LUND UNIVERSITY

Tongue and Jaw Movement Patterns in Heavy Metal Growling

An EMA Case Study

Thea Johansson

Supervisor: Mechtild Tronnier, Malin Svensson Lundmark

Centre for Language and Literature, Lund University
MA in Language and Linguistics, Phonetics
SPVR01 Language and Linguistics: Degree Project – Master's (Two Years) Thesis, 30 credits
January 2025

Abstract

Research into voice qualities that employ supraglottal structures has enlightened our knowledge on the relationship between the laryngeal and the oral articulators. One such insight, which is described within the laryngeal articulator model, is the relationship between the tongue and jaw, and laryngeal constriction. There is evidence to suggest that the tongue dorsum is actively involved in laryngeal constriction by narrowing the vocal tract through retraction. Additionally, it has been observed that laryngeal constriction is correlated with a more open jaw. The current thesis sought to deepen our insights into the relationship between the laryngeal and oral articulator by investigating tongue and jaw movements in modal voice versus heavy metal growling.

Growling in heavy metal necessarily involves supraglottic structures such as the aryepiglottic and ventricular folds. To use these structures, laryngeal constriction must be present. We thus hypothesised that growling would exhibit more tongue retraction and jaw-lowering compared to modal voice. To investigate whether this was true, we employed electromagnetic articulography to track the tongue and jaw movements of two participants. The participants only spoke English as an L2, and only ever growled in their L2. We collected data from their L1s as well (Italian and Greek) but conducted statistical tests on their L2s only. Both participants demonstrated two types of growls.

The results revealed that growl always has a more lowered ($p < 0,05$) and backed ($p < 0.001$) tongue dorsum, as well as a lower ($p < 0.001$) and backed ($p < 0.001$) jaw. We can thus summarise that our statistical tests agreed with our hypotheses. However, through visualisation of the data, we found one outlier in one participant's L1 (Italian) in which the tongue dorsum was less backed in one of the growls compared to modal voice. Additionally, the visualisation of the data might suggest that there indeed are two types of growls, which is a new finding. The results thus generally agree with the predictions we had based on concepts within the laryngeal articulator model but also generated several potential points of research for the future.

Keywords: Voice Quality, Growl, Laryngeal Articulator Model, Laryngeal Constriction, Tongue, Jaw, Metal Music, Extreme Voice Qualities, Nonmodal Voice Qualities

Acknowledgements

Firstly, this thesis would not have been possible without my supervisors *Mechtild Tronnier* and *Malin Svensson Lundmark* who continued to support me throughout this project. I have gained much knowledge and experience thanks to them. I also want to extend many thanks to *Johan Frid* for assisting with the EMA recordings when no one else was available. Had he not agreed to assist with the EMA, it would have been impossible to use it. Thank you also to *Shinichiro Ishihara* for your comments on this thesis. I also want to express that I appreciate all the work that has gone into, and goes into, the *Master's Programme in Languages and Linguistics* at *Lund University*.

I also want to express my gratitude to my participants. They were very willing to communicate about schedule changes that could not be avoided. Without them, this thesis would not exist.

My friends and family deserve acknowledgement for their encouragement and willingness to listen to me talk about this project endlessly. My parents, *Anna Johansson* and *Benny Johansson* deserve my warmest thanks for always encouraging me even when I felt doubt. Here, I also want to thank Lunds Fontänhus, and my contact there.

Finally, I wish to extend a big thank you to myself for never giving up.

Abbreviations

AIC	Akaike Information Criterion
CVT	Complete Vocal Technique
DV	Dependent Variable
EGG	Electroglottography
EMA	Electromagnetic Articulography
EMG	Electromyography
f0	Fundamental Frequency
G	Growl
IV	Independent Variable
JW	Jaw
LAM	Laryngeal Articulator Model
LE	Left Ear
LMM	Linear Mixed-effects Model
M	Modal Voice
MRI	Magnetic Resonance Imaging
NR	Nose Ridge
Psub	Subglottal Pressure
RE	Right Ear
RP	Received Pronunciation
SFT	Source-Filter Theory
TD	Tongue Dorsum
VVFC	Vocal-Ventricular Fold Coupling

Table of Contents

List of Figures

List of Tables

1. Introduction

Metal is a music genre which is associated with several interesting voice qualities. Studying these voice qualities may lead to interesting linguistic implications. In this study, we are observing tongue and jaw movements during modal voice and *growl*, also called *grunt* and *death growl* (see e.g. Aaen et al., 2024; Eckers et al., 2009; Kato & Ito, 2013; Sadolin, 2021). Based on previous research into voice qualities as well as pharyngeal involvement, particularly through the lens of the Laryngeal Articulator Model (LAM) (e.g. Esling et al., 2019), the tongue dorsum (TD) may play an active role in the production of growling, and the jaw may be lowered.

Growling may be considered a type of *nonmodal voice qualities* (further discussed in section 2.1). These voice qualities are relevant to phonetic research to learn about different voice sources, pharyngeal and epiglottal articulation, and about the impact of the laryngeal articulator on the oral articulator, and vice versa. With a good understanding of how voice qualities are produced, we can develop new hypotheses to deepen our understanding and eventually construct precise labels and/or conceptualisations for the different voice qualities produced by the laryngeal articulator. This would make it easier to have discussions within and across the field, potentially in studies of vocal health (speech therapy), and of course in studies into voices for the sake of creative performance. In this study, contributions are made to the field of articulatory phonetics by observing the tongue and jaw movements during modal voice and growl.

1.1 Definitions

Growl is associated with subgenres of metal sometimes referred to as *extreme metal*, such as *death metal* and *black metal* (see Herbst and Mynett, 2023, pp. 36-37). However, it should also be mentioned that *growl* is used to describe harsh-sounding voice qualities in genres other than metal as well, such as jazz, blues, and pop (see Sakakibara et al., 2004). This means that *growl* is a rather broad term. Furthermore, it seems likely that *growl*, when it is specified to be the vocal style used in death metal (Eckers et al., 2009), may be the same as *death growl* (Kato & Ito, 2013), as well as *grunt* in the vocal paradigm Complete Vocal Technique (CVT) (Aaen et al., 2020; Aaen et al., 2024; Sadolin, 2021). This is further elaborated upon in section 2.2. At the very least, *growl* in death metal, *death growl*, and *grunt* appear to be voice qualities associated with metal music and particularly the death metal genre, and thus contrast with growl associated with, for example, jazz. Studies which use a

broader definition of *growl* (e.g. Guzman et al., 2014; Guzman et al., 2019), or which do not mention metal in relation to *growl* (Sakakibara et al., 2004), have varying degrees of relevance for the current study. Throughout the thesis, whenever *growl* is used, we are thus referring to a voice quality associated with metal music specifically, which is investigated in studies on *growl* in death metal, *death growl*, or *grunt* as it is defined in CVT.

Throughout the thesis, the term *voice quality* frequently turns up. There is by no means agreement amongst researchers about the terminology regarding the notion of voice quality; however, we can summarise that *voice quality* is used in a narrow sense and a broad sense (Esling et el., 2019, pp. 1-2; Garellek, 2022, p. 1; Kreiman and Sidtis, 2011, pp. 5-6; Laver 1980, p. 1), which is discussed further in 2.1.3. For now, suffice to say that *voice quality* in the current thesis does not solely refer to phonation at the glottis, but includes other supraglottic voice sources as well. When phonation produced solely by the vocal folds is discussed, that is referred to as *phonatory quality*.

1.2 Current Study

In the current study, we observe the tongue and jaw's movements during the style of growl used in metal music compared to modal voice. In this, we seek to learn more about laryngeal-oral relationships during voice qualities which require laryngeal constriction to set supraglottal structures into movement. The goal is to acquire completely new data about the movements of the oral articulators during a little-studied voice quality and find out how the tongue and jaw may be involved in producing growl.

We employ electromagnetic articulography (EMA) to investigate the tongue and jaw movements during production of growl and modal voice performed by amateur metal vocalists. EMA is capable of recording movements along the vertical and horizontal dimensions of individual sensors which are attached to the participants' relevant articulators. Based on the literature review, no previous studies have set out to examine the tongue's or jaw's movements during growl. As such, this study is explorative in nature and novel. However, the prevalence of certain tongue and jaw movements may be predictable in these voice qualities. Based on research on metal music vocal styles (see e.g. Aaen et al., 2024; Caffier et al., 2018; Eckers et al., 2009; Guzman et al., 2014), and phonetics (see e.g. Baer et al., 1988; Edmondson & Esling, 2006; Esling, 2005; Takano & Honda, 2007), it is possible that the tongue dorsum retracts (moves down and back) during growling, and that the jaw

lowers. As such, in the current study, we are investigating how tongue and jaw movement differ between modal voice and growling.

2. Background

Voice quality has historically been a relatively understudied field. Abercrombie (1967) stated that, in the research on speech production, voice quality was the least investigated, which Laver (1980, p. 1) considered to be a justified statement. Laver (1980) expanded upon Abercrombie's work (Esling et al., 2019; Stuart-Smith, 1999, p. 2553). Laver's (1980) work on voice quality is undoubtedly important, as explicitly stated also by Stuart-Smith (1999, p 2553). Esling et al. (2019) also remark that "[t]he comprehensive history of voice quality presented by Laver (1975, 1979, 1980, 1991) chronicles the earliest origins of the concept of voice quality in phonetic theory." (p. 9). Laver's (1980, pp. 109-132) analysis included several basic types of phonations: modal, falsetto, whisper, creak, harshness, and breathiness. These can be combined into *compound voice qualities* to create even more voice qualities such as *whispery creak* (Laver, 1980, pp. 135-140). Esling et al. (2019), in their description of the laryngeal articulator model (LAM), build upon Laver's (1980) initial system and demonstrate a variety of endoscopic footage of voices produced in isolation, that may or may not originate at the vocal folds. The book by Esling et al. (2019) contains illustrations of the larynx during the production of various voice qualities and might be good to reference for a reader who appreciates images.

Before diving into different views on tongue- and jaw-movement during growl, we must (1) establish the phonetic research that the views are connected to, (2) detangle the vocabularies used to describe growl, and (3) examine the anatomy which we use to vocalize and (4) investigate how it relates to the tongue and jaw. This chapter starts with the first goal by presenting some well-established conceptualisations of, and research on, voice qualities in phonetic research. Following this, we examine the ways that growling is produced and see how it relates to names of growling and adjacent voice qualities. Then, the physiology of the tongue, jaw, and laryngeal tissues which are used to generate sound are presented, as well as their connection with each other. Finally, the connections between the laryngeal and oral articulators are demonstrated through phonetic research in language specific examples.

2.1 Voice Qualities and Frameworks in Phonetic Research

The senses of *voice quality* in phonetics can be linked to different theories of speech production. This section presents views regarding how to best systematize voice qualities from both traditional phonetics following source-filter theory (SFT; Fant, 1960) as well as the laryngeal articulator model (LAM; Edmondson & Esling, 2006; Esling et al., 2019). An

important difference between these is their view of the larynx and the resulting consequences for how researchers of each tradition conceptualise voice qualities. We will see that the narrow sense of voice quality, where voice quality equals phonatory quality, fits better with frameworks that follow SFT, while the LAM employs a broader notion.

2.1.1 Source-Filter Theory and the Breathy-Modal-Creaky Continuum

Source-filter theory (Fant, 1960) is the more traditional model of speech production, described by Fant (1960), while the laryngeal articulator model is younger, described by Edmondson & Esling (2006) and applied on voice qualities by Esling et al. (2019). According to source-filter theory, the larynx has one job, namely, to provide phonation (see Fant, 1960, pp. 15-16, 18-20). The phonation by the larynx generally corresponds to the *source* in source-filter theory, while oral articulation is the *filter* (see Fant, 1960, p. 17). It only generally corresponds to the source since the theory also recognises that the source in, for example, voiceless trills is the place of articulation rather than the vocal folds (Fant, 1960, p. 18).

Voice qualities have been described both categorically and on a continuum (see a discussion on continuous vs categorical categorisation in Gerratt & Kreiman, 2001; compare Gordon & Ladefoged, 2001). The breathy-modal-creaky voice continuum put forth by Ladefoged (1971 pp. 6-22) is a description of voice qualities that can be conceptually generated at the same place – the vocal folds. The continuum describes voice qualities, or in Gordon and Ladefoged's (2001, p. 384) words, *phonation types*, relative to how they differ from modal voice according to the aperture between the arytenoid cartilages (see Gordon & Ladefoged, 2001; Ladefoged, 1971). As such, breathy and creaky voice qualities differ from modal voice in how close the vocal folds are together. Gordon and Ladefoged (2001) summarise the full continuum as: "*[most open] voiceless – breathy – modal – creaky – glottal closure [most closed]* "(p, 384). Defining voice quality according to the aperture between the arytenoid cartilages is a narrow definition of voice quality, as Garellek (2022, p. 1) points out. In the narrow sense, *voice quality* strictly refers to phonation produced by the vocal folds, that is, to phonatory quality or phonatory type (Esling et el., 2019, p. 2; Garellek, 2022, p. 1).

