study as data was collected abroad in two different labs. Further, EGG has, to the best of my knowledge, not yet been used to study the stød, despite being a popular option for studies on voice quality in other languages. As such, comparing EGG measurements of the stød to other ways of capturing the stød previously used would provide an enlightening novel methodological perspective. It would also contribute to exploring how CQ values correspond to contrastive voice quality cross-linguistically by providing empirical data on a language not measured yet in this way.

The link between the EGG waveform and other articulatory measures is strong and well-established (Fourcin 1974; Lecluse et al 1975; Pedersen 1977, Anastaplo & Karnell 1988, Baer et al 1983a; Childers et al 1990; Childers & Krishnamurthy 1985, Berke et al 1987, Dejonckere 1981, Kitzing 1983, Titze et al 1984, Kitzing et al 1982, Rothenberg 1981, Rothenberg & Mahshie 1988). A limitation of using EGG, however, is that spatial discrimination of the laryngeal tissues is poor, i.e. the signal only records changes in the vocal fold contact area but does not reveal if these changes are anterior, posterior, high or low in the glottis. Further, any laryngealisation not caused by manipulating the glottal shape will possibly not be captured, a potential issue that might be a limitation in this study as the acoustic and electroglottographic findings were somewhat divergent. As outlined in section 2.1.2. and 2.1.3. there are many ways to impact voice quality apart from the vocal fold patterns, including ventricular fold activity and aryepiglottic constriction. Indeed, Fischer-Jørgensen (1989) found ventricular fold activation during the stød when taking EMG measurements from a subset of speakers. As EGG is traditionally used to quantify vocal fold contact patterns, our understanding of how the use of other laryngeal structures impact the EGG output, if indeed at all, is poor. The lack of high correlation between acoustic and articulatory significance, particularly in terms of H1-H2, in the current study suggests that the lack of capturing partially non-glottal laryngealisation is an issue and that EGG might not be appropriate for all types of laryngealisation.

The tracking was done via an automated customised Praat script kindly written by Stefano Coretta. This used the derivative of the EGG signal, the dEGG, to calculate the CQ of each cycle. When inspecting these values, it was evident that some tokens were better tracked than others and that some speakers were better tracked than others. When collecting the data it was difficult to ensure appropriate placement of the electrodes on participants' necks due to the lingering Covid-19 restrictions requiring the collector to keep a physical distance to the subjects. This meant they were responsible for fitting the EGG equipment onto their own necks. There are some guidelines to optimal electrode spacing of around 3-4 cm (Titze 1990) and these were followed to the degree possible under the circumstances. However, even if optimal initial electrode placement is achieved, a drawback of EGG is that the larynx may move up or down during speech production, causing the vocal folds to move above or below the range of the electrodes, rendering the EGG signal unusable (Gick et al 2013: 90). Any absolutely glaring CQ tracking errors were dealt with quantitatively by excluding tokens with any values outside of 0 and 1. In future work it might be useful to filter out outliers based on what CQ values one could reasonably expect in the context of laryngealisation. However, given the stød has not been measured via EGG before, there is no precedence suggesting what language-specific reasonable values are. A suggestion could be to filter out CQ values under 0.40 as they would certainly be too breathy to characterise laryngealisation. It is uncertain, however, whether this would have made a difference for the analysis and further, if this would potentially have filtered out important trends in the data. Given that the lowest values in this study were around 0.45, it is unlikely that such filtering would have made a difference. Another option to obtain more precise CQ values would be to test the effect of different ways to extract these from the EGG signal. There are several methods that have been used to do this previously (see Herbst 2020 for an overview) and further, within the same method, different thresholds can be set for when the EGG amplitude is considered to reach maximum peaks and troughs, taken as the initiation of either the contacting or de-contacting phase in a cycle. Depending on the source of error, it might be fruitful in future work to test specific thresholds for different data and/or derive the CQ values using a few different methods. However, without any other comparable physiological measures in the same study, e.g. glottal airflow from inverse filtering, the researcher would still be in position where deciding which CQ values best reflect the true vocal fold behaviour is difficult.

When analysing the EGG data it was found that the CQ values were lower than expected based on unambiguous cases of creaky voice previously reported. There were, however, some changes and differences in the CQ trajectories. The non-stød trajectories in both dialects were between 0.45 and 0.50 in the mean smoooths. This was also true for the tonal stød CQ values in Aarhus. The tonal stød in Copenhagen rose to CQ values just below 0.60 which was also the case for the regular stød when the stød basis was a long vowel in an open syllable, whereas the regular stød rose to 0.55 when the stød basis was a short vowel and consonant in a closed syllable. In Aarhus the regular stød tokens ended at 0.53. The question in this regard is how much a CQ value needs to increase to be relevant as a production of a laryngeal contrast and at what threshold this can be reliably perceived as such by listeners.

Going by the results from the GAMM model comparison, the CQ values were only statistically significantly different between the Copenhagen non vs regular stød tokens which amounted to an increase from around 0.50 to 0.60. The non-stød tokens declined slightly to 0.45 towards the end but the difference smooth showed significance from around the 18% timepoint, meaning this dip was not necessary for statistical significance for the trajectory

This suggests that a reasonable threshold for significance might be around a 0.10 CQ increase or decrease, supported by the fact that an increase from 0.50 to 0.55 was not large enough to yield a significantly different result in the model comparison between the regular vs tonal Copenhagen stød. An issue remains, however, with the absolute CQ values between dialects. Only dialect-internal values were compared in the models but considering a CQ of 0.53 for the regular stød in Aarhus which does not appear particularly laryngealised based on auditory impression when going through the data, compared to a 0.60 CQ value for the Copenhagen regular stød which is highly audible as laryngealised, it raises the question of how this inter-dialectal difference in CQ values of 0.07 is very auditorily significant when the intra-dialectal comparison found a 0.05 increase to be insignificant? And further, it raises the issues of what range of CQ values can be expected for laryngealisation outside of prototypical creaky voice, both in Danish and in languages with similar contrastive phonation?

Turning to other studies for answers, Avelino (2010) measured laryngealised vowels in Yalálag Zapotec and found that they had a larger closed quotient than modal vowels with changes in a similar range to the present study: 64% for laryngealised vowels versus 59% for modal vowels. This was measured and reported as closed quotient percentages but should be highly comparable to Contact Quotient proportions used in the present study. Using one-way ANOVAs Avelino (2010) showed the observed differences to be statistically significant, suggesting that what appears to be relatively small changes in vocal fold patterns, i.e. 5% difference, might be enough to constitute a phonatory contrast in this language. In another study on contrastive phonation DiCanio (2009) measured the Open Quotient for different registers in a handful of speakers of Takhian Thong Chong. Results showed a fairly large inter-speaker variation in value ranges, but most between 45-60%. It should be noted that DiCanio (2009) measured the Open Quotient, not the Contact Quotient, and that most of these phonatory contrasts were not laryngealised except for the tense register. However, the reported ranges still serve as an indication of the expected differences in EGG values in contrastive phonation and these were, just like Avelino's (2010), very similar to the overall range in the present study. The fact that even modal phonation is reported in Avelino (2010) as having vocal fold contact phases in each glottal cycle as high as 59% makes it even more difficult to interpret the findings in the present study as the regular stød in Copenhagen was taken to be laryngealised at a CQ of 0.60, based on comparing this value to the acoustic correlates and the auditory impression of laryngeal stød vs no stød. The relative difference between Avelino's (2010) modal and laryngealised vowels was 5%, highly equivalent to the

0.05 CQ difference observed between the regular and tonal stød in Copenhagen. In the present study, this difference was not significant, whereas Avelino (2010) found that it was for their data. The contrasts, of course, were different, given that Avelino compared modal to laryngeal phonation, whereas the comparison here was regular vs tonal stød which are both hypothesised to be laryngealised. The statistical tests used were also different. Both of these facts could possibly influence the result and thus constitute possible reasons for the difference in significance of statistical outcomes. There is also a possibility that contrastive phonation ranges are language-specific to a degree, or even phenomenon-specific, perhaps changing depending on not only the type of phonation but also the subtype of such contrast. This is an intriguing proposition but very difficult to gauge from the existing literature as most research on contrastive non-modal phonation does not divide these into subtypes but rather investigates them as one single phenomenon. However, as a preliminary observation in this regard, a cross-linguistic acoustic phonetic study on four languages, Gujarati, Jalapa Mazatec, White Hmong and Yi, all with multiple phonation type contrasts, found that the acoustic correlates of these contrasts did not cluster by contrast (Keating et al 2010). This indicates that phonation types do have a degree of language specificity as breathy vowels in one language are acoustically distinct from breathy vowels in another language. This may offer an explanation for how the maximum observed CQ value in the present study is 0.60 on laryngealisation whereas Avelino (2010) reports a 59% closed quotient for modal phonation. This makes it even more compelling to enrichen the empirical data pool of cross-linguistic phonetic differences in phonological contrasts to enable a robust account for both language-specificity and phonation subtype specificity.

It would be interesting to follow these results up with a study testing the correlation between CQ values and perception of the stød as, ultimately, what matters most in a phonemic contrast is how it is perceived by speakers to make meaning of the speech signal they encounter. The results also suggest that there are articulatory mechanisms active during the stød that are simply not captured by using EGG. This hypothesis is supported by Fischer-Jørgensen's (1989) findings of muscle activity in the ventricular folds during the stød and further, supported by the LAM suggesting that all effects that can be characterised as laryngeal are produced with involvement of the aryepiglottic constrictor mechanism, with tongue retracting and larynx raising as secondary mechanisms (Esling et al 2019). These articulatory gestures are unlikely to be precisely captured by EGG, if they can be captured at all, and more work is needed to establish whether EGG can detect ventricular fold activity and/or other possible laryngeal mechanisms relevant to laryngealisation. Based on the fact that some of these mechanisms, like larynx raising, might obscure the electrode placement during EGG data collection, this method is probably not appropriate for all types of laryngeal contrasts, and it is also entirely possible that individuals employ different articulatory strategies to create contrastive laryngealisation, meaning some speakers might be appropriate for EGG, some might not. The relatively heterogenous result regarding inter-speaker variation in how the CQ values were tracked suggests that this might be the case.

7.4. Acoustic and EGG data correlation

For the direct correlation between the acoustic and electroglottographic measurements, these were time-aligned and random forests were fitted for each speaker with CQ values as the outcome variable and each acoustic measure as a predictor variable. The resulting variable importance rankings were fairly homogenous between dialects, all except one favouring H1-H2, the outlier being the Aarhus non-stød vs regular stød contrast favouring intensity. However, a weakness of testing variable importance with random forests was that significance of each variable for predicting the outcome variable is not indicated, meaning even the highest ranked predictor might be a poor predictor, or that even the lowest ranked predictor might be significantly correlated with the CQ values. To uncover this relationship, other methods would be needed for future work. One possibility is to fit individual GAMMs modelling CQ values against acoustic variables to each speaker individually for each stød contrast in the same way that was done for all speakers combined in the GAMM analyses. This, however, was not feasible for the present study and this method would also suffer from the reverse issue of the random forests, namely that all acoustic predictors might be deemed statistically significant and thus the question of which is the best correlate cannot be answered. This scenario is very likely based on the results presented here and the fact that all measures are hypothesised to capture laryngealisation. This could be overcome by combining individual GAMMs with the rankings from the random forest models, running the GAMM model comparisons first and then only including acoustic predictors into the random forest that were statistically significant, thus only ranking variable importance for variables that are proven to be significant in predicting the difference between stød types.