Modal voice is a voice produced with minimal effort that leads to optimal vocal fold vibration. This is considered the most optimal and common voice. The vocal folds are open and closed for about equally as long and the folds themselves are relatively thick (Gick et al., 2013, p. 79). Modal voice has also been referred to as *normal voice* (see e.g. Hollien, 1974, p. 126), but scholars such as Hollien (1974, p. 126) advocate for the term modal voice instead of

normal voice since the term normal voice suggests that other voices are abnormal. Hollien (1974, p. 126) intended the term to cover the fundamental frequencies we normally use when we speak or sing. Laver (1980), in agreement with Hollien's (1974) motivation, uses the term modal voice to describe "[t]he neutral mode of phonation […] where the vibration of the true vocal folds is periodic, efficient, and without audible frication." (p. 94). Linguists have brought up the question if modal voice should be divided into two voice qualities, or registers. That discussion is outside of the scope of the current study. Interested readers can read, for example, Hollien (1974) and Laver (1980, pp. 93-94, 109-111). What we can say is that *register* is a rather vague term (Hollien, 1974, p. 125; Laver, 1980, p. 93), and will be left out of the current study. The term *nonmodal phonation* comes from the idea of voice qualities that differ from modal in some way; however, how exactly they differ, and what modal voice is, is unclear (Gerratt & Kreiman, 2001, pp. 365-366, 377-378).

While many languages utilise a breathy, modal, and creaky voice quality, some scholars find that describing voice qualities on a continuum is problematic (see e.g. Edmonson & Esling, 2006; Gerratt & Kreiman, 2001). When we look at how Gordon and Ladefoged (2001) argue for this continuum, we can see that they mostly, but not solely, appear to argue from the perspective of perception (i.e. which voice qualities, based on our perception and ability to perceive their differences, are employed by languages in some way), and based on vocal fold differences between voice qualities. However, Gordon and Ladefoged (2001) themselves also acknowledge that there are voice qualities which do not overtly fit within their description. One example is the strident voice quality in !Xóõ (Tu language, Botswana) which does not fit in the continuum because it involves the aryepiglottic folds (Gordon & Ladefoged, 2001).

We should also note that while this continuum describes voice qualities relative to the states of the vocal folds, creaky voice can include the ventricular folds as part of its production (Moisik et al., 2015). As such, creaky voice differs from modal voice not just in the aperture between the arytenoid cartilages but also in the addition of some other structure. This suggests that we must consider if creaky voice fits into this continuum. Gerratt and Kreiman (2001) also argue that creaky voice is its own category. Gerratt and Kreiman (2001) hold that speakers' perceptions should be the determining factor in which voice qualities exist and conclude that creaky voice is its own category as speakers perceive creaky voice to be clearly different from modal. In contrast, speakers have a harder time distinguishing between breathy and modal voice, which lead Gerratt and Kreiman (2001) to argue that only breathy and modal voice are a continuum.

In sum, we can see that arguments against the continuum conceptualisation of voice qualities can be made from both a production and perception perspective. This also problematises a narrow view of *voice quality*. An alternative way of describing voice qualities is found in the laryngeal articulator model (LAM).

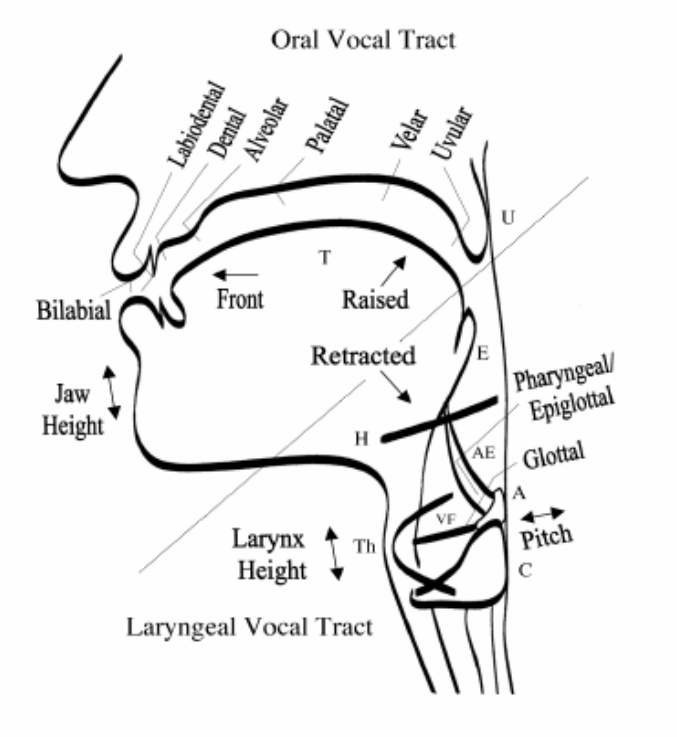
2.1.2 The Laryngeal Articulator Model

The LAM is a newer model of speech production in which the larynx, named the laryngeal articulator, is attributed a bigger role in speech production compared to what it had in the source-filter model (Esling et al., 2019). The LAM includes a two-vocal-tract model (see Figure 1) wherein the laryngeal articulator is the articulator of pharyngeal and epiglottal consonants, as well as the producer of pharyngealisation and epiglottalisation effects, laryngealisation, and many other vocal tract effects (Esling et al., 2019, p. 5). The two vocal tracts here refer to the oral and laryngeal parts of the vocal tract (Esling et al., 2019, pp. 5-6), which is visualised in Figure 1. As we shall see, the LAM also describes laryngeal-oral relationships during speech which neatly allows us to explore and describe how the tongue is connected to, and affects, laryngeal articulation. The view on vowels in the LAM also has interesting connections with laryngeal constriction.

Instead of conceptualising a single place in which all voice sources originate, the LAM describes the laryngeal vocal tract as consisting of six valves, with the first valve at the vocal folds and all others successively above (Edmondson & Esling, 2006, p. 159). Esling et al. (2019, p. 79) express that a conceptualisation of voice qualities as existing on a single plane as suggested by Gordon and Ladefoged (2001) is inaccurate. All voice qualities we may perceive are not created at a single point but rather by a specific set of postures throughout the vocal tract (see Edmondson & Esling, 2006). Here, a brief description of the valves is given, but detailed descriptions are given throughout this chapter when they become relevant.

Figure 1

The "two-vocal-tract" Model



Note: A reconceptualisation of the vocal tract into the "two-vocal-tract" model, employed in the LAM. The image is from Moisik et al. (2007, p. 374).

All six valves interact with each other in complex manners during speech, and there are physiological rules about how they may interact (see Edmonson and Esling, 2006). Valve 1 is at the glottal folds and as such includes adduction and abduction of the folds (Edmondson & Esling, 2006, pp. 159-160). Valve 2 is at the ventricular folds and preforms a constriction of the ventricular folds above the vocal folds while they are closed (Edmondson & Esling, 2006, p. 161). This is called ventricular incursion and may occur in two degrees, namely, partial or complete (Edmondson & Esling, 2006, p. 161). Valve 3 consists of a sphincteric mechanism or, in Edmondson and Esling's (2006, p. 162) words, an aryepiglotto-epiglottal constriction, which they term the *laryngeal sphincter (mechanism)* or the *aryepiglottic sphincteringmechanism*. This sphincteringconstriction may occur while there is tension at valve 1 together with or without 2 (Edmondson & Esling, 2006, p. 162), that is, with or without phonation, and if there is phonation with or without ventricular incursion. Edmondson and Esling (2006, pp. 163-164) have found that valves 1 to 3 are all active as a triple seal during epiglottal stops in several languages. Edmondson and Esling (2006, p. 188) also point out that

valve 1 and 3, at the glottal level and the aryepiglottic constrictor, are particularly potent sources of energy. This does not mean that less efficient voice sources, even non-laryngeal ones, do not exist, such as buccal voice or 'Donald Duck Voice' (see Gick et al., 2013, p. 107).

Valve 4 has two phases in its most extreme form (Edmondson & Esling, 2006, p. 164). It both (1) engages valve 3 and (2) backs two active articulators (the tongue and the epiglottis) towards a passive articulator (the pharyngeal wall) which can result in occlusion, frication, and trilling (Edmondson & Esling, 2006, p. 164). As such, we can see how backing the tongue assists in producing a sound which arises in the larynx (e.g. trilling). Edmondson and Esling (2006, p. 164) also suggest that valve 4 could be said to begin at the epiglottal stop position. In that position, the tongue and epiglottis continue to move, until the occlusive stage of valve 4 is reached where the epiglottis and tongue cover valves 1-3 (Edmondson & Esling, 2006, pp. 164-165). One example of valve 4 in language can be found in Amis (Austronesian) (Edmondson & Esling, 2006, pp. 166, 182-183). In Amis, we find an example of aryepiglottic constriction (valve 3) followed by the occlusive stage of valve 4, where the epiglottis and sides of the tongue are active articulators that create a stop by meeting the pharyngeal wall (Edmondson & Esling, 2006, pp. 182-183).

Next, valve 5 typically raises, but may also lower, the larynx, and assists with both the aryepiglottic sphinctering of valve 3 as well as with tongue retraction (Edmondson & Esling, 2006, p. 166). Finally, valve 6 narrows the pharynx, and is also affected by valve 3 (Edmondson & Esling, 2006, p. 166). While there can be vertical stretching of certain muscles (typically with laryngeal lowering) at this valve, valve 6 refers to the constriction of the laryngeal constrictor muscles around the pharynx, and this constriction typically occurs as an enhancement to otherwise strong constriction in the laryngeal articulator (Edmondson & Esling, 2006, pp. 166-168).

Valves 3, 4, and 5 together correspond to the *laryngeal constrictor* (Edmondson & Esling, 2006, p. 162). In other words, the laryngeal constrictor consists of the laryngeal sphincter at the level of the aryepiglottic folds and the epiglottis (valve 3), tongue backing and epiglottal backing (valve 4), as well as raising of the larynx (valve 5). This may occur with additional narrowing throughout the pharynx (valve 6). The Laryngeal Constrictor Mechanism is interesting for the current study as it describes a synergistic relationship between potential

articulators of growl phonation (discussed further in section 2.2) up to tongue and jaw movement.

The LAM naturally permits a broad definition voice quality which can include voice/noise from places other than the vocal folds, or together with the vocal folds performing modal or nonmodal phonation (see Esling et al., 2019). Garellek's (2022, p. 8) view agrees with this as he also links the LAM to a broad notion of voice quality. The broader sense of voice quality includes someone's long-term vocal characteristics, including non-laryngeal features such as nasality, as well as the variation of phonation within an individual (Abercrombie, 1967, p. 91; Esling et al., 2019, p. 1-2; Garellek, 2022, p. 1). It also includes variation in phonation that happen in the short-term as even short-term variations can convey a long-term impression on the listener (see Esling et al., 2019, p. 1-2; Garellek, 2022, p. 1). Short-term changes here refer to changes in voice quality that may be contrasting within a language (Garellek, 2022, p. 8).

In relation to the notions of narrow and broad voice qualities, we can briefly recall that two or more voice qualities can be combined into compound voice qualities, such as *whispery creak* (Laver, 1980, pp. 135-140), or *harsh whispery creaky voice* (Esling et al., 2019, p. 15). Since the narrow notion of voice quality only refers to phonatory quality, it would be difficult to adequately investigate and explain the mechanisms behind compound voice types while working within a framework that uses it. The LAM, however, with its reconceptualisation of voice qualities as sets of postures, and its broad notion of voice quality, can accommodate such voice qualities.

In sum, we can see that the notion of voice quality can refer to some identifiable voice quality in the time domain either being long or short, and can include both laryngeal features beyond just vocal fold phonation and non-laryngeal features. However, it can also refer to purely vocal fold phonation. The voice qualities, or techniques, of interest in this study at least partially originate in structures other than the vocal folds, such as the ventricular or aryepiglottic folds (see e.g. Aaen et al., 2024; Caffier et al., 2018; Eckers et al., 2009). We thus employ a broad definition of voice quality. Some scholars hold that our perceptual judgement should be the ultimate judge of which voice qualities exist, with little to no consideration of their production (see e.g. Kreiman, 2024, for a discussion about approaches to describing voice qualities). However, in the current study we recognise that the production of voice qualities has an impact on speech in such a way that it is crucial for phoneticians to

recognise them, in line with, for example, Edmondson and Esling (2006). Finally, while the current study is not testing the LAM itself, researchers who are working in this framework are frequently referred to as they produce much research about laryngeal articulation and laryngeal-oral interaction.

2.2 Growling

Growling is particularly related to subgenres of metal referred to as extreme metal (see Herbst & Mynett, 2023, pp. 36-37). However, as we are discussing an artistic expression, growling may of course also occur in other subgenres of metal as well, and the vocal style can also change somewhat within the genre. Herbst and Mynett (2023, pp. 36-37) specifically mention death metal, which emerged in the 1980s, as an example of a genre in which growling or grunting vocals are used. there were different styles of the genre which were associated with certain places, namely Florida (USA, specifically Morrisound Studio in Tampa), Stockholm (Sweden), and Gothenburg (Sweden) (Herbst & Mynett, 2023, pp. 37-39). Florida style became associated with death metal featuring technically demanding structures, sometimes known as "brutal", and included bands like *Death, Morbid Angel, Cannibal Corpse, Deicide, Obituary*, and *Malevolent Creation* (Herbst & Mynett, 2023, p. 37). In Sweden, two styles of death metal emerged, with a Stockholm style producing a less technical raw sound that favoured groove over complexity, and a Gothenburg style which was more melodic with folk-music influences (Herbst & Mynett, 2023, pp. 37-38). An example of the Stockholm style is found in the album *Left Hand Path* by Entombed, and for examples of the Gothenburg style see *Slaughter of the Soul* by At the Gates, *The Gallery* by Dark Tranquillity, and *The Jester Race* by In Flames (Herbst & Mynett, 2023, pp. 37-38).

Another style in the extreme metal category is black metal, which emerged in the 1980's (Herbst & Mynett, 2023, p. 40). In contrast to death metal, black metal favours a low production approach (Herbst & Mynett, 2023, p. 40), or in Hagen's (2023, p. 222) words low-fidelity or unorthodox recording qualities, which may also feature a variety of rough sounding vocals. With the genre's musical choices, it focuses on creating a particular atmosphere (Herbst 2023, p. 222). We should reiterate that there can of course be variation in the vocal techniques used both within a subgenre and outside of it. Growling can be found outside of these genres as well, and there may be other voice qualities present within the subgenres. But this summary should give the reader a general understanding of what growling could sound like.

Moving on, we will consider the definition of *growl* in phonetic and voice research. In research on growl, as with other voice qualities, it is sometimes difficult to discern if one voice quality is described with different names or if they are separate. Generally, the production-perception correspondence between what is perceived as growling and how it is produced is likewise unclear. Perhaps this is partially because *growl* itself seems to be a relatively wide term that may encompass several voice qualities (see Aaen et al., 2020; Aaen et al., 2024; Bailly et al., 2014; Caffier et al., 2018; Eckers et al., 2009; Guzman et al., 2014; Guzman et al., 2019; Kato & Ito, 2013; Sakakibara et al., 2004). Sakakibara et al. (2004) provides the following examples of growl: (1) the jazz, blues and gospel singers using a style similar to singers like Louis Armstrong and Cab Calloway, (2) Brazilian Samba singers, (3) Elza Soares, (4) Bruno and Marrone, (5) Enka singers such as Harumi Miyako, (6) the Xhosa vocal tradition *umngqokolo*, and it is also occasionally present in (7) *Noh* percussionists. Note that metal music is not included in the list. The studies which mention metal music in addition to the examples by Sakakibara et al. (2004) are Caffier et al. (2018), Guzman et al. (2014) and Guzman et al. (2019). Studies which employ this broad notion of voice quality thus have varying amounts of relevance to the subject of growl in metal music specifically.

Growl has also been used in a more specific sense by Eckers et al. (2009) who use *death metal singing* to describe their vocal style of interest. What appears to be the same notion of *growl* is also referred to as *death growl* by Kato and Ito (2013, p. 460). An example vocalist provided by Kato and Ito (2013, p. 460) is Chris Barnes of the band *Cannibal Corpse*, which is a death metal band (Herbst and Mynett, 2023, p. 37). As such we can see that Eckers et al. (2009) and Kato and Ito (2013) relate their voice quality of interest to the genre death metal. Furthermore, some of the studies are also designed according to, and their findings are related to, an already established paradigm of singing/vocalising techniques with a specific set of labels: *Complete Vocal Technique* (CVT) (see Aaen et al., 2020; Aaen et al., 2024; Caffier et al., 2018). According to our understandings of previous studies, it is probable that *death growl* denotes the same, or a very similar, voice quality as *grunt* in the vocal paradigm CVT (see Aaen et al., 2020; Aaen et al., 2024; Caffier et al., 2018; Eckers et al., 2009; Kato & Ito, 2013, p. 460). Examples of grunt given include Dimmu Borgir's song *Progenies of the great apocalypse* and Arch Enemy's *My Apocalypse* (Aaen et al. 2024). Aaen et al. (2024) did not specify which vocalist of Arch Enemy they are referring to. The song *My Apocalypse* was performed by Angela Gossow on the studio album *Doomsday Machine* (Arch Enemy, 2005), but the band's current vocalist Alissa White-Gluz performs it on the live album *As the Stages*

Burn! (Arch Enemy, 2017). Moving on, it should also be noted that, while Caffier et al. (2018) are mentioned here when discussing grunt, Caffier et al. (2018) only referred to authors within the CVT paradigm when they described grunt, and did not explicitly state that they themselves recruited singers who had trained grunt through the CVT paradigm. It is therefore probable that Caffier et al.'s (2018) understanding of grunt follow research on CVT, but it is simultaneously unclear if their recruited singers had trained within CVT.

CVT also includes a voice quality named *growl*, but that denotes a voice quality used in jazz rather than metal (see Aaen et al., 2020). Examples of growl in CVT include Louis Armstrong's *What a wonderful world*, Toni Braxton's *Unbreak my heart*, and Christina Aguilera's *Fighter* (Aaen et al., 2024). Note that Sakakibara et al. (2004) also mentioned jazz artists. Therefore, whenever the studies on voice qualities in CVT are referred to, that is, the studies by Aaen et al. (2020), Aaen et al. (2024), and Caffier et al. (2018), the results of interest are those regarding grunt rather than growl. For more examples of these voice qualities, it is also possible to listen to samples of CVT growl and grunt at Complete Vocal Institute (n.d.).

Finally, grunt is briefly mentioned in some research outside of CVT as well, but not studied (Eckers et al., 2009). This suggests that there may be a grunt vocal style which is perceived as a separate voice quality from growl within the harsh-sounding vocal styles in metal music. However, there seems to be no consensus on any official differences between growl and grunt in either phonetic or vocal research. This is, naturally, excluding CVT where grunt is a defined vocal effect (see e.g. Aaen et al., 2024). While we maintain that grunt in CVT is more interesting than growl, and that grunt is a better description of the style of singing in metal music that we are interested in, future research may reveal further distinctions between similar vocal styles that have not been studied.

2.2.1 How to Produce a Metal-Style Growl

Generally, growling employs supraglottic structures such as the ventricular and aryepiglottic folds (see e.g. Aaen et al., 2020; Aaen et al., 2024; Caffier et al., 2018; Eckers et al., 2009). Eckers et al. (2009) used non-simultaneous electroglottography (EGG) and laryngoscopy to investigate *growl* and found two main forms of growling: ventricular and aryepiglottic. Furthermore, Eckers et al. (2009, p. 1750) found that the vocal folds were oscillating in every production of growling. In other words, all growling was voiced. However, it must be considered that Eckers et al. (2009) had their participants going back and forth between

modal voice and growling on a sustained vowel, which could have encouraged voiced growling rather than voiceless. In any case, the ratio at which the vocal folds vibrated relative to the ventricular or aryepiglottic folds varied between 2:1, 3:1, or 4:1 (Eckers et al., 2009). Eckers et al (2019) did not specify which ratios were relevant for which sets of folds, or which type of growl, but they did point out that Sakakibara et al. (2004) found that the ratio between the vocal and aryepiglottic folds were 2:1. However, as we have seen, Sakakibara et al.'s (2004) notion of growl does not include metal vocal style. Finally, Eckers et al. (2009, pp. 1749-1750) point out that they could not conclude whether there was also ventricular vibration during the aryepiglottic vibration. Thus, we can summarize that growling in metal may be produced with either ventricular or aryepiglottic vibration.

Another study on growl was conducted by Guzman et al. (2019). Guzman et al. (2019) investigated the aerodynamic characteristics of growl and found that growl had a higher glottal airflow rate compared to no growl ($p < 0{,}001$), as well as a higher subglottal pressure (Psub), and less glottal resistance. Based on their results, Guzman et al. (2019) suggested that growl is produced with decreased vocal fold adduction coupled with a high Psub and speculated that a high Psub could be what produces the vibration of the supraglottic structures. Guzman et al. (2019) further suggest that if there had been vocal fold adduction, the necessary glottal airflow would be impossible to reach, meaning that it would be impossible to create the necessary vibration of the supraglottic structures.

Regarding grunt in CVT, it has been shown to consist of vibration of the whole supralaryngeal structure at a low frequency, with some variation between vocalists (Aaen et al., 2020). About half of Aaen et al.'s (2020) participants exhibited large amplitude variations at the vocal folds, and the other half did not exhibit any vocal fold oscillations. Furthermore, according to Aaen et al.'s (2020) findings, grunt may include aryepiglottic and ventricular vibration simultaneously (see Aaen et al. 2020). Additionally, Aaen et al. (2024) found that, in some vocalists, grunt affected the vocal fold oscillation so that it become irregular, which Aaen et al. (2024) attributed to the turbulent airflow necessary to produce grunt. In another study which investigated grunt, amongst some other voice qualities, Caffier et al. (2018, p. 343), who employed laryngoscopy and EGG, concluded that during grunt, the larynx was vibrating in a very open position without any oscillation anywhere. This vibration included the vocal folds, which vibrated decoupled, that is, independently from each other (Caffier et al. 2018, p. 343). In sum, studies on grunt in CVT generally find much vibration of the supraglottic structures, but no oscillation. Note that Guzman et al.'s (2019) suggestion that

vocal fold adduction would make it impossible to reach the necessary airflow agrees with the results with of Caffier et al.'s (2018) study, as Guzman et al. (2019) also point out. We can also add that Aaen et al.'s (2020) results likewise agree.

Although Guzman et al. (2019) speculate that vocal fold oscillation cannot be present during growling, we need to remember that Eckers et al. (2009) did find vocal fold oscillation during both aryepiglottic and ventricular growling. This tells us that while Caffier et al.'s (2018) and Guzman et al.'s (2019) participants may have growled without oscillation anywhere, the specific production of growl which they observed may only be one kind of growl. Alternatively, it could be that Caffier et al.'s (2018) and Guzman et al.'s (2019) participants were most easily physically capable of growling with a more open vocal tract. Additionally, it is possible for the vocal folds to oscillate during aryepiglottic trilling (see Esling et al., 2019, p. 74; Moisik et al., 2010, pp. 1551-1552), and there may be vocal-ventricular fold coupling (VVFC) while aryepiglottic trilling is occurring, as in harsh voice (see Esling et al., 2019, p. 74). We can thus conclude that supralaryngeal structures can be set into movement and significantly contribute to the resulting voice quality even if the vocal folds are oscillating.

In sum, *growl* has a wide definition in research. In research on growl in metal, death growl, and grunt, it is produced with aryepiglottic and/or ventricular fold vibration (Eckers et al., 2009), or vibration of the whole supraglottic structure (Aaen et al., 2020; Caffier et al., 2018; Guzman et al., 2019). Research also suggests that growl requires a high airflow rate and Psub (Caffier et al., 2018). If we recall the narrow/broad notions of voice quality, we can see how growl would require a broader notion of voice quality, and how it is best described within a model such as the LAM. It is also possible that growl could be described as a compound voice quality; however, more research is needed to discern which voice qualities growl would consist of in such a case. Additionally, in the literature, growl does not normally appear to be discussed in this way. This might be because most research into growl is produced by researchers in the creative field rather than the linguistic one. For now, growl is thus best considered its own voice quality. Either way, we can see that supraglottic structures are of crucial importance in our voice quality of interest. Theories like the LAM, which explore speech production throughout all of our speech apparatus, thus provide great insight into voice qualities like growl and harsh voice. With an understanding of the underlying theory of the LAM, we can begin exploring how the laryngeal articulator interacts with the oral articulator.

2.3 The Physiology of the Tongue and the Jaw

2.3.1 The Tongue

The tongue can move in many complex ways. Evidence for tongue movement and articulation comes from both electromyography (EMG) studies (see e.g. Baer et al., 1988 for a study on American English vowels; Honda, 1996 for a model of tongue articulation based on EMG; Smith, 1977), and imaging techniques like x-ray, Magnetic Resonance Imaging (MRI), and ultrasound (see Takano and Honda, 2007 for an MRI study on the Japanese vowels). We will limit ourselves to focusing on the most important parts of the tongue required for understanding this study, which is mainly when the tongue and larynx affect each other in some way, and how the tongue moves up/down and forward/backward. This is because the EMA, which is used to gather data in this study, can provide data on how the sensor glued to the tongue is moving on the vertical and horizontal dimension (i.e. up/down and front/back movement).

Generally, the tongue is moved in two ways: pulling and squeezing. The pulling is easily understood as the tongue is moved by muscles pulling it in certain directions. But the tongue is also like a hydrostat (Gick et al., 2013, p. 167; Takano & Honda, 2007, p. 56), and can be, to borrow Gick et al.'s (2013, p. 167) example, likened to a water balloon: if one part of the tongue is squeezed another part of the tongue gets bigger.

The muscles that control the tongue may be more tongue-moving or tongue-shaping. According to Esling (2005, p. 19), the tongue is moved in different directions by three different extrinsic muscles. The extrinsic muscles also play a part in shaping the tongue (Smith, 1977, p. 10). They are *extrinsic* because they originate from bone structures outside of the body of the tongue and attach to those structures as well as the tongue (Gick et al., 2013, p. 152; Smith, 1977, p. 10). The tongue has *intrinsic* muscles as well, which mostly shape the tongue (Gick et al., 2013 p. 167; see Smith, 1977, p. 11). Because the intrinsic muscles primarily shape the tongue, rather than move it, only the extrinsic muscles are of immediate interest to the current study.