When deciding whether laryngealisation was different between the respective stød types in the analysis there was a second issue with the fact that the acoustic measures and the CQ values were not always aligned in terms of significance. To remind the reader where the discrepancies lie, each measure and its associated significance is given in a table below for each stød type per dialect:

Dialect	Contrast	f0	Intensity	H1-H2	CPP	HNR	CQ
Copenhagen	Non vs regular stød	+	+	+	+	+	+
Copenhagen	Regular vs tonal stød	-	+	-	+	+	-
Aarhus	Non vs regular stød	+	+	+	+	+	-
Aarhus	Regular vs tonal stød	-	+	-	+	+	-

Table 22. Significance of each acoustic and articulatory measure indicated with a + for significant and a – for not significant, aggregated by stød type and dialect.

The only contrast with significance for all measures is the Copenhagen non vs regular stød, a result that was congruent with the expectation of contrastive laryngealisation during the stød. For the Copenhagen regular vs tonal contrast, acoustic and articulatory measures were not expected to be significantly different, yet intensity, CPP and HNR seemed to distinguish the stød types, whereas f0, H1-H2 and CQ measures did not. For the Aarhus dialect, the non vs regular stød contrast was distinguished by all acoustic measures but not the CQ values. The regular vs tonal contrast was distinguished by intensity, CPP and HNR but not f0, H1-H2 and the CQ values. In Keating et al's (2015) characterisation of creaky voice subtypes they use H1-H2 as a primary correlate of glottal constriction. This fits well with the pattern that the regular vs tonal stød in both dialects have a non-significant H1-H2 and non-significant CQ values. However, for variable importance, H1-H2 was the best correlate for CQ values. As discussed in section 9.3., previous studies have generally shown a good correlation between phonation and EGG output, so it must be considered why this does not appear to be the case in the present study. The sources of these discrepancies could be some of the methodology-related technical errors discussed in section 7.1., 7.2. and 7.3., including electrode placement, cycle tracking, acoustic measurement tracking and so on, despite the methodology generally being similar to that of previous work on contrastive phonation in other languages. An interesting reason suggested earlier in the discussion for the lack of acoustic and articulatory correlation was that it might indicate that the stød can be articulated with a noise source independent of glottal constriction and thus would not be captured well via EGG. This is supported by (i) the finding that the ventricular folds are also active during the stød (Fischer-Jørgensen 1989), (ii) that the timing of compression in the voice during the stød can be independent of the onset of irregular vibrations in the signal (Hansen 2015), and (iii) that the LAM shows that entire larynx can be an active articulator with lots of ways of creating phonetic variation (Esling et al 2019). It was also hypothesised that phonation types have a degree of language specificity, supported by findings from Keating et al (2010), and that there might even be a degree of phonation sub-type specificity. This would mean that glottal constriction is perhaps more salient in the non vs regular stød contrast compared to the regular vs tonal contrast for instance. Ultimately, more research is needed to tease out these effects and in doing so, it could be insightful to supplement a quantitative approach with a more qualitative approach, inspecting subgroups of tokens of different phonation types. Further, methods that more directly visualise the laryngeal structures would be beneficial to separate the effects of methodology from the effects of non-glottal laryngeal articulations affecting voice quality. This would be particularly enlightening to establish why acoustic correlates of laryngealisation, particularly signal noise, during the stød are significant whereas EGG measures are mostly not.

Having discussed the more study-internal factors with which to interpret the results, the attention will now be turned to the broader perspectives, starting with a discussion of the findings within the Danish stød research, and then continuing with discussing broader perspectives on cross-linguistic contrastive phonation and the results reported here.

7.5. Comparative perspectives on previous research on Danish stød

The major findings from previous phonetic research on the stød were outlined in section 2.5. and related sub-sections. These findings will briefly be discussed here in relation to the results of the present study. This thesis was concerned with measuring phonetic variation in the stød in two different dialects to explore variation in phonological voice quality contrasts using Danish as a case study. Previous phonetic studies on the stød have been conducted only on the Copenhagen dialect and as such, previous work can only be directly compared to the data from the Copenhagen speakers in the present study. However, having data on the Aarhus speakers nevertheless provides a novel and interesting perspective on what the stød is and can be phonetically, and in extension of this, what range and type of phonetic variation phonological contrasts can exhibit in their production. The overall results confirm that intensity is a relevant acoustic correlate of the stød. It should be noted, however, that previous studies (Riber Petersen 1973, Fischer-Jørgensen 1989) have approached the stød syllable a bit differently by dividing it into two parts, an initial phase followed by the stød phase ‘proper’. In the initial phase there was a strong tendency for f0 to be higher in words with stød, particularly in disyllables with a long vowel. The initial f0 peak was clearly highest in words with a vowel + a sonorant consonant. In the second phase of the stød f0 tended to fall on approximately half of the stød tokens whereas for non-stød tokens f0 were flat or slightly rising. The degree of f0 drop was very variable, however. For intensity, a falling trajectory for syllables with stød was reported and the intensity minimum was lower compared to non-stød syllables. Fischer-Jørgensen (1989) linked the f0 and intensity patterns to articulatory mechanisms via obtaining EMG measurements of laryngeal muscles. The cricothyroid muscle showed greater activity in the initial phase of the stød compared to non-stød which matched the higher f0 as the cricothyroid muscle lengthens the vocal folds, making them thinner and more tense which in turn raises f0. Cricothyroid activity generally did not, however, fall in the second phase of the stød, meaning a lack of correlation between the f0 decline and de-activation of the cricothyroid muscle. The lateral crico-arytenoid muscle also helped raise f0 in the initial stød phase and both this muscle and the vocalis muscle had activation peaks constituting a peak in these syllables, suggesting these muscles combined are responsible for achieving a degree of glottal constriction during the stød. This in turns lowers f0 and intensity.

This pattern is not entirely consistent with those observed for the Copenhagen dialect in the present data where the f0 trajectory in stød syllables is characterised by starting unusually high but do have a peak and a decline. Despite this decline, the trajectory still ends higher than the non-stød tokens. It was speculated that the relative f0 movement rather than the absolute values were relevant for the contrast. The EGG data for the Copenhagen speakers suggested a degree of glottal constriction also congruent with Fischer-Jørgensen’s measurements of muscle activity in the larynx.

Neither Riber Petersen (1973) nor Fischer-Jørgensen (1989) measured HNR, H1-H2 or CPP, meaning the results of the present analysis of these measures are not directly comparable to their findings. Assuming that HNR is a good correlate of irregular vibrations, however, it can be noted that Fischer-Jørgensen (1989) found irregular vibrations on 70% of stød tokens and that HNR did distinguish the Copenhagen stød contrasts, but HNR was not a good predictor of the CQ measures. The more recent study on the stød by Hansen (2015) used different acoustic measures to capture the stød, among these H1-H2, Center of Gravity

(CoG) and H1-H1½. He generally found the stød to be extremely varied and not consistently well-captured by any of these measures, except for H1-H1½ which was interpreted as diplophonia being present on most tokens. As mentioned previously, this analysis was only done on one speaker and thus might be unique to their voice quality. The main hypothesis tested in Hansen's (2015) study was that the stød is realised as a brief change in voice quality towards a compressed or creaky voice and back, i.e. a dynamic voice quality gesture involving an adequately large change during an adequately short time interval. This hypothesis was rejected based on the results, a finding congruent with the present study as neither the acoustic nor the articulatory measures correlating with the stød where found to be transient within the syllable as the measures did not return to any pre-stød-onset level after their initiation. Rather, if the stød initiated the decline or increase of a measure, this pattern overwhelmingly continued until syllable termination. Despite this, Hansen (2015) stressed that the stød vs non-stød tokens still differed in the compression course and degree, the stød tokens being more dynamic and compressed to a higher degree, so while the specific hypothesis is rejected, he argued there is still much to gain from measuring the compression degree in voice quality during the stød. Indeed, the present study confirms this, particularly when adding different dialectal perspectives on the manifestation of the stød as shown by very variable and dynamic patterns in the acoustic GAMM smooths.

The finding that the stød appears to be more tonal in its production in the dialect of Aarhus provides an interesting perspective on the interplay between laryngealisation and tone. Section 2.6. provided some historical perspectives on how the stød developed and how it is related to the contrastive tonal Scandinavian word accents. There appears to be most evidence in favour of the stød developing from an originally tonal accent in Old Scandinavian. The other Scandinavian languages retain a tonal contrast in word Accent I and Accent II, and the stød is most phonologically similar to Accent II. Having looked at evidence of a tonal stød in Aarhus in the present study it is tempting to compare it to the tonal contrast in the Scandinavian word accents. However, the phonological requirements between the two are fairly different as the word accents require at least two syllables for production whereas the tonal stød tokens in Aarhus were produced on monosyllables and the stød is a property of the syllable rather than the word. Nonetheless, the stød has been analysed phonologically as a phonetic realisation of an underlying falling tone on one syllable (Ito & Mester 1997) due to Danish having High-Low pitch contours associated with the stressed syllable. Other scholars reject this interpretation as Modern Standard Copenhagen has a Low-High-Low intonational contour (Grønnum 1992, Vazquez-Larruscaín & Basbøll 2013) and further, the stød appears to autosegmentalise on a separate tier from sentence intonation (Goldshtein 2020). A proposal outlined in section 2.6. was that that the source of the stød was likely to be the phonologisation of a historically non-contrastive laryngealisation on tonal accents (Goldshtein 2020). Articulatory biases rooted in aerodynamic correlates of prosodic boundaries implemented by a low tone resulted in irregular phonation being perceived as laryngealisation which then entered the pool of phonetic variance. Laryngealisation enhanced the prosodic boundary of the tonally manifested contrast and was gradually reinterpreted as a phonological target rather than a phonetic variant due to a shift in cue weighting. This proposal is relevant to the finding that PSC occurs even on a handful of tonal stød tokens in the Aarhus dialect which seems peculiar given that there is no apparent laryngealisation to promote any continued post-syllabic creak. However, if low tones are likely to be enhanced by laryngealisation, this makes the PSC observation on these tokens less surprising, although it still does not account for why PSC is not found on all tonal tokens. Further, the tonal realisation of the stød in Aarhus was interpreted primarily via the CQ values and the auditory impression of their production, and as such were not quantified by any measurable tonal behaviour but rather by the absence of articulatory evidence of laryngealisation. The hypothesis that the contrast is tonal and that laryngealisation might enhance this contrast would need to be researched separately, with more stringent ways of quantifying the tonal contrast by measuring e.g. semitone intervals or other ways to quantitatively account for the f0 patterns that manifest the contrast and how it is different from the general word pitch of the accents. The f0 patterns generally did not look as expected for the Aarhus tonal contrast so more research is certainly needed to investigate and support these suggestions.