2.3.1.1 The Genioglossus and Styloglossus

According to electromyography (EMG) research, the genioglossus and styloglossus are associated with the front and raised vowels (Baer et al., 1988, pp. 14-15.; Esling, 2005, pp. 19-20). The genioglossus is divided into an anterior, middle, and posterior part (Takano & Honda, 2007, p. 50). It pulls the tongue body forward and is active in the production of

closed, and front, vowels (see Esling, 2005, p. 19; Gick et al., 2013, p. 153; Takano & Honda, 2007, pp. 56-57; Smith, 1977, p. 12). However, a study by Takano and Honda (2007) also suggested that it plays a role in tongue backing as well. Since the tongue behaves as a hydrostat, contraction of the anterior part of the genioglossus would cause bulging in the tongue body towards the back so long as both the medial and posterior parts of the genioglossus are inactive, and thus, it plays a role in back vowels as well by causing tissue to bunch towards the back rather than pull tissue back (Takano & Honda, 2007, p. 56).

The styloglossus is thought to pull the tongue body up and back (Esling, 2005, p. 19; Gick et al., 2013, p. 154), although Gick et al. (2013, pp. 154, 156-157) mention that the newer evidence suggests that the styloglossus functions more as a stabilising muscle, while the movement of the tongue body up and back is cause by the genioglossus and some intrinsic tongue muscles. However, it does seem to contribute to the bunched shape of the tongue body seen in back vowels (Takano & Honda, 2007, p. 57).

2.3.1.2 The Hyoglossus

The hyoglossus is connected vertically into the sides of the tongue from the greater horns of the hyoid bone, and in many speakers, some fibres may run forwards along the lateral backside of the tongue, possibly to the tongue tip (Gick et al., 2013, p. 154; Takano & Honda, 2007). When the hyoglossus contracts, it can move both the hyoid bone and tongue body (Esling, 2005, p. 20). It moves the tongue down and back (i.e. retracts the tongue), and has an antagonistic relationship with the posterior genioglossus (Baer et al., 1988, p. 15; Honda, 1996, p. 43; Takano & Honda, 2007, p. 56). Its posterior part also causes bulging of the tongue dorsum in open back vowels (Perkell, 1996, as cited in Takano & Honda, 2007, p. 56). As such, the hyoglossus is generally important for articulating open-back vowels (Esling, 2005, p. 20; Gick et al., 2013, p. 154; Takano & Honda, 2007, pp. 56-57). Since the hyoglossus is responsible for retracting the tongue, we can see that it plays an active role in voice qualities that employ laryngeal constriction to set supralaryngeal structures into movement. Esling (2005, p. 38) likewise states that the hyoglossus is likely responsible for tongue retraction in laryngeal constriction. We can also note that it might be tricky to produce a vowel which employs the genioglossus, due to its antagonistic relationship with the hyoglossus, if one is producing voice qualities which strongly use the aryepiglottic articulator.

2.3.2 The Jaw

While the jaw can be moved relatively independently of the laryngeal articulator, some research suggests a connection between a higher f0 and jaw movement (Erickson et al., 2017), and there is a tendency for activation, or contraction, of the aryepiglottic constrictor to be correlated with the jaw opening (Esling, 2005, pp. 40-41; Esling et al., 2019, p. 68). This is interesting for growl since growl involves laryngeal constriction, and because it tells us that the movements of the jaw may affect the structures in the larynx. We can see this tendency of the jaw lowering during laryngeal constriction in non-linguistic situations such as swallowing a piece of food (Esling et al., 2019, p. 68). When we swallow, the jaw closes reflexively as food is brought over the tongue, but it may open again once the aryepiglottic constrictor closes (Esling et al., 2019, p. 68). Similarly, while we vomit, it opens reflexively while the aryepiglottic constrictor is kept closed (Esling et al., 2019, p. 68).

The connection between a lowered jaw and laryngeal constriction can also be noticed in the production of some languages. A possible example of this can be seen in the phonological rules of Sephardic Modern Hebrew, wherein, according to Pariente (2015), there are several rules to avoid non-open vowels near pharyngeals. This includes placing a syllable boundary between a pharyngeal and a non-open vowel and adjusting the vowel quality to a more open vowel if a pharyngeal and non-open vowel happen to appear in the same syllable (Pariente, 2015).

Finally, we shall briefly note the mentioned connection between f0 and jaw movement suggested by Erickson et al. (2017). Erickson et al. (2017) investigated jaw movements in contrastive emphasis, in American English, and found that the emphasised syllables had (1) a lower jaw, (2) a more protruded jaw (in 4 out of 6 speakers) as well as (3) a higher f0 (Erickson et al. 2017, pp. 141-142). Erickson et al. (2017, pp. 142-144) tentatively suggest a hypothesis which states that the forward movement of the jaw mechanically counteracts the effects of jaw-lowering and aids in producing the higher f0. In simplified terms, jaw-lowering could mechanically change the state of the cricothyroid joint so that it is positioned in a way that works against the vocal-fold lengthening needed to produce a higher f0 (Erickson et al., 2017, p. 147). By protruding our jaw, we may counteract this effect (Erickson et al., 2017, pp. 147-148). We can also note that jaw-lowering on its own can lower f0 (Erickson et al., 2017, pp. 147-148).

To summarise, unlike the tongue, the jaw does not directly contribute to laryngeal constriction. However, there is a general tendency for the jaw to lower when there is laryngeal constriction present. This tendency can be seen in non-linguistic situations, and possibly in linguistic situations as well. Finally, mandibular movement have been found to have some effect on f0 by affecting the vocal folds. Although it should be clarified that the current study does not investigate the connection between f0 and jaw movements.

2.4 Laryngeal-Oral Interactions in Speech

Recall that, in section 2.2.3, the extrinsic tongue muscles (genioglossus, styloglossus, hyoglossus) may all play a role in producing (closed) back vowels. However, the genioglossus and styloglossus by themselves may not be sufficient to produce these vowels. The hyoglossus, on the other hand, is active in producing open lower back vowels.

Esling (2005) has considered that the tongue's laryngeal connection via the hyoglossus has implications for how we describe vowel production. Where the source-filter model describes vowels as *back*, LAM considers those same vowels to be either *raised* or *retracted* (Esling, 2005, p. 19), citing, for example, Honda's (1996) EMG based model of tongue movement. In Esling's (2005, p. 23) view, the raised vowels are [ɯ u ʊ ɤ o], the retracted vowels [ʌ ɔ ɑ ɒ], and one vowel is either raised or retracted [ɐ]. The raised vowels are mainly produced with the styloglossus (raising the back of the tongue), and the retracted ones with the hyoglossus (retracting the back of the tongue) (Esling 2005, pp. 19-20, 22-23). However, as discussed in section 2.3.1, the genioglossus might be involved in raised vowels as well. For a detailed overview over the muscles of the tongue and their relationships of one another, see Honda (1996) and Baer et al. (1988).

There is an interesting connection between retracted vowels and *harsh voice*. Harsh voice is achieved by constricting the aryepiglottic constrictor mechanism (valve 3), and especially lower epilaryngeal tightening leading to ventricular adduction (valve 2), or alternatively VVFC with enough subglottal pressure that it supersedes creaky phonation (Edmondson & Esling, 2006, p. 162; Esling et al., 2019, p. 67). It also employs valve 1, because the vocal folds themselves also produce aperiodic noise (Edmondson & Esling, 2006, p. 162). The degree of harshness has been noted to increase the more vowels become open and retracted, with some variation depending on consonantal and syllabic environments (Rees, 1958 as cited by Esling et al., 2019, p. 67). This means that when the tongue is retracted, it affects the resulting voice quality in a way that makes it harsher. The tongue also tends to be retracted

during harsh voice, which Esling et al. (2019, p. 68) point out is consistent with the observed relationship between perceived degree of harshness and retracted vowels.

Finally, the relationship between tongue retraction and laryngeal constriction has also been noted in non-vocalic speech sounds which, according to Esling et al. (2019, pp. 8, 28) and Esling (2005, p. 26), are produced with the aryepiglottic constrictor, such as pharyngealisation (Al-Tamimi, 2017; Colarusso, 1985, p. 367; Rose, 1996, p. 74). For example, in Semitic, during pharyngealisation, consonants are realised further back in the vocal tract, so that /k/ is realised as [q] (Colarusso, 1985, p. 367).

2.5 Research Questions and Hypotheses

When we use the laryngeal articulator, particularly if the aryepiglottic constrictor mechanism is very contracted, there is a high tendency for the tongue to retract and for the jaw to lower. Studies on metal vocal techniques reveal that the vocal, ventricular, and aryepiglottic folds are used in growl/death growl/grunt, or all supraglottic structures. Because the aryepiglottic constrictor mechanism likely must contract for aryepiglottic vibration to occur, and because constriction must occur at the level of the ventricular folds as well, it is likely that some degree of tongue retraction and jaw-lowering occurs during the production of these techniques. Furthermore, the tongue might be actively involved in producing growl, as is the case with, for example, aryepiglottic trilling.

Because any voice quality called *growl* in metal music likely uses the same articulatory mechanisms to some degree (the laryngeal constrictor), in the hypotheses H1 and H2, we are assuming that any type of growl would exhibit the predicted changes in tongue and jaw movements. In H1 and H2, retraction and lowering respectively is the downwards movement on the *y*-axis and backwards movement on the *x*-axis. In H3, we are assuming the null hypothesis about the vowels, that this expected difference happens regardless of vowel.

RQ1: How does degree of tongue retraction in vowels differ between modal voice and different types of growl?
H1: Vowels become more retracted during the production of various types of growl compared to modal voice.

RQ2: Are there any differences in jaw movement between modal voice and growl?
H2: The jaw is more lowered during the production of various types of growl compared to modal voice.

H3: The difference in mean tongue dorsum or jaw movement across the open/close dimension and front/back dimension happens regardless of vowel.

3. Method

The study was conducted in three parts: an interview, a simultaneous kinematic and sound recording, and a follow-up interview. The kinematic recordings were performed with an EMA (AG501) at the Lund University Humanities Laboratory. An interview was preferred over other methods of collecting background information as the terminology in the field is somewhat vague. As such, a questionnaire, or something similar, can be confusing for both the participant and the researcher. Interviewing a participant is also preferred as it can make participants feel more engaged with the project. The experiment itself was a repeated measures design wherein one participant repeats one piece of stimuli several times per condition.

3.1 Participants

Participants who perceived themselves to produce growling in any metal genre were welcome to participate. Two participants were recruited via word of mouth. Prior to participation, both participants read and signed a consent form. The interview was recorded and then transcribed. Participant 1 (male, age 32) had Italian as his L1, and participant 2 (male, age 44) had Greek as his L1, and both knew English as an L2. Both participants expressed that they had done all their growling in English, and none in their L1. Neither participant expressed that they were experiencing any current voice issues. In the recording, both participants preformed their most used growl (GA) and a variant growl (GB). Further results from the interviews are found in Appendix B and will be referred to when applicable.

3.2 Interviews

Information collected during the first interview was about (1) the vocal techniques the participants could perform and what they themselves called these techniques, (2) if they were inspired by any particular vocalists, (3) the participants backgrounds in growling as well as singing, (4) the participants own ideas or knowledge about how they perform the techniques, and (5) how the participants preferred to move during growling and if it is difficult to sit down (see Appendix A). (1) , (2), and (3) are important for understanding how to best communicate with the participants about their techniques and for analysing the data. (4) has a similar purpose, but it can also be something to return to during the follow-up interview as a way of giving back to the participants. (5) is important because the participants must sit during the lab recording. If there were any issues with this, it could be addressed during the interview. During the second interview, any additional questions that arose upon listening to

the first interview were asked, as well as if there were any particular stimuli which were tricky to growl. In the first interview, both participants were asked the same questions, but the second interview had some questions unique to each participant because they were clarifying things which had come up in the first interview. The interview questions are found in Appendix A. The participants were also informed about the project in person after the second interview.

3.3 Experiment

The EMA recording was done at Lund University Humanities Laboratory. The EMA at this lab is a Carstens (AG501). It has a sampling rate of 250 Hz/1250 Hz and can record 16 sensors simultaneously (Svensson Lundmark, 2020, p. 49). This means that AG501 can record many positions, in rapid movement, in detail, and generate a lot of data. AG501 has three transmitter coils which emit three electromagnetic fields through which sensors, which are glued to the participant's articulators and face/head, can be detected (Svensson Lundmark, 2020, p. 49). AG501 then calculates the distance between the coils and the sensors (Svensson Lundmark, 2020, p. 49). It is possible to collect data in three dimensions simultaneously. For the purposes of the current study, we collected data on the vertical and horizontal dimensions. Additionally, an external condenser microphone (t.bone EM 9600) was used simultaneously as the EMA recording. The speech signal was used to identify the relevant parts of the EMA signal.

Three reference-sensors were used: one behind each ear (LE = left ear, RE = right ear), and one on the nose ridge (NR). These sensors are used as reference points for the other sensors. This means that the participant can move their head around without disrupting the data collection. For the sensors measuring our points of interest, one sensor was placed on the jaw (JW, in the mouth on the gums just under the teeth, on the outside of the teeth), one on the tongue blade (TB), and one on the tongue dorsum (TD). The TB sensor was placed in case there was time to view it as well, but unfortunately, there was not time to analyse it. When placing a sensor on the TD, to avoid causing discomfort to the participant, the participant was asked to stick out their tongue as far as comfortable, and then make an indent in their tongue with their teeth. The sensor was the put at the indentation, in the middle of the tongue.

3.3.1 Stimuli

The experiment included two kinds of stimuli. The first type of stimuli consisted of several CVC nonce words to be read in modal voice and growl: four in English (Meem, Myym,

Maam, Moom), four in Italian for participant 1, and five in Greek for participant 2. These stimuli were designed to bring forth realisations of vowels close to the vocalic phonemes /iː, ɪ, ɑː, uː/, which are found in the vowel system of received pronunciation (RP) (Giegerich, 1992, pp. 48-49, 51, 100). These phonemes were picked since they are far away from each other in the closed/open front/back dimensions. “myym”, which could potentially be realised as [mɪːm], was added in case the stimulus “meem” [miːm] would be accidentally amusing as it is pronounced identically to the word “meme” (Meme n.d.). According to the Cambridge Dictionary, one definition of “meme” is “An image, video, piece of text, etc., typically humorous in nature, that is copied and spread rapidly by internet users, often with slight variations.” (Meme n.d.).

In addition to stimuli in English, stimuli were also created for the participants’ L1’s in case influences from their L1’s could explain potential differences between the participants. As with the English stimuli, we constructed stimuli that could be realised close to phonemes which are far away from each other in the closed/open front/back dimensions. Standard modern Greek only has five vocalic phonemes (Arvaniti, 1999, p. 169; 2007, p. 118; Lengeris, 2016; Ruge, 1974, p. 3), so we chose to include all of them. Lastly, a bilabial nasal was chosen as consonant because it does not interrupt the airflow and is easy to articulate.

The second kind of stimuli consisted of a single long vowel, described as, for example, ‘aa̰a’ (see Table 1), wherein the participant was instructed to start vocalising in modal voice, switch to growl, then back to modal voice, all while sustaining the vowel. Growl here is signified with the “◌̰“ symbol below the growled section. The stimuli are summarised in Table 1. The “Vowel” columns display which vocalic phoneme each stimulus is associated with. Naturally, we were aware that our participants might realise the English phonemes with slight variation as they were L2 speakers of English. The English vowels in Table 1 are based on the vocalic phonemes of RP (Giegerich, 1992, pp. 48-49, 51, 100). For the Greek vowels we refer to Arvaniti (2007, p. 118) and Ruge (1974, p. 3), and for the Italian vowels, Bertinetto and Loporcaro (2005, pp. 136-137) and Vietti and Mereu (2023). For Greek letters, we refer to Ruge (1974, p. 3).

Table 1: Stimuli and Corresponding Vocalic Phonemes

Vowel	English 1	English 2	Vowel	Italian	Vowel	Greek
/iː/	e̯e̯e	meem	/i/	miim	/i/	μιμ
/ɪ/	yy̯y	myym				
			/e/ or /ɛ/	meem	/ɛ/	μεεμ
/ɑː/	a̯a̯a	maam	/a/	maam	/ɐ/	μααμ
					/o/ or /ɔ/	μοομ
/uː/ or /ʊː/	o̯o̯o	moom	/u/	muum	/u/	μουμ

3.3.2 Procedure

The experiment was expected to take about 45 minutes. The participants were instructed to read all nonce words in modal voice as many times as they were displayed, and then do the same but growling them. In other words, the participants produced the entire set of stimuli in modal voice, and then that same set in GA, and then in GB. The single vowel stimuli were presented as its own set. All data in English was collected first, and then data on the participants' L1' s was collected.

During the EMA recording, the program Praat (Boersma & Weenink, 2024) displayed the stimuli on a screen which the participants were instructed to produce out loud. This was programmed by Dr. Johan Frid and Dr. Susanne Schötz at Lund University Humanities Lab, and was first used in research by Schötz et al. (2013). Each stimulus was displayed for 8 seconds.

The participants were instructed to pick one way to pronounce the vowels and stick to it. Furthermore, they were instructed to produce each voice quality at a pitch that felt comfortable to them. Since certain vowels may affect growling, or may be difficult to produce while growling, we judged that it is important not to instruct the participants too much. If given too detailed instructions, there is a risk that the participants would change their voice quality to accommodate specific vowels, thinking that this is what they are supposed to do. Consequently, during the recording, the participants were not corrected on their pronunciation even if they seemed to deviate to some degree. However, during the experiment for participant 1, the participant himself requested direction on how to pronounce a word on two occasions, once during "maam" and once during "meem". In response, we instructed them to pronounce it like he had previously.

The procedures differed between the participants in some ways. Because participant 1 requested direction, but we did not want the participants to focus too much on doing things as we expected, a practice set was introduced for participant 2 wherein words that contained the target vowel for each nonce word were repeated. Additionally, for participant 1, each stimulus was repeated five times per voice quality, but for participant 2, we collected six repetitions. See Appendix C for further information.

3.3.3 Analysis

The articulography data was analysed with a praat script created by Dr. Johan Frid, researcher at the Lund University Humanities Lab (see the script in Appendix D). The script provided the audio (.wav) in one file and sensor data in another file (.pos). The sensor data described the movement of a specified sensor, along the vertical or horizontal dimensions. In praat, it is possible to retrieve a coordinate numerically so that the *x*-axis represents time and the *y*-axis represents the chosen dimension, which is visualized as a curve (see Figure 2). For example, you can choose to view the tongue dorsum (TD) data of the horizontal movements of the sensor. In the resulting curve, lower values on the *y*-axis corresponds to backwards movement of the sensor over time, while higher values correspond to forward movement (Figure 2).

One of the most important things to decide during the analysis was naturally which points to measure. To get a good view of the tongue's movement during the vowels, five points in time on were measured. These five points will be referred to as A, B, C, D, and E. Point A and E refer to the beginning and end of the vowel segment. Point C is the middle, and points B and D are evenly spaced between the other points. In Figure 2, an example of the segmentation in the spectrogram is shown alongside the EMA data. In addition to these points, the lowest and highest point was measured, which corresponds to the most and least lowered/backed position. As such, the 5 points were measured equally in every vowel, but the other two points vary in position. The modal-growl-modal recordings were similarly annotated so that the first modal vowel had points A, B, C, D, and E, and the middle and last vowel had the same. However, these, of course, shared some points as, for example, point E in the first vowel was the same as point A in the second vowel.

Because the tongue is in constant motion during the vowel segment, it is assumed to be closer to the vowel target in one of the middle measurement landmarks. After data collection and annotation, to decide which data point(s) (A, B, C, D, or E) to perform statistical tests on, points A and E were thus ruled out as interesting. Finally, to decide between B, C, and D, we

decided that C would best represent the vowel, and this was the point we used for running the statistics. We also decided not to use the min/max positions as these varied in position quite a bit, and sometimes we ended up with, for example, multiple maximally backed positions within one vowel. Similarly, in the modal-growl-modal stimuli, we picked point C in each one of the three vowels.

Figure 2

Example of Annotation Set-Up

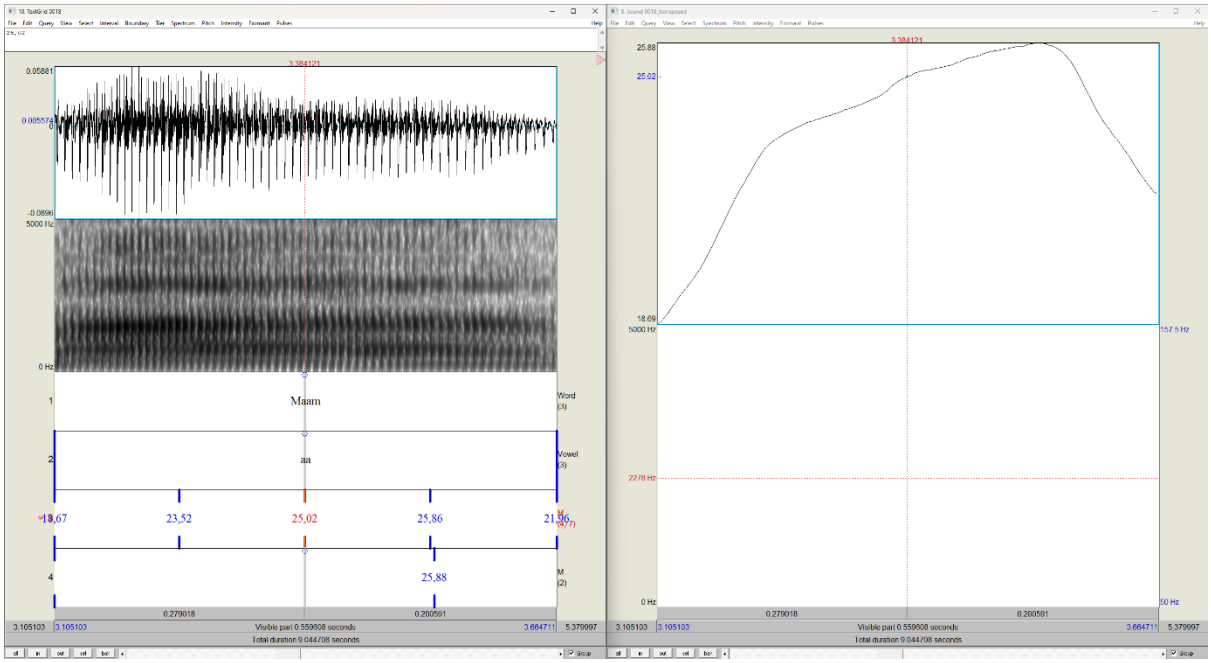


Figure 1: Spectrogram and TextGrid (left) showing the nonce word 'maam' produced in modal voice with an English vowel produced by participant 1, and the corresponding EMA data of the TD-sensor moving forwards/backwards. The x-axis in the EMA window shows the time domain, and the y-axis shows the TD's movement. The curve's upwards movement indicates that the sensor on the TD is moving forward. The position in the centre of the vowel (point C) is selected.

3.3.4 Statistics

In the current thesis, the experiment has four dependent variables (DVs) per RQ (jaw or tongue movement coordinates at the up/down and front/back dimensions). We are further interested in categorical interactions between four independent variables (IVs) that each have multiple levels: *voice quality* (modal, growl), *vowel* (4 vowels), *participant* (1, 2) [1], and *set* (4 different sets). These IVs are all within-subject factors. The variable *set* refers to if the coordinate comes from a nonce-word type stimuli or a "aa̰a"-type stimuli (which is counted as three levels, a.a̰.a).

[1] The participants also represent two different L1's, and since there are only two participants, there is no need for an additional variable 'language'

A Linear Mixed-effects Model (LMM) was used to investigate the interactions between our IVs, and their effect upon our DVs. Mixed-effects Models can take dependences between IV's into account, and can be employed on non-normalised data containing multiple regressions, by including random effects (see Winter 2020, pp. 232-235). Random effects are also separated into random intercepts and random slopes. To briefly explain *random intercepts* and *slopes*, if we were to only test if our coordinates (DV) change between modal voice and growl (fixed effect, IV), then we ignore other things which may significantly impact the results such as variation between the participants. By adding "participant" as a random effect to the model, the data is allowed to vary by participant. If "participant" is just an intercept, this allows the data from each participant to have its own baseline. In Winter's (2020) words: "You can think of this as assigning each participant a deviation score which describes how much that person's intercept deviates from the population intercept." (p. 237). If "participant" is also a slope, then each speaker is allowed to affect the DV in ways that disregard the average slope of the data. For example, our initial test without random effect might indicate that the TD is lower in growl compared to modal voice. If "participant" is included as a slope, then the model is allowed to capture that, for example, participant 1 shows significantly more TD lowering than participant 2, or perhaps that one participant has more TD lowering in modal voice compared to growl.

To briefly demonstrate how the LMM can describe the data, we can consider phonemic variation in the data. It is clear, and expected, that the participants exhibit phonemic variation, which is visible when comparing the standard deviation of vowels when speaker data is combined versus separated (see Appendix E). We dealt with this by telling our LMM to take participant into account. This told us if there is a significant difference between the voice qualities when "meem" is produced by different participants (who are influenced by different L1's, and may be influenced by different English accents).

Statistics were performed in the statistical software R (R Core Team, 2024), and R studio (Posit team 2024). We also decided to get *p*-values with the LMM. To retrieve a *p*-value from the results generated by the `lmer` function in `lme4`, the `dplyr` package was used (Wickham et al. 2023). Firstly, all five points were visualized in all sensor/axis combinations by creating interaction type plots in R. Plots were generated with the `interaction.plot` function included in R. As described in section 3.3.3, the statistical tests were run on data extracted from measuring point C. For the LMM, the package `lme4` was used (Bates et al. 2015).

When determining which LMM model best describes the data, several models of various complexity (i.e. with varying numbers of IVs, intercepts, and slopes) are created. We started with a simple model and then successively added complexity. To find the best model, each new model is compared with the older model. In line with Wieling and Tiede (2017), we compared the models' Akaike Information Criterion (AIC) and accepted the newer model if it had an AIC value that was lower than the previous model by 2. If the AIC value was lower in model 2, we added complexity to model 2 and called it model 3. If model 3 was not better, that is, if model 3 did not show a lower AIC value, the simpler model (model 2) was decided to explain the data best. In R, the function `anova(Model_1,Model_2)` was used to retrieve this number.