To sum up the results overall, the present study divided the stød into types rather than phases and introduced a comparative perspective on dialectal variation, making it less directly comparable to previous mono-dialectal studies. However, findings generally mirrored previous results on the phonetics of the stød well, suggesting that it correlates with intensity and that it is also highly variable in its acoustic manifestation. Findings in f0 correlated less well with previous studies as the f0 behaviour in the present data did not show the prototypical peak and steep decline expected. The stød was also found to also correlate with declines in HNR and CPP in both dialects. The EGG data analysis was a novel technique in the context of the stød and showed some laryngealisation in Copenhagen compliant with previous studies and this was not found in Aarhus to the same degree. Since data on Aarhus is lacking, it is difficult to discuss this aspect in the perspective of previous stød research. However, other novel findings were obtained, particularly in terms of the potential segmental span of the effects of the stød and how it is timed. These are discussed in the following section.

7.6. Perspectives on stød timing patterns

Research question 3 in this study explored the timing of non-modal phonation production within the syllable on which it is contrastive. This was motivated by previous studies on other languages which have found this timing to be very variable. Thus, documenting variation in stød timing in Danish as a case study was a way to contribtue more empirical data on the possible ranges of this variation, adding to the cross-linguistic typology of phonetic variation in phonological contrasts of voice quality. It has previously been attested that the stød timing is highly variable (Fischer-Jørgensen 1989, Grønnum & Basbøll 2001) but some general trends could be teased out. Riber Petersen (1973) reported that in cases where the stød manifested as irregular vibrations, the duration of the stød was consistently around 1/3 of the total vowel. She also reported that when the stød was articulated on the vowel in closed syllables the f0 declined later and reached its minimum closer to the termination of the vowel segment. Fischer-Jørgensen (1989) found f0 to decline quite late in the stød syllable and quantified the timing with measurements of the vocalis muscle which consistently peaked for 5 speakers around 20-40 ms into the vowel with stød. The drop observed in f0 had its onset immediately before or right at the vocalis muscle activity peak, suggesting that the stød in these speakers is timed 20-40 ms into the vowel. Hansen (2015) found stød tokens to generally have a rising-falling compression course based on the trajectory of the intensity, but the timing of it varied in two major ways so that either (i) the compression peak and the intensity minimum occurred close to each other and simultaneously with irregular vibrations, or (ii) the compression peak occurred before the intensity minimum (approximately 80-100 ms before) and before the onset of irregular vibrations. Whether this timing is random or related to which segment(s) constitute the stød basis is not clear from these results but would be an interesting comparative perspective to adopt on findings of the stød subtype division employed here to explore any correlations between Hansen's (2015) type division and the regular vs tonal division.

The current study found variability in stød timing which could be divided into two groups: 1) Measures that were statistically significant throughout the entire syllable, and

2) Measures that were statistically significant mostly in the latter part of the syllable.

There was a tendency in both dialects for measures in the non vs regular stød contrast to belong to group 1 (f0, intensity, H1-H2, HNR) and a tendency for measures in the regular vs tonal stød contrast to belonging to group 2 (intensity, CPP, HNR). This suggested timing differences between stød types more than between dialects, meaning phonetic variation in timing likely depends on the stød basis.

The timing, however, was not measured in the same way as the previous studies cited which all used milliseconds from either vowel onset or milliseconds before the onset of irregular vibrations to quantify the stød timing. In the present study timing was instead quantified using the difference smooths from the GAMMs in the acoustic and articulatory analysis which indicated where in the trajectories the stød types differed the most. The approach of measuring milliseconds on the stød basis is a method that is only feasible on tokens which have a clear demarcation of when the stød occurs. This was difficult to determine in many cases, a trend mirrored in previous research. Fischer-Jørgensen (1989) only found irregular vibrations on 70% of stød tokens, meaning if adopting Hansen's (2015) approach to measuring the stød via the timing of the intensity peak relative to irregular vibrations, this would exclude 30%, or about one third of the data from being accounted for. Further, as at least one variant in Aarhus was hypothesised to not have any laryngealisation at all, measuring the timing of the stød relative to irregularity in the signal would not be feasible.

Measuring timing based on difference smooths offers the advantage of not having to qualitatively decide when the onset of the stød is and what it correlates with, avoiding the ambiguity of its variability. However, this method also offers some disadvantages, besides making the results harder to compare to previous findings. Firstly, it only measures differences between contrasts rather than measuring trends in the individual trajectories. These patterns are of course related but it is possible that there are tokens where the onset of the stød and onset of the significance of the difference smooth are staggered. Another disadvantage is that it pools all tokens together to create an average trajectory for each dialect stød type. This means there might be significant inter- and intra-speaker variation not uncovered with this approach. In a study on phonetic variation this is obviously a limitation. However, given the amount of tokens analysed, accounting for individual token variation would be a messy affair unless a few very clear patterns emerge which is entirely possible based on Hansen's (2015) findings of two major types, with the caveat that this data was from one speaker only, meaning patterns could be rather idiosyncratic.

With this in mind the Copenhagen non vs regular stød contrast is likely the most directly comparable to previous work on the stød and thus mostly relevant to discuss in the light of this research. This stød contrast had f0 and intensity differences in the GAMMs starting at the onset of the stød basis and lasting more or less the entire remaining syllable. This is different than previous work outlined above reporting that f0 declines rather late in the syllable during the stød, although a relative decline was seen in the mean GAMM f0 smooth. The Copenhagen non vs regular stød had an onset of CQ value divergence at the 18% timepoint. Comparing this to Fischer-Jørgensen's (1989) findings that the f0 and intensity patterns correlated well with the vocalis muscle activity, the CQ values here suggest the onset of laryngealisation extremely early in the syllable. The Aarhus non vs reg stød also had a significantly different f0 throughout the entire syllable up until 90%. The CQ values here were not significantly different, however.

Variability in timing is tempting to explain by the segmental differences in the stød basis, being either a long vowel or a short vowel and sonorant consonant. However, Grønnum and Basbøll (2001) did not find that the onset of the stød generally differed between syllables with long vowels or short vowels + consonantal rhymes, although they did find minor timing differences on closer examination, reporting that in long vowels the stød onset timing occurred around the vowel midpoint, whereas in short vowel + sonorant consonant rhyme, the stød onset was in the last third portion of the vowel. This study did not, presumably, include the third option, stød basis consisting of a long vowel but occurring in a closed monosyllable which in the present study was labelled tonal stød.

An additional sub-analysis of timing was carried out on the segmental scope of the phonetics of the stød, prompted by the discovery that there was visible and audible prototypical creaky voice after the segment receiving stød on a surprising number of tokens. This was labelled PSC and the analysis of it accounted more systematically for this phenomenon of creaky voice lag after the stød, noted previously by e.g. Grønnum and Basbøll (2007: 199). It was found that the phonetic effects of the stød could span up to four segments after the stød segment, but most prominently occurring on the segment immediately adjacent to the stød. What made the finding particularly interesting was that PSC was statistically significantly correlated with both dialect and stød type, being much more prevalent in the Copenhagen dialect, and on the regular vs tonal stød contrast. There was also large inter-speaker variability in PSC prevalence, making it a potentially good candidate for further sociophonetic and forensic phonetic research. That its prevalence seems highly varied can be confirmed by a very recent study also reporting on stød timing which found that the stød phonetics extended to the following word in a majority of cases. This study by Peña (2022) investigated the timing of the stød relative to the syllable on the assumption that it is biphasic, much like Fischer-Jørgensen (1989) who divided the stød into an initial phase and the stød phase ‘proper’. Peña (2022) adopts the framework of Articulatory Phonology to investigate the timing and proposes a division of the traditional ‘glottal’ tier in Articulatory Phonology into two tiers to account for the biphasic stød, one related to phonation and one related to f0 to capture the varying independent effects of multiple glottal configurations on phonation and f0. Data from the present study possibly confirms the utility of such an approach as acoustic differences in intensity, CPP and HNR were found between measures for the Copenhagen regular vs tonal stød despite no significant difference in CQ values, suggesting differences between the types are achieved by more detailed articulations from other sources than purely the glottis shape/vocal fold contact area. It should be noted, however, that the regular and tonal stød types were not distinguished by f0 differences, somewhat questioning the utility of the assumption that independent effects of multiple glottal configurations affect f0. Participants in Peña’s study were 9 speakers from Copenhagen, i.e. highly comparable in both dialect and number to the present study. Based on the results, timing descriptions were reported as the first phase being timed relative to the first half of the syllabic rhyme, the second phase being timed relative to the second half of the syllabic rhyme. Very interestingly for the PSC observed in the present study, Peña (2022) found that the stød extended to the following word in 88.36% of tokens, making PSC the norm rather than the exception in the data. This was calculated by pooling all stød occurrences together whereas in the present study the PSC occurrence was accounted for by dividing the stød into two types with the regular type in Copenhagen having PSC in 21.7% of tokens and the tonal stød having PSC 58.7% of tokens. If pooling both stød types together, the PSC in Copenhagen occurs on 29.2% of tokens, considerably less than in Peña’s speakers. Peña measured the occurrence of the stød as periods of decreased f0 and intensity or the visual presence of creaky phonation in the spectrogram. As such, the difference in PSC prevalence is quite surprising as Peña’s methods were not dissimilar to those of the current study, although for the PSC analysis here only visual and auditory cues were used to quantify the phenomenon. This again demonstrates the variability of the phenomenon. However, it might also be partly attributed to the difference in the stød syllables elicited. Peña (2022) had included a type not included in the present study, namely syllables with a CVːˀO structure as in the word *gås* [g̊ɔːˀs] ‘goose’. A syllable with an unvoiced obstruent would not constitute a suitable minimal pair for the regular vs stød contrast and thus this type was not included in the present data elicitation.

The variable findings in timing provide compelling evidence that there is reason to investigate stød timing more thoroughly, uncovering the extent of inter- and intra-speaker variation, the effects of stød basis and its utility for other linguistic fields. These findings also provide a valuable contribution to the typology of cross-linguistic segmental span of contrastive non-modal phonation, defying previously reported patterns where the phonation contrast usually spans less than the segment duration, presumably for maximum contrast in perception. In extension of such a cross-linguistic perspective, the next section discusses the results from the current study in the light of other languages with contrastive non-modal phonation more broadly to offer a perspective on how the data here compares to the trends outlined in section 2.4.

7.7. Comparative cross-linguistic perspectives

In section 2.4. some cross-linguistic patterns in phonologically contrastive non-modal phonation were outlined, the purpose of which was to uncover some of the phonetic variation in the manifestation of these contrasts in different languages. This allows the findings from the current research project to be situated within the already existing pool of typological knowledge and enables the identification of how the patterns found in the Danish stød compare. This should be prefaced by a reminder of the fact that the stød is perhaps less like other contrasts phonologically as it is classified as a phonological property of the entire syllable. In many other languages with voice quality contrasts these take the characteristics of more conventional suprasegmentals where the contrast is a property of the segment it occurs on. As we have seen, however, despite this phonological difference, the stød both can and has been said to be 'received' by a single segment in the syllable with stød basis (e.g. Fischer-Jørgensen 1989). Indeed, the entire division of the stød into types in this study was motivated by the fact that the stød basis does not always span the entire syllable, creating phonetic variation. As such, this variation in stød patterns is arguably still highly relevant to a developing typology on cross-linguistic patterns in contrastive non-modal phonation.