4. Results

4.1 Interview results

Participant 1 could produce one main type of growl (GA) and potentially one more (GB), but he was uncertain if GB was truly different from his main growl or just a modification of it. The GB he produced in the current study was a (modified) growl intended to sound lower or deeper than GA. Participant 2, however, stated that he could produce two types of growls. Participant 2 further named his growls Gothenburg style (GA) and Florida style (GB). Both participants expressed that sitting down affected their growling due to how it affected their breath support. They also mentioned that sitting down while growling, to some extent, caused a feeling of being disconnected to their entire body, including legs and feet. They also thought that fronted vowels were the difficult to produce, and thought it was somewhat unclear how to produce “myym”. We will refer to Appendix B when discussing additional information collected during the interviews.

4.2 Tongue Retraction and Jaw-Lowering

While the experiment yielded enough data for our analysis, due to technical issues, some data had to be excluded. The issue caused the timing of the .wav and .pos files to not match in all files. The extreme cases were completely excluded, but cases with very minor de-syncing (less than 10 ms) were included in the analysis. Because of the number of data points we are including, and because we are not looking at the precise timing, but degree of displacement, we judged that the data is still useful. For context, in Figure 2, it could be that the right window with the EMA data should be moved less than 10 ms forwards or backwards in time, while the audio remains where it is. Since the de-syncing is that small, we can still see the relevant information that we need. Once the unusable recordings were removed, including the ones where a sensor had fallen off (Recording participant 1: 8 files; Recording participant 2: 13 files), we had a total of 387 data points (170 for participant 1, 217 for participant 2).

In the statistics, GA and GB were one single category. This is because our hypotheses are about modal voice versus multiple kinds of growl, and because it allows us to get more datapoints per category. However, the data was visualised with GA and GB separated so see if any interesting patterns would appear. This initial visualisation of the data suggests that growl always correlates with a lower and more backed TD as well as a lower JW, in line with

the hypotheses. The effect can be seen in both participants. The LMM found a significant results in all four tests (summarised in Table 2).

Table 2: P-values and the Significance of TD and Jaw movement up/down and front/back between Modal Voice and Growl

LMM results	**Tongue Dorsum**	**Jaw**
Up and Down movement	*p*-value 0.0283*	*p*-value 2.11e-05 ***
Front and Back movement	*p*-value <2e-16 ***	*p*-value < 2e-16 ***

4.2.1 Linear Mixed Model Results

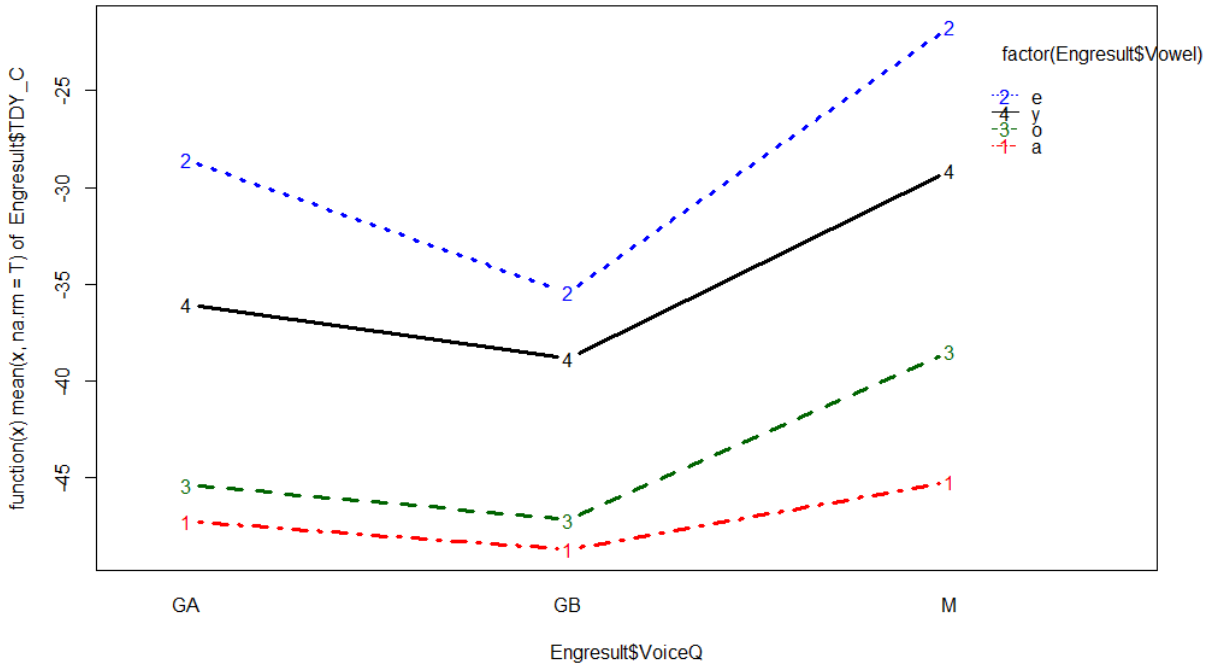
4.2.1.1 Up/Down and Front/Back movement of the Tongue Dorsum

In Figure 3, we can suspect that the TD is lower in both types of growls compared to modal voice. The LMM supports this conclusion ($p < 0.05$). The model which best described the data from the TD sensor moving up or down, at time point C, includes all our random effects, indicating that they all affect the data in some way. These effects were both intercepts and slopes between the IV "voice quality" and the IV's "participant", and "vowel". However, it only includes the IV "set" as an intercept. This indicates that participants and vowels affect the relationship between our voice qualities and coordinates so that participants 1 or 2 could exhibit a stronger pattern than the other. The general difference between our participants can be seen in the visualisation of our data, where participant 1's TD appears to be slightly less affected by GA and GB compared to participant 2 (see Figure 7). Similarly, different vowels could exhibit different levels of strength. In Figure 3, where data from both speakers is combined, "a" /ɑː/ appears to be relatively equal between GA and M, but M still has the highest TD position. This contrasts with "e" /iː/, which appears to be affected by GA and GB more strongly.

When we look at the front/back movement for the TD, visualised in Figure 4, we also get a highly significant result ($p < 0.001$). The best model, like the one for the up/down dimension, included all random effects, but as intercepts only. This indicates that the difference between modal voice and growl was not significantly influenced by either "participant" or "vowel" in this dimension.

Figure 3

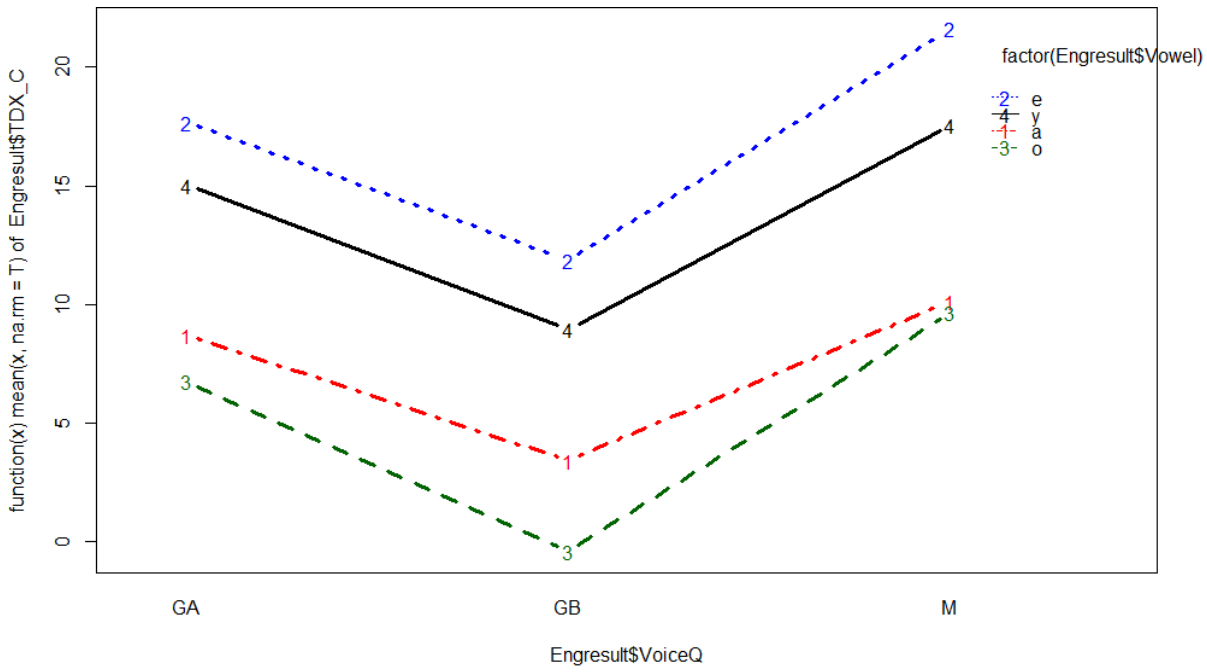
TD Coordinates in the Up/Down Dimension



Mean coordinates at time point C for the TD moving up/down, in each voice quality. The vowels are separated. Data from both speakers is combined. Lower values represent a lower TD position.

Figure 4

TD Coordinates in the Front/Back Dimension



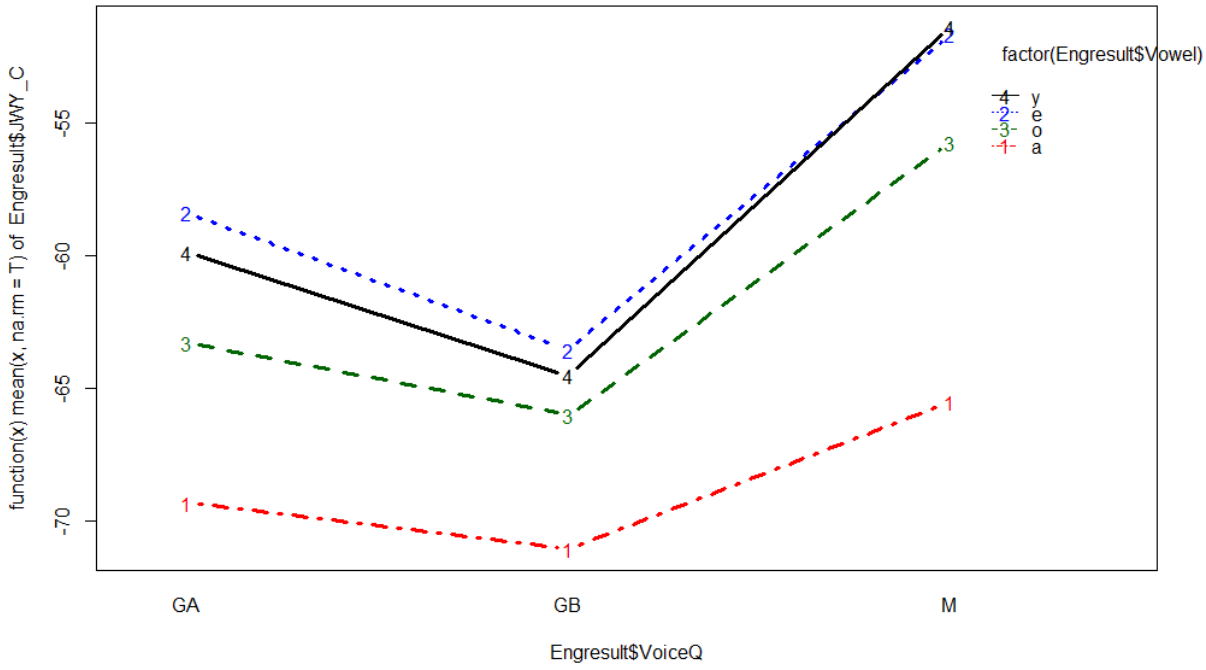
Mean coordinates at time point C for the TD moving front/back, in each voice quality. The vowels are separated. Data from both speakers is combined. Lower values represent a more backed TD position.

4.2.1.2 Up/Down and Front/Back movement of the Jaw

When we look at the results for the jaw (Figures 5 and 6), the LMM shows a significant result in both up/down and front/back movement. In both dimensions, we got highly significant *p*-values: up/down ($p < 0.001$), and front/back ($p < 0.001$). Both models included all random effects as intercepts. Only one random effect was also a slope, namely, "vowel" in the best model for the up/down dimension. In Figure 5 below, we can see that, once again, "a" /ɑː/ appears to be the least effected by the growls, while the closed vowels appear to display a larger difference.