Some predictions for the patterns of contrastive phonation were outlined where Silverman (1995/1997) proposed that it is sequenced to ensure optimal perception and that this could manifest itself in a few ways: (i) placing the phonation contrast on the first portion of the vowel for early auditory detection for the listener, (ii) having a transient portion of non-model phonation in between two stretches of modal phonation for maximum contrast, or (iii) non-modal phonation types are in themselves salient enough that they do not require perceptual enhancement and therefore may last the entire duration of a vowel. The onsets of the acoustic differences between stød types in the present study were fairly varied between types but were divided into two groups:

1) Measures that were statistically significant throughout the entire syllable, and

2) Measures that were statistically significant mostly in the latter part of the syllable.

There was a tendency in both dialects for measures in the non vs regular stød contrast to belong to group 1 and a tendency for measures in the regular vs tonal stød contrast to belonging to group 2. This suggested timing differences between stød types more than between dialects. In terms of Silverman's three options, group 1 could be said to be compatible with pattern (iii) of spanding the entire duration of the segment. Silverman's pattern (ii) is slightly ambiguous as it can refer to either a very brief transient non-modal stretch within the segmental boundary surrounded by modal phonation, or a transient non-modal stretch in a given proportion of an utterance which may span multiple syllables. The former interpretation is not congruent with the Danish stød patterns here as the laryngealisation, in the vast majority of cases, persist throughout the remaining syllable duration after their onset. However, the second interpretation largely describe the patterns of stød timing where they start somewhere after syllabic onset and eventually are discontinued as in group 2. The PSC analysis showed that when the stød manifests as audible and visible creak, these effects may span up to four segments after the stød segment. Such post-segmental phonatory effects on adjacent segments have previously been reported for non-modal phonation occurring on consonants in other languages, labelled post-glottalisation. In these cases, the creak starts approximately in the middle of the consonant and lasts well into the following sound (Hargus 2016, Howe & Pulleybank 2001, Ladefoged & Maddieson 199, Silverman 1995/1997). Similar patterns were evident for the stød where PSC on the adjacent segment to the stød was by far the most prevalent PSC pattern. However, the segments were different to previous reports in other languages. The most common stød type to have PSC was the tonal type where the stød belongs to a long vowel which is followed by a consonant in the syllabic coda. This is the reverse pattern of the reports of post-glottalisation where other languages had a consonantal contrast extending into the following vowel. Whether languages with vocalic rather than consonantal suprasegmental non-modal phonation would also exhibit such effects is not known. It can, however, be said that the post-segmental phonation can appear regardless of segment type, meaning the segment type is not the sole cause of the occurrence.

From a typological angle, the variable timing of the stød makes it difficult to make any sweeping conclusions about how it patterns compared to other languages. What can be said, however, is that the variability might in itself be a characteristic of the stød, but seemingly not of phonation contrasts in other languages, or at least to a lesser extent. Section 2.4. reported how in Chong, the voice quality changes started to manifest around 50% into the vowel (DiCanio 2009). This was also the case for Santa Ana del Valle Zapotec where laryngealisation was realised towards the end of the vowel (Esposito 2003). Both of these patterns were quite consistent, meaning less variability than the stød realisation. Variability can also be a more defined feature of other languages, e.g. Yucatec Maya which is reported as having three timing variations, namely (i) laryngealisation articulated as a full glottal stop separating two surrounding stretches of speech with modal phonation, (ii) vowels with modal phonation with a transient stretch of laryngealisation in the middle of the vowel and (iii) vowels starting with modal phonation and increasingly receiving laryngealisation towards the end of the vowel (Avelino 2010). This three-way categorisation, however, suggests that the variability is more structured than for the stød with more discernible patterns in the phonetic variation.

Another issue with timing comparisons is deciding what exactly marks the phonation contrast, an issue that might be particular to the stød but possibly also for other languages. Some measures in Aarhus differentiated the stød contrast not by the final decline in measure but rather by a much higher early peak in one type but not the other. Such early stage differences were also found by Fischer-Jørgensen (1989) who described the stød realisation phonetically in two distinct phases, stage one having raised f0 and intensity. That the first part of the stød syllable might be important is supported by a pilot study on the perception of the Modern Standard Copenhagen stød which found that the acoustics of the stød correlated with the perception in unexpected ways (Thorsen 1974), ways that are relevant to the question of when the stød can be said to have its onset. Part of the study aimed to test how much of the syllable is necessary for perception of the stød by cutting out successively larger and larger chunks of the syllable and testing when perception was impaired. Stød tokens with visible irregular vibrations in a spectrogram were not perceived until the onset of these vibrations which, according to Fischer-Jørgensen's (1989) study is rather late in the syllable, meaning a good stretch of duration is needed for perception. However, some stød tokens were spectrographically invisible, meaning no visible irregular vibrations or any other indication that they contained stød, but nonetheless highly perceptible as stød syllables. For these tokens, the perception occurred much earlier in the syllable, shortly after the onset of the vowel. Thorsen (1974) suggested that f0 might be used as a secondary cue for stød perception due to differences in intonation between word pairs. These results, while concerned with perception rather than production, support that there are different subtypes of stød and suggest that the timing of the stød is different based on perception, but that it might be difficult to capture these. It also suggests that for some stød types, phase one is critical, whereas for others, the stød phase ‘proper’ is the relevant phase. For the tonal stød in Aarhus, this becomes an even more interesting perspective as these are assumed to be cued by f0, but it has yet to be systematically studied how the timing differs from standard Copenhagen stød realisation subtypes where f0 is used as a perceptual cue. In summary, comparing stød timings in a cross-linguistic typology is difficult because it is difficult to establish what, and when, the stød is. It appears that there is some interplay between non-modal phonation and tone that is key to characterising the phonetic variation, however, and this interplay will be discussed in the remaining portion of this section.

Section 2.4. also described languages with an interaction between laryngealisation and tone. A recent and updated overview of cross-linguistic patterns in phonation types (Esposito & Khan 2020) provides an interesting typological perspective on languages with interacting phonatory lexical contrasts relevant to the dialectal variation found in Danish in this study. According to this review, languages broadly pattern into two groups where they can have (i) tone and phonation as independently contrastive, or (ii) a contrast in phonation, usually on the vowel, but no contrast in tone. For the latter, the authors note that this division is not so clear cut, however, as the line between phonation and tone is often blurred. They report that in almost all languages with a phonation contrast, f0 is a major part of the realisation specification, despite tone not being lexical. Thus, the classification of languages into categories of being either contrastively tonal or contrastively phonatory can appear arbitrary when the realisation of the contrast clearly involves both phonetically. Interestingly, Esposito and Khan (2020) specifically mention Danish as an exception to this realisation specification, mentioning that f0 is not a reliable cue for laryngealisation during the stød. Thorsen’s (1974) findings suggested otherwise. The current study found f0 to distinguish the non vs regular stød but not the regular vs tonal stød, partially supporting that, at least in terms of statistical significance in production, f0 might be an unreliable cue.

There is general tendency in the literature on the stød to make statements about it as though it is one thing, despite the rich amount of variation, probably because it is phonological and thus can be described as a phonological unit rather than a phonetic chaos. The current study suggests that such generalised statements might be true for some stød types but not for others. This makes it more complicated to fit the phonetic behaviour of the stød into a typological classification in a cross-linguistic comparison but it also means that there is a rich source of variation to explore from many different perspectives and it further encourages more research into previously attested languages with phonological voice quality contrasts to explore dialectal, or other sociophonetic, variation.

Some dialectal variation has already been uncovered, e.g. in Scottish Gaelic that was outlined in Section 2.4. as having dialectal variation in the phonetic realisation of a phonological Class 1 and Class 2 distinction based on metric structure. An interesting consequence of the metric analysis was the observation that the Class 1 vs Class 2 distinction is dependent on the length of the vowel – if it is short, the form is Class 1, if it is long it belongs to Class 2 (Morrison 2019: 395). The relationship between vowel length and type is also relevant for the stød as vowel length distinguishes the stød basis, and thus the stød type, in monosyllabic, closed syllable words. As such, the tonal stød occurs if the vowel is long and a regular stød occurs if the vowel is short. The difference between Scottish Gaelic patterns and Danish patterns are is that in the former, each dialect has the same phonetic realisation in the two classes, whereas in Danish, the change is within the dialects themselves.

In summary, the stød compares to other languages on some parameters but also appears to be much more variable in many respects. This makes a study on its phonetic variation a valuable addition to the already existing literature on phonetic realisation of phonological voice quality contrasts. The stød had more variation in timing of the contrast in particular but also on the fact of having distinct types with relatively indistinct acoustic correlates.

7.8. Perspectives on the LAM and suggestions for future directions

As many questions as this research project have sought to answer, just as many new questions have arisen in the pursuit to do so. This section sums up some of the unexplored future directions this research has exposed.

The framework for the study was the LAM that proposed the larynx as an active articulator in laryngealisation types. In the data some discrepancies were found between significance for different acoustic measures, for instance the finding that CPP distinguished the regular vs tonal contrast in both dialects but H1-H2 did not, despite both measures being correlates of laryngealisation. Keating et al (2015) use H1-H2 as a measure of glottal constriction whereas CPP in the present study measured noise in the signal. Because of this, these findings can be interpreted as suggesting the introduction of noise from another source than the shape of the glottis. These findings offer a very interesting perspective on the proposal by Esling et al (2019) that the entire larynx is an articulator with lots of possibilities of creating phonetic variation using structures beyond the vocal folds such as the ventricular folds and the aryepiglottic constrictor mechanism. In this regard, however, it has not been established whether EGG can detect ventricular fold activity and/or other possible laryngeal mechanisms relevant to laryngealisation, or indeed if they affect the signal at all. Some articulatory mechanisms like larynx raising might obscure the electrode placement, making EGG inappropriate for some types of laryngealisation and further, individuals might employ different articulatory strategies to create contrastive laryngealisation, meaning some speakers might be appropriate for EGG, some might not. Future work is needed to be certain how the EGG signal is affected by non-glottal laryngeal mechanisms if possible to measure. It is likely that this will involve some idiosyncrasy but even quantifying this would be a step forward. Generally, the idiosyncrasy of articulatory strategies in laryngealisation is not particularly well-studied, making it difficult to determine the appropriateness of EGG for all speakers even if research finds that it is affected by non-glottal articulatory gestures.

In summary, the results of this study appear to be highly compatible with the proposals of the LAM framework and warrant more research into methods gauging laryngeal articulatory strategies via non-invasive methods and the correlation between these and what is already known.