Figure 5

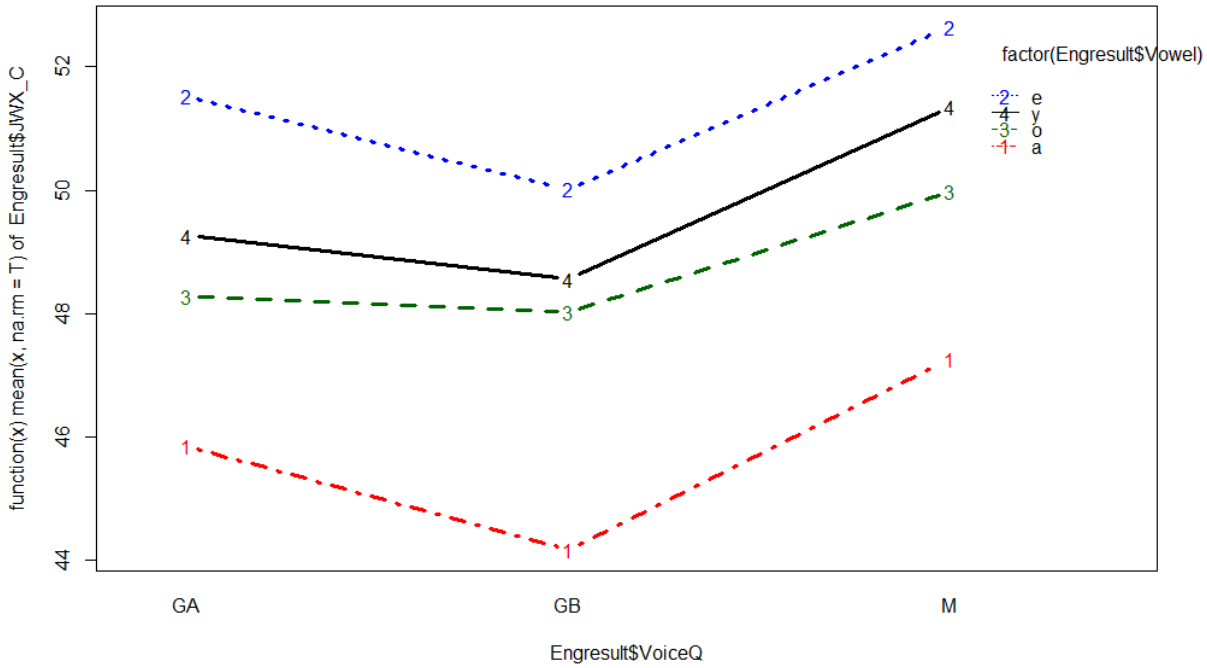
Jaw Coordinates in the Up/Down Dimension



Mean coordinates at time point C for the Jaw moving up/down, in each voice quality. The vowels are separated. Data from both speakers is combined. Lower values represent a lower Jaw position.

Figure 6

Jaw Coordinates in the Front/Back Dimension



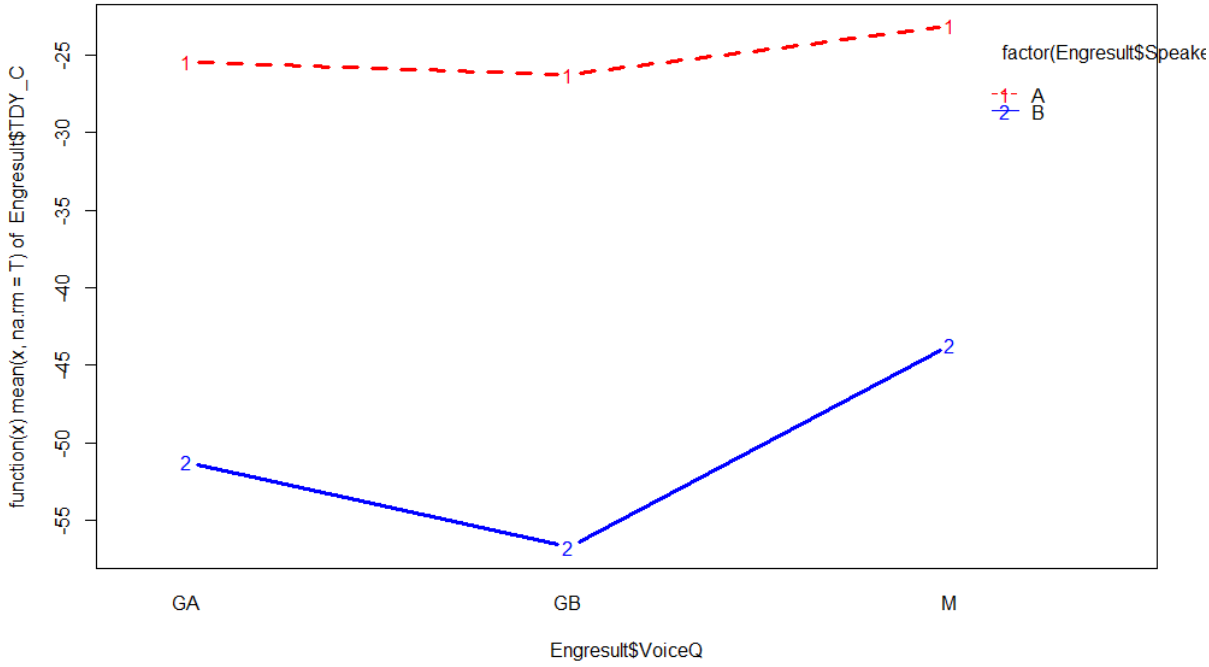
Mean coordinates at time point C for the Jaw moving front/back, in each voice quality. The vowels are separated. Data from both speakers is combined. Lower values represent a more backed Jaw position.

4.2.2 Differences Between Growls

Visualisation of the data revealed differences between the participants, as well as differences between GA and GB. When comparing the overall coordinates for the TD moving up/down in the participants in Figure 7, wherein all vowels are collapsed, we can see that the participants differ in the overall space in which they are moving their articulators, perhaps due to, for example, the size of their articulators. Here, we can also see that participant 2 makes a clearer distinction between GA and GB than participant 1, and that there is, overall, more TD lowering between both GA/GB and M. This is also supported by the fact that our model for the TD in the up/down dimension found some difference between participants, although the model, of course, does not state exactly what the difference is. Note that, as mentioned, GA and GB were one single category in our test, and so cannot support the difference between GA and GB, only the difference between the participants.

Figure 7

TD Coordinates in the Up/Down Dimension with all Vowels Collapsed, Comparing Participants

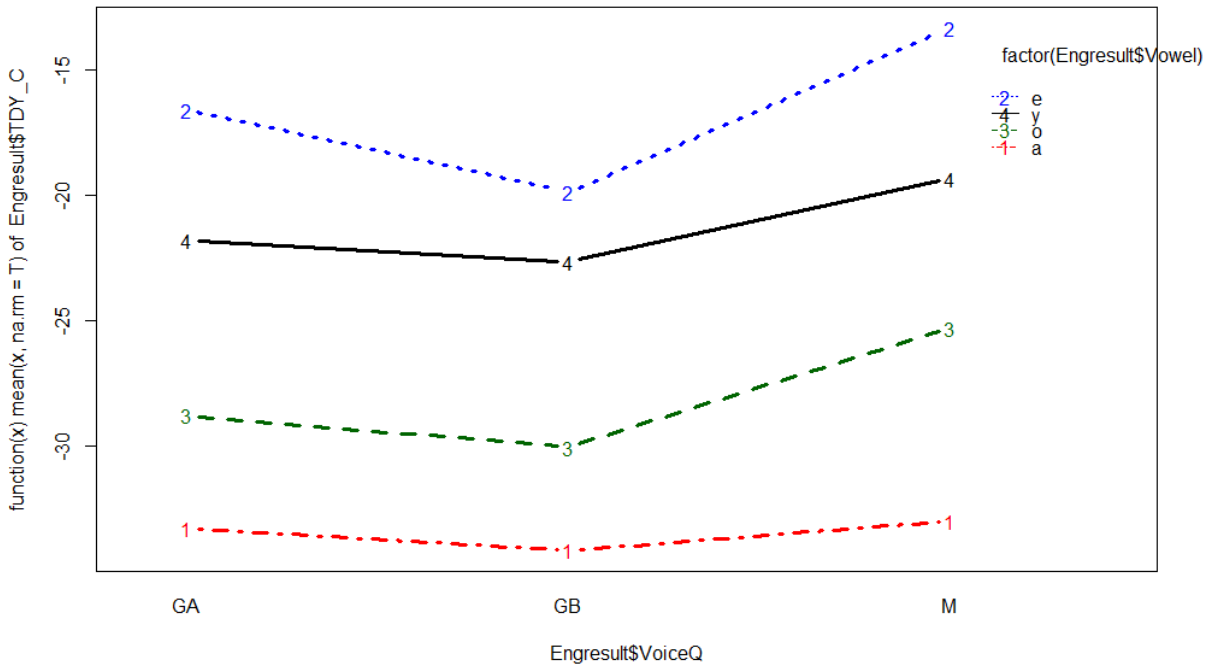


Mean coordinate at time points C in every vowel collapsed, in each voice quality, at the TD. Participant data is separated. A (red line) is participant 1, and B (blue line) is participant B. Lower values represent a lower TD position.

When we look at the participants separately (Figures 8, 9), the difference between GA and GB generally persists, particularly for participant 2 – with one outlier "o" /uː, ʊː/ (Figure 9). The difference between GA and GB does not appear to be as strong at the jaw, particularly not in the participants' L1s (see Figures 5, 6, and Appendix F). In participant 2, there are two clear outliers at the jaw, in both dimensions, namely, the Greek vowels "ε" /ɛ/ and "α" /ɐ/ (Figures F6, F7).

Figure 8

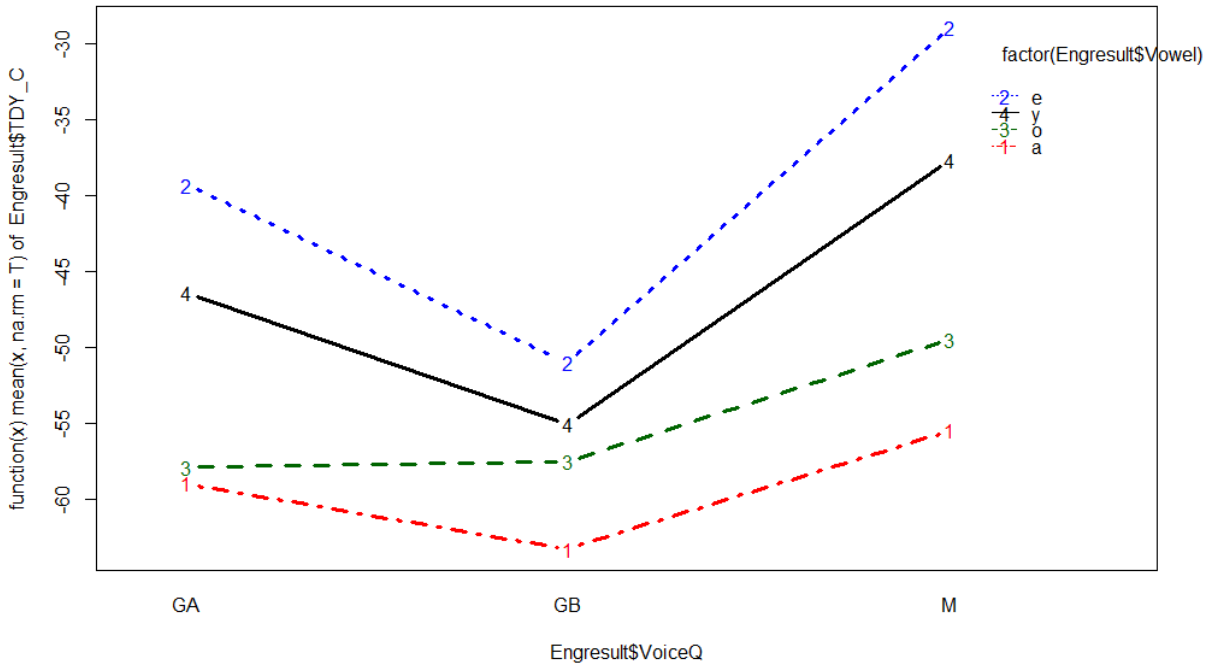
TD Coordinates in the Up/Down Dimension from Participant 1



Mean coordinate at time point C in every vowel separated, in each voice quality, at the TD. Data is from participant 1. Lower values represent a lower TD position.

Figure 9

TD Coordinates in the Up/Down Dimension from Participant 2



Mean coordinate at time point C in every vowel separated, in each voice quality, at the TD. Data is from participant 2. Lower values represent a lower TD position.

4.3 Outliers

Broadly, when we separate the participants and investigate the patterns for each vowel, the same pattern is observed, namely, that the TD is more retracted, and the jaw is more lowered and backed in growl compared to modal voice. However, there are some other patterns present which may be interesting to note. Firstly, as mentioned in section 4.2.2, in all dimensions, there appears to be a tendency for vowels to be more retracted, and exhibit more jaw-lowering, in GB compared to GA, and M is always the least retracted or open. However, occasionally GA and GB appear to be more equal as we can see with “o” /uː, ʊː/ and “a” /ɑː/ in Figure 3 regarding the TD, and “o” /uː, ʊː/ in Figure 6 regarding the jaw. In some instances, when the participants are separated, some vowels appear to contradict this pattern more strongly. For example, in section 4.2.2, we noted that, in “o” /uː, ʊː/, the TD of participant 2 appears to be less lowered in GB compared to GA (Figure 9). Other than this example, the TD appears to have little outliers, but when we look at the jaw, there is more variation. Participant 1, for example, has an opposite pattern with “o” /uː, ʊː/ at the jaw wherein GB is less backed than GA (see Figure 10). Participant 1 also exhibits this pattern in the Italian “a” /a/ vowel, and (more subtly) in the “u” /u/ vowel (Figure F3).