Another general source of uncertainty pertained to threshold levels for acoustic measurements. When characterising the stød according to Keating et al's (2015) subtypes it was difficult to determine e.g. how low a low H1-H2 needed to be to indicate glottal constriction. This was an important issue as the subtypes are defined and distinguished by, among other things, their degree of glottal constriction. It is generally assumed, however, that H1-H2 should be negative, but whether there is a perceptual effect of just lowering is uncertain. In a similar vein, in the analysis of the Aarhus speakers there was uncertainty about when f0 could be said to be low for the non-constricted creak subtype. For some subtypes, like prototypical creaky voice, there have been proposals for absolute value thresholds, such as an f0 below 100 Hz (Laver 1980). For less common subtypes such as non-constricted creak, no such values have been suggested and it is unclear what Keating et al's (2015) criterion of a low f0 for this subtype is compared to prototypical creaky voice. A similar issue of what constitutes an absolute value threshold was raised in regard to the articulatory measurements. It was unclear whether an increase in CQ values by 10% or even less is relevant to distinguish phonation types. Or, to put it simply, it is not fully known at what exact threshold modal phonation becomes non-modal. To investigate these thresholds in future work it would be relevant to include perception studies as well as more specifically designed production studies to uncover these aspects. It is also worth considering the degree of language specificity in such thresholds. Researching this would require knowledge about the baseline restrictions imposed by physiology and aerodynamics to tease out the scope of possible differences that cannot be explained by such restrictions as these would naturally be the same for all languages. Further, the finding of different acoustic correlates being significant regardless of CQ value significance suggests that there might even be a phonatory subtype specificity to these thresholds, something research has yet to investigate systematically. This was also supported by findings in Aarhus where the regular stød in one contrast had CQ values rising to about 0.53, but regular stød in another contrast did not have increased CQ values, despite both being labelled as regular stød. This was taken to indicate that the stød in Aarhus might not be just one type in terms of its laryngealisation. In summary, based on the current study, some future work could favourably research relevant thresholds for various measures, particularly combined with perception of contrastive phonation to explore the correlation between the two.

Having provided some broader perspectives on the findings for future research related largely to methodological uncertainties, this section ends with some things left to discover about the Danish stød specifically. An important topic for variation was whether the tonal stød in Aarhus is realised purely through tone or with a degree of laryngealisation. F0 as an acoustic correlate was somewhat inconclusive as it did not significantly distinguish the regular and tonal stød, nor were its mean trajectory declining in the way expected of a tonal realisation. Compared to the regular stød CPP and HNR were higher for the tonal tokens but did have a decline towards syllable termination indicating a degree of increasing noise. This could mean the presence of laryngealisation on both stød types but to different degrees, or that laryngealisation is absent from (i) either the tonal tokens or, (ii) both the regular and tonal tokens. Auditorily, both types in Aarhus gives the impression of being relatively non-laryngealised compared to Copenhagen standard stød realisation, but if both types were articulated with modal phonation, due to their cue being tone, it would not be expected that CPP and HNR are declining compared to non-stød tokens. Inspecting the CQ values, however, these suggested that the tonal stød tokens in Aarhus were articulated with modal phonation. It could be argued that the Aarhus tonal stød is tonal based on higher periodicity compared to the regular stød tokens and the lack of evidence of increase vocal fold contact in the glottal cycles. However, research visualising the larynx more directly would be extremely valuable in supporting this observation. Until such research exists, there will remain a level of doubt in this matter.

8. Conclusions

The aim of this thesis was to explore phonetic variation in phonologically contrastive laryngealisation through the comparative lens of dialectal differences within a language using the LAM as a guiding framework. This was studied uilising the Danish stød as a case study due to previous reports of rich variation. The research was centred around three main research questions:

A) How much gradient phonetic variation in voice quality occurs during the Danish stød, a binary phonological contrast?

B) Which acoustic and articulatory measurements correlate with the stød?

C) How are the voice quality changes timed?

To categorise different phonetic types of laryngealisation, the five subtypes described in Keating et al (2015) were used as a reference, assumed to be appropriate as the stød is often described as a kind of creaky voice and Keating et al's (2015) paper describes exactly this – different kinds of laryngealisations. These are characterised by combinations of the acoustic correlates f0, H1-H2, HNR and SHR which were also used in this study. Some additional correlates were added, namely intensity and CPP, as well as an articulatory correlate via obtaining EGG measurements. The stød was divided into two contrastive pairs, allowing each pair to have a comparative baseline, and the differences between them were analysed using two different statistical methods, GAMMs and random forests.

All acoustic measurements tested were statistically significant in differentiating the non vs regular stød contrast in both dialects expect for the SHR. For the regular vs tonal stød contrast most measures were also statistically significant except for f0 and H1-H2 in both dialects. None of the subtypes in Keating et al (2015) fitted the data perfectly for either dialect but both were found most similar to non-constricted creak. In Copenhagen, however, glottal constriction was indicated by CQ values showing increased vocal fold contact in each cycle for the Copenhagen regular stød reaching 0.60 towards syllable termination, a value higher than the typical modal phonation value of around 0.50. Despite this, the criterion of a low H1-H2 was not fulfilled when comparing the regular to the tonal stød tokens. As such, there were some discrepancies between the acoustic and articulatory measures that further complicated the categorisation. For the Aarhus dialect, however, the CQ values did not indicate much glottal constriction. A question was asked for the Aarhus dialect on whether the tonal stød had evidence of laryngealisation or not as this would be interesting to account for in terms of possible phonetic variation in contrastive phonation. Acoustic measures of f0 did not look as expected for a tonal realisation due to a lack of a steep decline. Measures of increased noise such as HNR and CPP were also declining, indicating laryngealisation to some degree but as mentioned, the CQ values did not confirm the finding of laryngealisation in terms of increased vocal fold contact.

Overall, the acoustic analysis showed that the parameters patterned similarly for the dialects, both in terms of significance in the presence of the stød and also in terms of variable importance for which acoustic variable was the best at predicting this difference. This means that phonetic variation was less in this regard. The highest degree of dialectal variation was found in the timing of the stød. This generally patterned into two groups based on either a full syllabic span or a partial syllabic span. These patterns were related to stød type more than dialect.

As well as dialectal variation, large inter-speaker variation was found for the phenomenon labelled post-stød creak which described how the phonetic effects of the stød can have an increased segmental scope, lasting up to four segments after the relevant stød basis. These effects have not been systematically studied yet, but the initial account for the phenomenon here showed that it was statistically significant for predicting both dialect and stød type. More research into this phenomenon both in Danish and cross-linguistically would be highly relevant to both cross-linguistic typology of non-modal voice quality contrasts and to sociophonetic research due to the high correlation with dialect and also due to the high amount of inter-speaker variation which makes the phenomenon potentially important for forensic applications if large idiosyncrasy in its use can be documented consistently.

The variability in timing and lack of correlation in some acoustic vs articulatory measures suggested that the stød can be articulated with a noise source independent of glottal constriction, meaning other articulatory mechanisms beyond the shape of the glottis are utilised to create contrast. This provides a very interesting perspective supporting the LAM proposed by Esling et al (2019) that the entire larynx is an articulator with lots of possibilities of creating phonetic variation using structures beyond the vocal folds such as the ventricular folds and aryepiglottic structures. Much more research is needed to establish exactly how these structures interplay during contrastive laryngealisation in the world's languages and further, which methods are most appropriate to capture such articulations.

To conclude, the most important findings that this research project has contributed to the existing literature can be summed up as follows:

(i) Investigating dialectal differences and types of contrastive laryngealisation has proven to be a rich resource from which to widen our empirical understanding of phonetic variation in phonological voice quality across different languages and what can be the source of this variation.

(ii) The stød generally patterns much more variably than contrastive laryngealisation in other languages based on its timing. This seems to be related to the fact that the stød has different types of stød basis as its domain despite being classified as a phonological syllabic prosody.

(iii) The stød in Danish is not just one type of stød, and even the standard Copenhagen variant exhibits differences in articulation when analysed as two types rather than one phonetic entity.

(iv) The presence and prevalence of post-segmental phonetic effects of the stød, post-stød creak, is not just a random curiosity, nor the result of predictable co-articulation, but is in fact significant in distinguishing both different dialects and types of stød. Further, it seems to display inter-speaker idiosyncratic behaviour worth investigating more systematically in future research projects.

(v) The findings from this study support the notion that the larynx is an active articulator and that various structures within it can affect voice quality independent of activity in the glottis. This encourages more research into how these different laryngeal structures interplay and in what ways this affects commonly used acoustic and articulatory correlates of these mechanisms.

9. Bibliography

Adank, P., R. Smits & R. van Hout. 2004. A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America* 116(5), pp. 3099–3107.

Aichinger, P., I., B. Roesner, M. Schneider-Stickler, M. Leonhard, D. Denk-Linnert, W. Bigenzahn, A. K. Fuchs, M. Hagmüller & G. Kubin. 2014. Towards Objective Voice Assessment: The Diplophonia Diagram. *Journal of Voice* 31(2), pp. 253.e17–253.e26.

Anastaplo, S. & M. P. Karnell. 1988. Synchronized videostroboscopic and electroglottographic examination of glottal opening. *Journal of the Acoustical Society of America* 83, pp. 1883–1890.

Andersen, H. 1962. Om stødets alder. Sprog og Kultur 16. 103–106.

Andruski, J. E., & M. Ratliff. 2000. Phonation types in production of phonological tone: The case of Green Mong. Journal of the International Phonetic Association 30(1–2), pp. 37-61.

Avelino, H. 2010. Acoustic and Electroglottographic Analyses of Nonpathological, Nonmodal Phonation. *Journal of Voice* 24(3), pp. 270-280.

Awan. S. N., N. Roy, M. E. Jetté, G. S. Meltzner, R.E. Hillman. 2010. *Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V*. Clin Linguist Phonetics 24(9), pp. 742-758.

Awan, S. N., A. R. Krauss & C. T. Herbst. 2015. An Examination of the Relationship Between Electroglottographic Contact Quotient, Electroglottographic Decontacting Phase Profile, and Acoustical Spectral Moments. *Journal of Voice,* vol. 29(5), pp. 519-529.

Baer, T., A. Löfqvist & N.S. McGarr. 1983a. Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques. *Journal of the Acoustical Society of America* 73, pp. 1304–1308.

Baer, T., I. R. Titze & H. Yoshioka. 1983b. Multiple simultaneous measures of vocal fold activity. In A. Bless (ed.) *Vocal Fold Physiology: Contemporary Research and Clinical Issues*. College Hill, San Diego, pp. 229–237.

Basbøll, H. 1988. The Modern Danish Stød and Phonological weight. In Bertinetto, P. M. & Lopocaro, M. (eds.): *Proceedings from the Cortona Phonology Meeting*. Certamen Phonologicum. Torino: Tosenberg & Sellier, pp. 119-152.

Basbøll, H. 2005. *The Phonology of Danish*. Oxford University Press.

Basbøll, H. 2008. Stød, Diachrony and the Non-Stød Model: Jørgen Rischel (1934-2007) in memoriam. NOWELE 54–55. 147–189.

Berke, G. S., D. M. Moore, D. R. Hantke, D. G. Hanson, B. R. Gerratt & F. Burstein. 1987. Laryngeal modeling: Theoretical, in vitro, in vivo. *Laryngoscope* 97, pp. 871–881.

Berry, D. A., H. Herzel, I. R. Titze & K. Krischer. 1994. Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions. *Journal of the Acoustical Society of America* 95, pp. 3595–3604.

Berry, D. A. 2001. Mechanisms of modal and non-modal phonation. *Journal of Phonetics*, vol. 29(4), pp. 431-450.

Blomgren, M., Y. Chen, M. Ng, H. Gilbert. 1998. Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *Journal of the Acoustical Society of America* 103(5 pt 1), pp. 2649–2658.

Boersma, P. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Institute of Phonetic Sciences, University of Amsterdam, *Proceedings* 17, pp. 97-110.

Boersma, P. & D. Weenink. 2021. Praat: doing phonetics by computer [Computer program]. Version 6.1.42, retrieved 17 April 2021 from http://www.praat.org/.

Bradley, D. 1982. *Tonation*. Canberra: Pacific Linguistics.

Breiman, L. 2001. Random forests. *Machine Learning* 45, pp. 5–32.

Brink, L. 2000. *Sound Changes in Late Middle Danish 1350-1500*. Written as a contribution to a German anthology by Oskar Bandle. Available at www.larsbrink.dk (26/01-2021)

Brink, L. 2018. Stød, tryk og tonegang. In *Dansk Sproghistorie: Ord for ord*, vol. 1. Aarhus: Aarhus Universitetsforlag.

Brink, L. & J. Lund. 1975. *Dansk rigsmål: lydudviklingen siden 1840 med særligt henblik på sociolekterne i København*. Copenhagen: Gyldendal.

Brøndum-Nielsen, J. 1950. *Gammeldansk Grammatik i sproghistorisk Fremstilling I*. 2nd ed. Copenhagen: J. H. Schultz.

Bruce, G. 1977. *Swedish word accents in sentence perspective*. [Doctoral Thesis (monograph), General Linguistics]. Liber.

Brunelle, Marc. 2005. *Register in Eastern Cham: Phonological, Phonetic and Sociolinguistic approaches*. Ithaca: Cornell, Ph.D. dissertation.

Brunelle, M. & J. Kirby. 2016. Tone and Phonation in Southeast Asian Languages. *Language and Linguistics Compass* 10(4), pp. 191–207.

Bye, P. subm. *The Evolution of Pitch Accent in North Germanic: Pragmatics, Phonology, Geography, and History*. In Pavel Iosad & Björn Köhnlein (eds.)

Bye, P. 2004. *Evolutionary Typology and Scandinavian Pitch Accent*. Available at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.602.9844 (26/01-2021)

Bye, P. 2011. Mapping Innovations in North Germanic with GIS. In *OSLa: Oslo Studies in Language* 3(2), pp. 5–29.

Childers, D. G., D. M. Hicks, G. P. Moore, L. Eskenazi & A. L. Lalwani. 1990. Electroglottography and vocal fold physiology. *Journal of Speech and Hearing Research* 33, pp. 245–254.

Childers, D. G., & Krishnamurthy, A. K. 1985. A critical review of electroglottography. *CRC Crit. Rev. Biomed. Eng.* 12, pp. 131–161.

d'Alquen, R. & K. Brown. 1992. The origin of Scandinavian Accents I and II. In Irmengard Rauch, Gerald F. Carr & Robert L. Kyes (eds.) *On Germanic Linguistics*. Berlin, New York: De Gruyter Mouton.

Dalton, M. & A. Ní Chasaide. 2005. Tonal Alignment in Irish Dialects. In *Language and Speech* 48(4)., pp. 441–464.

Davidson, L. 2019. The Effects of Pitch, Gender, and Prosodic Context on the Identification of Creaky Voice. *Phonetica* 76, pp. 235–262.

Davidson, L. 2020. The versality of creaky phonation: Segmental, prosodic and sociolinguistic uses in the world's languages. *WIREs Cogn Sci*; 2020: e1547.

Deguchi, S., Y. Kawahara & S. Takahashi. 2011. Cooperative regulation of vocal fold morphology and stress by the cricothyroid and thyroarytenoid muscles. *Journal of Voice* 25(6), pp. 255–263.

Dejonckere, P. 1981. Comparison of two methods of photoglottography in relation to electroglottoraphy. *Folia Phoniatr*. 33, p. 338.

Dejonckere, P. H. & J. Lebacq. 1983. An analysis of the diplophonia phenomenon. *Speech Communication* 2, pp. 47-56.

de Krom, G. 1993. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research* 36, pp. 254–266.

Draxler, C. & K. Jänsch. 2004. SpeechRecorder – a Universal Platform Independent Multi-

Channel Audio Recording Software. Published in *International Conference on Language Resources and Evaluation*, Lisbon, Portugal. Developed for Bavarian Archive for Speech Signals, Institut für Phonetik, Ludwig-Maximilians-UniversitŠt München, Germany, available for download at https://www.bas.uni-muenchen.de/Bas/software/speechrecorder/ (30/12/22)

Edmondson, Jerold A. 2010. The Kháng language of Vietnam in comparison to Ksingmul (Xinh-mun). *A Mosaic of Languages and Cultures: Studies Celebrating the Career of Karl J. Franklin,* ed. by K.A. McElhanon and G. Resink, 138–54. Dallas: SIL.

Edmondson, J. A. and J. H. Esling. 2006. The valves of the throat and their functioning in tone, vocal register, and stress: Laryngoscopic case studies. Phonology 23, 157–191.

Ejskjær, I. 1990. Stød and pitch accents in the Danish dialects. *Acta Linguistica Hafniensia: International Journal of Linguistics*, 22(1), pp. 49-75.

Esling, J. H. 1999. The IPA categories 'pharyngeal' and 'epiglottal': Laryngoscopic observations of pharyngeal articulations and larynx height. *Language and Speech* 42(4), pp. 349–372.

Esling, J. H., J. Clayards, J. A. Edmondson, Q. Fuyuan & J. G. Harris. 1998. Quantification of pharyngeal articulations using measurements from laryngoscopic images. *5th International Conference on Spoken Language Processing*, pp. 3091–3094. Sydney: ASSTA.

Esling, J. H. 2005. There are no back vowels: The Laryngeal Articulator Model. Canadian Journal of Linguistics 50, 13–44.

Esling, J. H, S. R. Moisik, A. Benner & L. Crevier-Buchman. 2019. *Voice Quality: The Laryngeal Articulator Model*. Cambridge Studies in Linguistics. United Kingdom: Cambridge University Press.

Esposito, C. 2003. Santa Ana del Valle Zapotec phonation. Master's thesis, UCLA.

Esposito, C. M. 2006. The effects of linguistic experience on the perception of phonation.

Ph.D. dissertation, UCLA.

Esposito, C. M. 2010. The effects of linguistic experience on the perception of phonation. Journal of Phonetics 38, pp. 306–316.

Esposito, C. M. 2012. An acoustic and electroglottographic study of White Hmong tone and phonation. Journal of Phonetics 40(3), pp. 466-476.

Faaborg-Andersen, K. K. 1957. *Electromyographic investigation of intrinsic laryngeal muscles in humans*. Copenhagen: Möllers Bogtrykkeri.

Fischer-Jørgensen, E. 1989. Phonetic Analysis of the Stød in Standard Danish. *Phonetica* 46(1– 3), pp. 1–59.

Fourcin, A. J. 1974. Laryngographic examination of vocal fold vibration. In B. Wyke (ed.) *Ventilatory and Phonatory Function*. Oxford University Press, London, pp. 315–326.

Gårding, E. & P. Lindblad. 1973. Constancy and variation in Swedish word accent patterns. In *Working papers,* Department of Linguistics Lund University, pp. 36–110.

Garellek, M. 2012. The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics* 40, pp. 152–161.

Garellek, M. 2013. *Production and perception of glottal stops*. PhD dissertation, University of California, Los Angeles.

Garellek, M. & P. Keating. 2011. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. Journal of the International Phonetic Association 41(2), pp. 185-205.

Gerratt, B. R. & J. Kreiman. 2001. Toward a taxonomy of nonmodal phonation. *Journal of Phonetics,* vol 29(4), pp. 365-381.

Gick, B., I. Wilson & D. Derrick. 2013. *Articulatory Phonetics*. Wiley-Blackwell Publishing: United Kingdom.

Goldsthein, Y. 2020. *Towards a natural history of stød*. Master's thesis, Aarhus University.

Grabe, E. 1998. Comparative Intonational Phonology. English and German. PhD thesis submitted at Nijmegen University.

Grønnum, N. 1992. *The Groundworks of Danish Intonation - An Introduction*. Copenhagen: Museum Tusculanum Press.

Grønnum, N. 2005. *Fonetik og Fonologi - Almen og dansk*. Copenhagen: Akademisk Forlag

Grønnum, N. 2007. *Rødgrød med fløde*. Copenhagen: Akademisk Forlag.

Grønnum, N. & H. Basbøll. 2001. Consonant length, stød and morae in Standard Danish. *Phonetica* 58, pp. 230-253.

Grønnum, N. & H. Basbøll. 2007. Danish Stød - Phonological and Cognitive Issues. In M. J. Solé, P. S. Beddor & M. Ohala (eds.) *Experimental Approaches to Phonology* (eds.). Oxford University Press, Oxford. pp. 192-206.

Grønnum, N., M. Vazquez-Larruscaín & H. Basbøll. 2013. Danish Stød: Laryngealization or Tone. *Phonetica* 70(1–2). 66–92.

Hanson, H. M., K. N. Stevens, H.-K. J. Kuo, M. Y. Chen & J. Slifka. 2001. Towards models of phonation. *Journal of Phonetics* 29, pp. 451–480.

Hargus, S. 2016. Deg Xinag word-final glottalized consonants and voice quality. In H. Avelino, M. Coler, & L. Wetzels (eds.) *The phonetics and phonology of laryngeal features in native American languages*, pp. 71–128. Leiden, Netherlands: Brill.

Hampala, V, M. Garcia, J. G. Svec, R. C. Scherer & C. T. Herbst. 2016. Relationship Between the Electroglottographic Signal and Vocal Fold Contact Area. *Journal of Voice*

30(2), pp. 161-171.

Hansen, G. F.. 2015. *Stød og stemmekvalitet: En akustisk-fonetisk undersøgelse af ændringer i stemmekvaliteten i forbindelse med stød*. Copenhagen: University of Copenhagen, Det Humanistiske Fakultet.

Hejná, M., P. Palo & S. Moisik. 2016. Glottal squeaks in VC sequences. *INTERSPEECH*, September 8–12, 2016, San Francisco, USA.

Hejná, M., P. Šturm, L. Tylečková & T. Bořil. 2021. Normophonic Breathiness in Czech and Danish: Are Females Breathier Than Males? *Journal of Voice* 35(3), pp. 498.e1-498.e22.

Heller Murray, E. S., A. Chao & L. Colletti. 2022. A Practical Guide to Calculating Cepstral Peak Prominence in Praat. *Journal of Voice*. 10.1016/j.jvoice.2022.09.002

Henrich N, C. d'Alessandro, B. Doval & M. Castellengo. 2004. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *Journal of the Acoustical Society of America* 115, pp. 1312-1332.

Heman-Ackah, Y. D., R. J. Heuer, D. D. Michael, R. Ostrowski, M. Horman, M. M. Baroody, J. Hillenbrand & R T. Sataloff. 2003. Cepstral Peak Prominence: A more reliable measure of dysphonia. Ann. Otology, Rhinology & Laryngology 112, pp. 324-334.

Herbst, C. T. 2020. Electroglottography – An Update. *Journal of Voice* 34(4), pp. 503-526.

Herbst, C. T. 2021. Performance Evaluation of Subharmonic-to-Harmonic Ratio (SHR) Computation. *Journal of Voice* 35(3), pp. 365-375.

Herbst, C. T., H. K. Schutte, D. L. Bowling & J. G. Svec. 2017. Comparing Chalk With Cheese—The EGG Contact Quotient Is Only a Limited Surrogate of the Closed Quotient. *Journal of Voice* 31(4), pp. 401–409.

Hirose, H. & T. Gay. 1973. Laryngeal control in vocal attack. *Folia phoniatria* 25, pp. 203-213.

Hirose, H., C. Y. Lee & T. Ushijima. 1974. Laryngeal control in Korean stop production. *Journal of Phonetics* 2, pp. 145-152.

Hirano, M. 1974. Morphological structure of the vocal fold and its variations. *Folia Phoniatr*. 26, pp. 89–94.

Hirano, M. 1988. Vocal Mechanisms in Singing: Laryngological and Phoniatric Aspects. *Journal of Voice* 2(1), pp. 51–69.

Hofmans, G. C. J. 1998. *Vortex sound in confined flows.* Unpublished ph.D. thesis, Eindhoven University of Technology, Eindhoven, Netherlands.

Hollien, H. 1974. On vocal registers. *Journal of Phonetics* 2, pp. 125-143.

Hollien, H., P. Moore, R. W. Wendahl & J. F. Michel. 1966. On the nature of vocal fry. *Journal of Speech and Hearing Research*, vol 9, pp. 245–247.

Holmberg, E. B., R. E. Hillman and J. S. Perkell. 1989. Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch. *Journal of Voice* 3, pp. 294-305.

Howe, D. & D. Pulleyblank. 2001. Patterns and timing of glottalisation. *Phonology,* 18(1), pp. 45–80.

Hunter, E. J., I. R. Titze & F. Alipour. 2004. A three-dimensional model of vocal fold abduction/adduction. *Journal of the Acoustical Society of America* 115(4), pp. 1747–1759.

Høysgaard, J. P. 1747. Accentuered og Raisonnered Grammatica. Reprinted in Henrik Bertelsen (ed.), *Gammeldanske Grammatikere 4*. Copenhagen: Gyldendal, 1920, pp. 249-488.

Iosad, P. 2016. Tonal stability and tonogenesis in North Germanic. In Ian Giles, Laura Chapot & Christian Cooijmans (eds.) *Beyond borealism: New perspectives on the North*. London: Norvik Press.

Iseli, M., Y. Shue & A. Alwan. 2006. Age- and Gender-Dependent Analysis of Voice Source Characteristics. *Proc. ICASSP*, Toulouse, May 2006, I-389.

Ito, J. & A. Mester. 1997. *Stødet i dansk*. Handout. Scandinavian Summer School in Generative Phonology, Hvalfjarðarströnd, Iceland, June 16–28.

Jespersen, O. 1897. *Stød og musikalsk accent*. 24 pages.

Johnson, K. 2003. *Acoustic and Auditory Phonetics*. Maldon, MA & Oxford: Blackwell.

Johnson, K. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34, pp. 485-499.

Keating, P., C. M. Esposito, M. Garellek, S. D. Khan, & J. Kuang. 2010. Phonation contrasts across languages. *UCLA Working Papers in Phonetics* 108, pp. 188–202.

Keating, P., M. Garellek & J. Kreiman. 2015. Acoustic properties of different kinds of creaky voice. Conference paper presented at the 18th International Congress of Phonetic Sciences: Glasgow, Scotland, 4 pages.

Kim, M. 2004. Correlation between VOT and F0 in the perception of Korean stops and affricates. *Proceedings of INTERSPEECH 8th*, pp. 49-52.

Kim, M. R., P. S. Beddor & J. Horrocks. 2002. The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics* 30(1), pp. 77-100.

Kiparsky, P. 1995 (In press). Livonian stød. In Wolfgang Kehrein, Björn Köhnlein, Paul Boersma & Marc Oostendorp (eds.): *Segmental Structure and Tone*. De Gruyter Mouton. Available at https://web.stanford.edu/~kiparsky/Papers/livonian.pdf (10/02-21)

Kitzing, P. 1983. Simultaneous photo- and electroglottographic measurements of voice strain. In I. R. Titze & R. C. Scherer (eds.) *Vocal Fold Physiology*. The Denver Center for the Performing Arts, Denver, pp. 221–229.

Kitzing, P., B. Carlborg & A. Loöfqvist. 1982. Aerodynamic and glottographic studies of the laryngeal vibratory cycle. *Folia Phoniatr*. 34, pp. 216–224.

Kock, A. 1885. *Språkhistoriska undersökningar om svensk akcent. Andra delen*. Lund: Gleerup.

Kristensen, M. 1899. *Stødet i dansk.* Arkiv for Nordisk Filologi 15, pp. 41–66.

Kristoffersen, G. 2004. *The development of tonal dialects in the Scandinavian languages*. Talk given at the conference Typology of Tone and Intonation, Cascais, Portugal 1.–3. april 2004. Available at https://bora.uib.no/bora-xmlui/handle/1956/17590 (26/01-2021)

Kroman, E. 1947. *Musikalsk Akcent i Dansk*. Copenhagen: Munksgaard.

Kyst, B. 2004. *Trykgruppens toner i århusiansk regionalsprog*. Unpublished Master's Thesis, Aarhus University.

Kyst, B. 2007. Toner i århusiansk regiolekt. In Torben Arboe (red.): *Nordisk dialektologi og sociolingvistik*. Fællestrykkeriet, Aarhus Universitet, pp. 238-247.

Ladefoged, P. 1996. *Elements of Acoustic Phonetics.* London: The University of Chicago Press.

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

Laver, J. 1980. *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Lecluse, F. L. E., M. P. Brocaar & J. Verschuure. 1975. The electroglottography and its relation to glottal activity. *Folia Phoniatr*. 27, pp. 215– 224.

Liberman, A. 1984. *Germanic Accentology*. Minneapolis: University of Minnesota Press.

Lobanov, B. M. 1971. Classification of russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America* 49(2B), pp. 606–608.

Lowell, S. Y. 2012. *The acoustic assessment of voice in continuous speech*. Perspect Voice Voice Disord 22(2), pp. 57-63.

Malling, A. 1982. Hvor gammelt er det danske stød? *Amsterdamer Beiträge zur älteren Germanistik* 17(1), pp. 79–87.

Mathieson, L. & M. C. L. Greene. 2001. *The voice and its disorders.* London: Whurr Publishing.

Martin, J. 2010. Isthmus Zapotec vowel formants. Work *Papers of the Summer Institute of Linguistics*, University of North Dakota Session: Vol. 50, Article 4.

McGowan, R. S. 1988. An aeroacoustic approach to phonation. *Journal of the Acoustical Society of America* 83(2), pp. 696–704.

Moisik, S. R., J. H. Esling, L. Crevier-Buchman, A. Amelot & P. Halimi. 2015. Multimodal imaging of glottal stop and creaky voice: Evaluating the role of epilaryngeal constriction. *Scottish Consortium* (ed.), paper 247.

Morrison, D. A. 2019. Metrical structure in scottish gaelic: Tonal accent, glottalisation and overlength. *Phonology* 36(3), pp. 391-432.

Nespor, M. & I. Vogel. 1989. On Clashes and Lapses. *Phonology* 6, pp. 69–116.

Oftedal, M. 1952. On the origin of the Scandinavian tone distinction. *Norsk Tidsskrift for Sprogvidenskap* 16, pp. 201–225.

Ogden, R. 2001. Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetics Association* 31, pp. 139-152.

Parsa, V. & D. G. Jamieson. 2001. *Acoustic discrimination of pathological voice*. J Speech, Lang Hear Res 44(2), pp. 327-339.

Patel, R. R., S. N. Awan, J. Barkmeier-Kraemer, M. Courey, D. Deliyski, T. Eadie, D. Paul, J. G. Švec & R. Hillman. 2018. Recommended Protocols for Instrumental Assessment of Voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. *American Journal of Speech and Language Pathology* 27(3), pp. 887-905.

Pedersen, M. F. 1977. Electroglottography compared with synchronized stroboscopy in normal persons. *Folia Phoniatr*. 29, pp. 191–199.

Perridon, H. 2006. On the origin of the Scandinavian word accents. *Zeitschrift für Dialektologie un Linguistikk Beihefte* 131, pp. 91–106.

Pickett, V. B, M. V. Villalobos & S. A. Marlett. 2010. Isthmus (Juchitán) Zapotec. *Journal of the International Phonetic Association* 40(3), pp. 365-372.

Pike, K. L. 1948. *Tone Languages*. Ann Arbor: University of Michigan Press.

R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Redi, L. & S. Shattuck-Hufnagel. 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29, pp. 407-429.

Riad, T. 1998. The Origin of Scandinavian Tone Accents. *Diachronica* 15(1), pp. 63–98.

Riad, T. 2003a. The origin of Danish stød. In Aditi Lahiri (ed.) *Analogy, Levelling, Markedness: Principles of Change in Phonology and Morphology*, pp. 261–301. Berlin: De

Gruyter.

Riad, T. 2003b. Diachrony of the Scandinavian accent typology. In Paula Fikkert & Haike Jacobs (eds.*) Development in Prosodic Systems,* pp. 91–144. Berlin, Boston: De Gruyter.

Riad, T. 2009. Eskilstuna as the tonal key to Danish. In *FONETIK*, 1–6. Dept. of

Linguistics, Stockholm University.

Riber Petersen, P. 1973. An instrumental investigation of the Danish "stød". *ARIPUC 7,* University of Copenhagen, Copenhagen: pp. 195-234.

Rothenberg, M. 1981. Some relations between glottal air flow and vocal fold contact area. *ASHA Rep.* 11, pp. 88–96.

Rothenberg, M. & J. J. Mahshie. 1988. Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech and Hearing Research*, pp. 31, 338–35.

Selbie, S. W., L. Zhang, W. S. Levine & C. L. Ludlow. 1998. Using joint geometry to determine the motion of the cricoarytenoid joint. *Journal of the Acoustical Society of America* 103(2), pp. 1115–1127.

Silva, D. J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23(2), pp. 287-308.

Silverman, D. 1995/1997. *Phasing and recoverability*. New York: Garland Press.

Silverman, K. E. A. & J. B. Pierrehumbert. 1990. The timing of prenuclear high accents in English. In John Kingston & Mary E. Beckman (eds.) *Papers in Laboratory Phonology*, pp. 72–106. 1st edn. Cambridge University Press.

Skautrup, P. 1944. *Det danske sprogs historie: Fra Guldhornene til Jyske Lov*. Vol. 1. Copenhagen: Gyldendalske Boghandel.

Slifka, J. 2006. Some Physiological Correlates to Regular and Irregular Phonation at the End of an Utterance. *Journal of Voice*, vol. 20(2).

Smith, S. 1938. *Zur Physiologie des Stosses*. APhS 12, 33.

Smith, S. 1944. *Bidrag til Løsning af Problemer vedrørende Stødet i dansk Rigssprog: en eksperimental-fonetisk Studie*. Copenhagen: Copenhagen University.

Sóskuthy, M. 2017. *Generalised Additive Mixed Models for dynamic analysis in linguistics: A practical introduction*. arXiv:1703.05339.

Storm, J. 1874. *Om Tonefaldet (Tonelaget) i de skandinaviske Sprog*. Christiania Videnskabs-Selskabs Forhandlinger for 1874, pp. 284–297.

Strobl, C., A. Boulesteix, T. Kneib, T. Augustin & A. Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9, pp. 307-318.

Sweet, H. 1873. On Danish Pronunciation. In *Transactions of the Philological Society*, pp. 94–112.

Tagliamonte, S. A. & R. H. Bayeen. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24, pp. 135-178.

Thorsen, N. G. 1974. Acoustical and perceptual properties of the Danish stød. *ARIPUC* 8, Københavns Universitet, København. pp. 207-213.

Titze, I. R., T. Baer, D. Cooper & R. Scherer. 1984. Automatic extraction of glottographic waveform parameters and regression to acoustic and physiologic variables. In A. J. Bless DM (ed.) *Vocal Fold Physiology: Contemporary Research Clinical Issues*. College Hill, San Diego, pp. 146–154.

van den Berg, J. & T S. Tan. 1959. Results of experiments with human larynxes. *Practical*

Otorhinolaryngologica 21, pp. 425–450.

Verner, K. 1872. *Letter to professor C.W. Smith.*

Verner, K. 1881. Review of Kock 1878. In *Anzeiger für deutches Altertum und deutsche Litteratur 7*, pp. 1–13.

Vogel, A. P, P. Maruff, P. J. Snyder, & J. C. Mundt. 2009. Standardization of pitch-range settings in voice acoustic analysis. *Behaviour Research Methods* 41(2), pp. 318-324.

Watts, C. R. & S. N. Awan. 2015. *An examination of variations in the cepstral spectral index of dysphonia across a single breath group in connected speech*. Journal of Voice 29(1), pp. 26-34.

Wetterlin, A. & A. Lahiri. 2015. The diachronic development of stød and tone accent in North Germanic. In *Historical Linguistics 2013: Selected Papers from the 21st International Conference on Historical Linguistics*, Oslo, 5-9 august 2013, pp. 53–67. Amsterdam: John Benjamins Publishing Company.

Winter, B. 2019. *Statistics for linguists.* Milton: Routledge.

Whitehead, R. L., D. E. Metz & B. H. Whitehead. 1984. Vibratory patterns of the vocal folds during pulse register phonation. *Journal of the Acoustical Society of America*, vol. 75, pp. 1293–1297.

Wood, S. 2007. *Generalized additive models: an introduction with R*. Boca Raton: CRC Press.

Xuan, Y. & Z. Zhang. 2014. Influence of embedded fibers and an epithelium layer on glottal closure pattern in a physical vocal fold model. *Journal of Speech, Language and Hearing Research* 57, pp. 416–425.

Yin, J. & Z. Zhang. 2014. Interaction between the thyroarytenoid and lateral cricoarytenoid muscles in the control of vocal fold adduction and eigenfrequencies. *Journal of Biomechanical Engineering* 136(11), 111006.

Yu, H. J. 2018. Tonal development and voice quality in the stops of Seoul Korean. *Phonetics and Speech Sciences* 10(4), pp. 91-105.

Yuasa, I. P. 2010. Creaky Voice: A new feminine voice quality for young urban-oriented upwardly mobile American women? *American Speech* 85(3), pp. 315-337.

Zhang, Z. 2016. Mechanics of human voice production and control. *Journal of the Acoustical Society of America* 140(4), pp. 2614-2635.

Zhang, Z., L. Mongeau & S. H. Frankel. 2002a. Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes. *Journal of the Acoustical Society of America* 112(4), pp. 1652–1663.

Zhang, C., W. Zhao, S. H. Frankel & L. Mongeau. 2002b. Computational aeroacoustics of phonation, Part II: Effects of flow parameters and ventricular folds. *Journal of the Acoustical Society of America*. 112, pp. 2147–2154.

Zhao, W., C. Zhang, S. Frankel & L. Mongeau. 2002. Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms. *Journal of the Acoustical Society of America* 112, pp. 2134–2146.

10. Appendix

Minimal pair sentences used for elicitation, non-stød vs regular stød:

Expected sentence stress is indicated in underlining for primary stress. Note that actual stress patterns differed from this in some recordings.

Non-stød (disyllables, open and closed)	**Regular stød (disyllables, open and closed)**	**IPA transcription of target word**	**Translation**
Der er piber i butikken		[ˈpʰiːb̥ʌ]	There are pipes in the store
	Vinden piber i sprækkerne	[ˈpʰiːˀb̥ʌ]	The wind whistles in the cracks
Urets viser står på tolv		[ˈviːsʌ]	The hands of the clock are at twelve
	Manden viser mig noget	[ˈviːˀsʌ]	The man is showing me something
Han har været ryger længe		[ˈʁ̥yːʌ]	He has been a smoker for a long time
	Han ryger når han drikker	[ˈʁ̥yːˀʌ]	He smokes when he drinks
Jeg er sulten i dag		[ˈsuld̥n̩]	I am hungry today
	Jeg mærker sulten i maven	[ˈsulˀd̥n̩]	I feel the hunger in my stomach
Jeg åbner slusen for vandet		[ˈsluːsn̩]	I open the sluice for the water
	Jeg fjerner lusen fra håret	[ˈluːˀsn̩]	I remove the louse from the hair
Jeg blev ramt af bygen i dag		[ˈb̥yːn̩]	I was hit by the (rain) shower today
	Jeg gik ud i byen i dag	[ˈb̥yːˀn̩]	I went out into the city today

Jeg fik en god griner med vennerne		[ˈg̊ʁ̥iːnʌ]	I had a good laughing fit with the friends
	Jeg griner med vennerne	[ˈg̊ʁ̥iːˀnʌ]	I laugh with the friends
De så en gyser i biografen		[ˈg̊yːsʌ]	They watched a horror film at the cinema
	Jeg gyser når vinden er kold	[ˈg̊yːˀsʌ]	I shudder when the wind is cold
Der er roer på marken		[ˈʁ̥oːʌ]	There are turnips in the field
	Hun er roer på højt plan	[ˈʁ̥oːˀʌ]	She is a rower on an elite level
Man må mene noget		[ˈme̹ːnə]	One must have an opinion
	Jeg mener det	[ˈme̹ːˀnʌ]	I mean it
Man må stole på ham		[ˈstoːlə]	One must trust in him
	Jeg stoler på ham	[ˈstoːˀlʌ]	I trust in him
Han er løber på højt niveau		[ˈløːb̥ʌ]	He is a runner at an elite level
	Han løber i alt slags vejr	[ˈløːˀb̥ʌ]	He runs in all kinds of weather
Vandet i mosen er brunt		[ˈmoːsn̩]	The water in the swamp is brown
	Jeg spiser mosen med en ske	[ˈmoːˀsn̩]	I am eating the mash with a spoon
Hun kører en tur		[ˈkʰøːʌ]	She is taking a drive
	Der er køer på marken	[ˈkʰøːˀʌ]	There are many cows in the field
Jeg er læser af et dagblad		[ˈlɛːsʌ]	I am a reader of a magazine
	Jeg læser et dagblad	[ˈlɛːˀsʌ]	I am reading a

			magazine
En knæler er et insekt		[ˈknɛːlʌ]	A mantis in an insect
	Jeg knæler på gulvet	[ˈknɛːˀlʌ]	I am kneeling on the floor
Jeg er låner på biblioteket		[ˈlɔ̝ːnʌ]	I am a borrower at the library
	Jeg låner bogen af ham	[ˈlɔ̝ːˀnʌ]	I am borrowing the book off of him
Jeg har en måler til vandtrykket		[ˈmɔ̝ːlʌ]	I have got a measuring device for the water pressure
	Jeg måler vandtrykket	[ˈmɔ̝ːˀlʌ]	I am measuring the water pressure
Hun gav taler til forsamlingen		[ˈtæːlʌ]	She gave speeches to the gathering
	Jeg taler til forsamlingen	[ˈtæːˀlʌ]	I am speaking to the gathering
Der er mange racer i dyreriget		[ˈʁ̞ɑːsʌ]	There are many species in the animal kingdom
	Befolkningen raser over beslutningen	[ˈʁ̞ɑːˀsʌ]	The population is raging over the decision
Det er tanken der tæller		[ˈtɑngŋ̍]	It is the thought that counts
	Jeg ser at tanken er tom	[ˈtɑnˀgŋ̍]	I see that the tank is empty
Jeg har taget den		[ˈtæːð̞]	I have taken it
	Fuglen er på taget, far	[ˈtæːˀð̞]	The bird is on the roof, dad
Der er mange onder at leve med		[ˈɔ̝ːnʌ]	There are many evils to live with
	Det er et under at han	[ˈɔ̝ːˀnʌ]	It is a wonder that he

	lever		is living
Jeg er tjener i restauranten		[ˈtʃɛ̝ːnʌ]	I am a waiter in the restaurant
	Jeg tjener en god løn	[ˈtʃɛ̝ːˀnʌ]	I earn a good salary

Minimal pair sentences, regular stød vs tonal stød:

Regular stød (closed monosyllables, short vowel)	**Tonal stød (closed monosyllables, long vowel)**	**IPA transcription of target word**	**Translation**
Jeg er vild med pasta		[vilˀ]	I am crazy about pasta
	Jeg får et hvil hver eftermiddag	[viːˀl]	I get a rest every afternoon
Med et lyst sind kommer man langt		[se̝nˀ]	With a positive mind one gets far
	En sen nat gik de ud	[se̝ːˀn]	A late night they went out
Jeg har selv sådan en		[sɛlˀ]	I have one of those myself
	Jeg så en sæl i går	[sɛːˀl]	I saw a seal yesterday
Et land er et afgrænset område		[lænˀ]	A country is a limited area
	Der var LAN-fest på netcaféen	[læːˀn]	There was a LAN party at the internet café
Maven er fuld efter middagen		[fulˀ]	The stomach is full after the dinner
	Jeg ser en fugl i haven	[fuːˀl]	I see a bird in the garden
Der er en lund med træer		[lɔ̝nˀ]	There is a grove with trees

	Jeg har et lån i min bank	[lɔ̞ːˀn]	I have a loan in my bank
Et ton er en måleenhed		[tʌnˀ]	A ton is a measuring unit
	Byens tårn er meget højt	[tʌːˀn]	The city's tower is very tall
Der var lam på marken		[lɑmˀ]	There were lamb in the field
	Der var larm fra naboen	[lɑːˀm]	There was noise from the neighbour