Figure 10 might also indicate another interesting thing that participant 1 did with the jaw moving front/back in English. Here we can see that the jaw appears to be backed a similar amount in all vowels except “a” /ɑː/, where it is instead much more backed. Because of how the jaw moves, if the jaw is lowered, the sensor on the jaw should also show that it is moving back. As we saw in Figure 4, and confirmed with our LMM, the jaw is significantly more lowered during growl, and simultaneously backed. In Figure F1, which visualises data from participant 1 only, we can further confirm that the jaw is lower in GA/GB compared to M. But based on Figure 10, we could tentatively speculate that participant 1 exhibits some jaw protrusion in all vowels but “a” /ɑː/. The same pattern cannot be seen in participant 2. Note also that Figure 10 is visualising English vowels, which were tested statistically. This means that the patterns we see in Figure 10 are significant.

Finally, there is one especially clear outlier, namely, the TD of the Italian “u” /u/ vowel moving front/back (see Figure 11). In fact, it is clearly more fronted in GB than in M. No other Italian vowel, or other vowel in this data set, exhibits such a pattern. In other words, in all other outliers, M is still the least retracted or lowered voice quality.

Figure 10

Jaw Coordinates of English Vowels in the Front/Back Dimension Produced by Participant 1

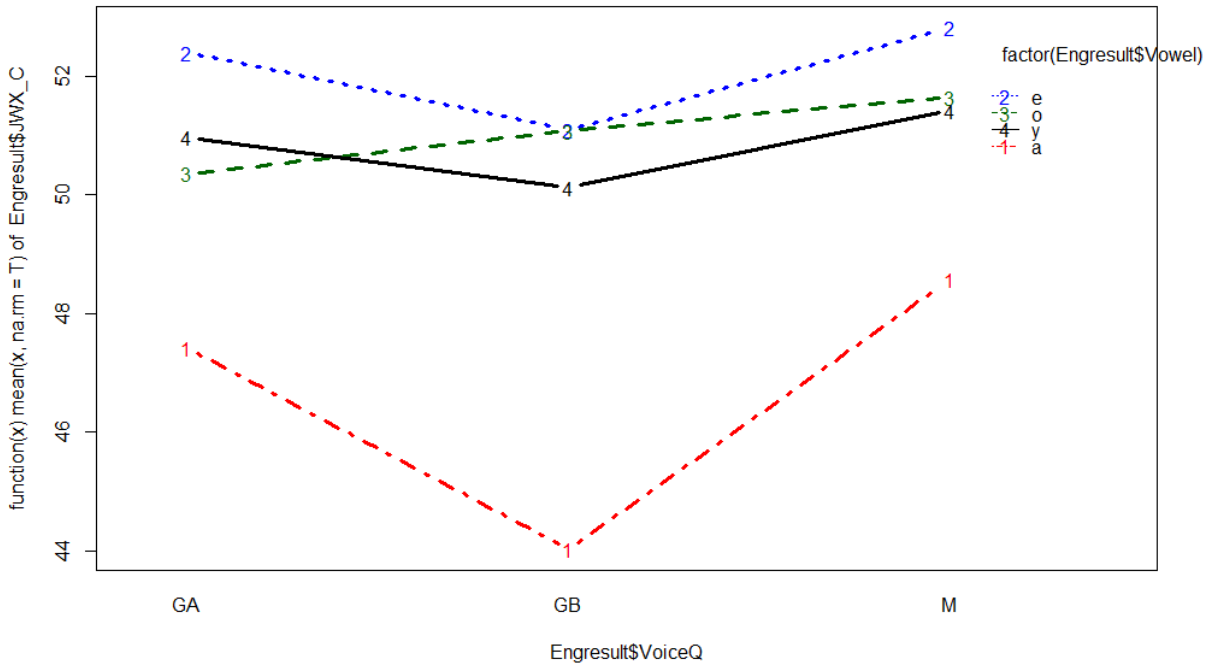
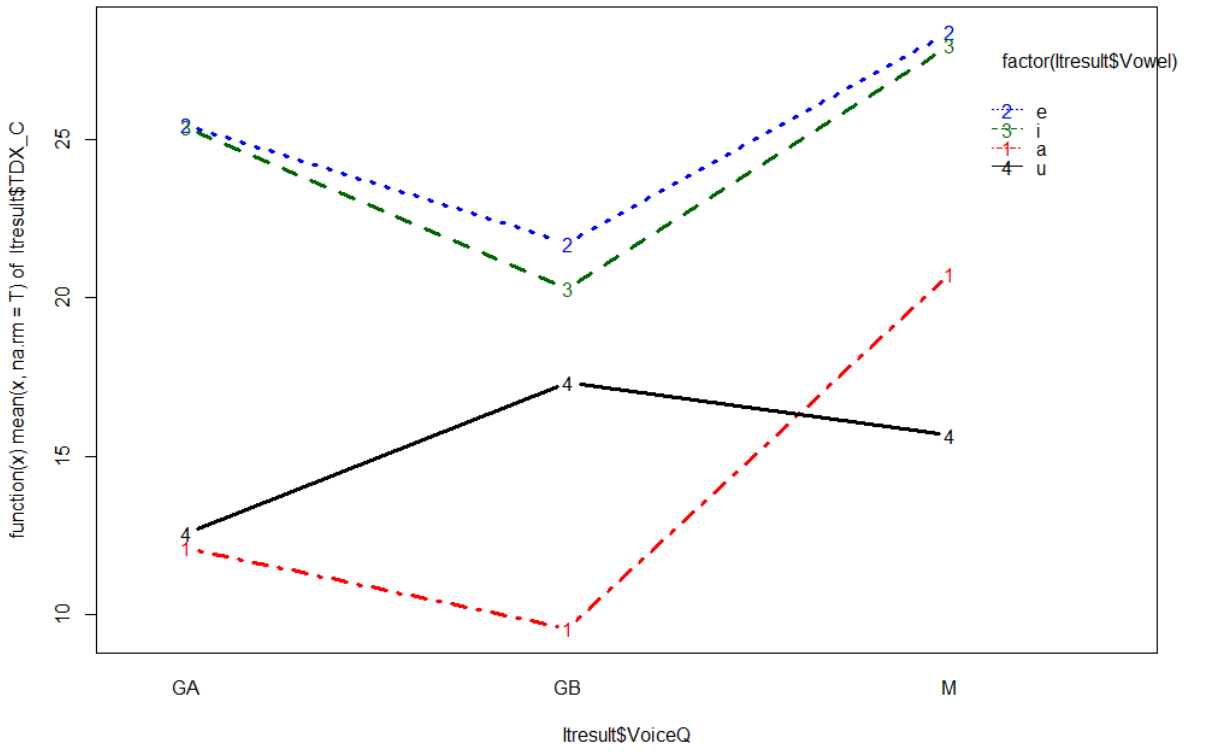


Figure 2: Mean coordinates at time point C for the Jaw moving front/back, in every vowel separated, in each voice quality. Data is from participant 1, and English vowels. Lower values represent a more backed Jaw position.

Figure 11

TD Coordinates of Italian Vowels in the Front/Back Dimension



Mean coordinates at time point C for the TD moving front/back, in every vowel separated, in each voice quality. Data is from participant 1, and Italian vowels. Lower values represent a more backed TD position.

5. Discussion

5.1 Laryngeal-Oral Interaction in Growling

5.1.1 Tongue Movement

In *growl*, the tongue was predicted to move both down and back with the hyoglossus muscle to assist in laryngeal constriction. In section 2.4, we noted that something similar may happen in harsh voice (Rees, 1958 as cited by Esling et al., 2019, p. 67). In the LAM, the tongue moving down and back to assist in laryngeal constriction is valve 4 (Edmondson & Esling, 2006, p. 164). The tongue's movement back and down found in the current study indicates that the TD actively contributes to the desired voice quality by retracting, and as such agrees with the description of valve 4 in the LAM. We can conclude that H1, "Vowels become more retracted during the production of various types of growl compared to modal voice", is fully supported in these speakers.

The fact that the tongue's movement in both dimensions is significant between modal voice and growling supports a broader view of *voice quality*, discussed in section 2.1. It is possible that the growls produced by the current study's participants is only possible to produce if the TD is significantly lowered and backed. This agrees with the observation that harsh voice is perceived as harsher the more backed and lowered vowels are (Rees, 1958 as cited by Esling et al., 2019, p. 67), mentioned in section 2.4. Future research could investigate perceived harshness in growling relative to vowel. Another potentially interesting project would be to do the opposite of the current study and clearly instruct the participants to pronounce a specific vowel as well as they can and observe the effects upon their growling.

5.1.2 Jaw Movement

In the Figures 5 and 6, we can see that the jaw seemingly exhibits some variation between all three voice qualities (M, GA, GB), but was always the leased lowered and backed in M. This was strongly supported by our statistical tests. As mentioned, it has been observed that laryngeal constriction correlates with jaw-lowering in non-linguistic situations such as swallowing, as well as in the speech production of certain languages (see Esling, 2005, p. 40; Pariente, 2015). The results of the current study strongly agree with this tendency. We can conclude that the results show a correlation between heavy metal growling, which is produced with laryngeal constriction, and jaw-lowering. Therefore, H2, "The jaw is more lowered during the production of various types of growl compared to modal voice" is

supported. Because the results for the up/down movement were highly significant ($p < 0.001$), it is not surprising that the results for the front/back movement also were highly significant ($p < 0.001$), since the jaw moves back as it lowers.

As we suggested in section 2.3.2, the jaw has a weaker connection to laryngeal constriction compared to the tongue since it is operating more independently of the laryngeal constrictor. If jaw-lowering is not necessary for producing the laryngeal constriction present in growling, then the jaw could be allowed to move more freely. However, since the results for the current study were so clear, it is possible that the jaw may also influence the growl in some important way. Here, we would like to point out that there may be another mechanical connection between jaw-lowering and the participants' GB - although we have not investigated this specifically, of course. In Appendix F, we can see that the jaw generally appears to be equally lowered or more lowered in GB compared to GA in both participants, although more clearly in participant 2 and with some outliers. Both participants had also expressed that their GB was either lower or "more base-heavy" (Appendix B). Additionally, as mentioned in section 2.3.2, there may be a mechanical synergy between lower f0 and a lower jaw, so that jaw-lowering might lower the f0 (Erickson et al., 2017, pp. 147-148). Since the participants were producing a growl that they perceived as lower, the notable jaw-lowering could have been further influenced by the relationship between the jaw and vocal folds, or simply a tendency to lower one's jaw when producing lower notes (whether necessary or not). Here, it must also be noted that both participants expressed that growling was done without vocal fold phonation (Appendix B). Consequentially, in their view, all growling that they did was voiceless. Future research could investigate the relationship between growls that are perceived as lower, low note vocal fold phonation, and jaw movement. If possible, employing instruments such as EGG could assist in determining whether vocal fold phonation is present or not.

5.2 Variations of Growl

As we noted in section 4.2.2, some differences between GA and GB have been captured in our descriptive data. GB (that is, *Growl 2* for participant 1 and *Florida Style* for participant 2) generally appears to exhibit more tongue retraction and jaw-lowering, especially in participant 2 (*Gothenburg Style* vs *Florida Style*). However, we can once again note that "o" /uː, ʊː/ is behaving slightly differently (see Figure 9).

Of the previous studies which explicitly investigate growling in metal (viz. Eckers et al., 2009; Kato & Ito, 2013), which they call death growl, none of them discuss if *death growl* encapsulates more than one variety of growling, although the findings by Eckers et al. (2014) include that the death growl can be produced in two ways: with the ventricular or aryepiglottic folds. In other words, when previous researchers have investigated growling in metal music specifically, they refer to it as one voice quality by the name death growl, but do not discuss if there are variations of growl within the genre metal, or within the term *death growl*. Likewise, Caffier et al. (2018), Guzman et al. (2014) and Guzman et al. (2019) discuss growl as one voice quality, although they are certainly aware of some variation based on the vocal style in the different genres associated with term *growl*.

If we are indeed looking at two kinds of growl within metal-style growling in two different participants, although this was clearer for participant 2, this is entirely new in phonetic research. As we can see from the literature review regarding musical voice qualities, the notion of *growl* is broad, but the voice quality we were interested in was the growl associated with metal music, which could be most like death growl and grunt. What we might have found is that even within this category of growl in metal music, there could be two metal growls.

Because GB appears to include more TD retraction than GA, we can guess that it is produced with more laryngeal constriction than GA, that is, that it engages valve 4 more than GA does. We should note that participant 2 expressed that, in Florida style, consonants disappear. Perhaps this is a natural consequence of producing the desired voice source in this growl? If the tongue needs to be retracted a lot for this type of growl to be produced, it might be difficult to produce, for example, those consonants which require the tongue tip to touch or reach the alveolar ridge. This could also, naturally, be somewhat determined by each individual's physiology.[2]

In connection with this, we should also add that this participant expressed that he imagined that he was shouting while growling (Appendix B). Recall that some research has found that growl, or rather grunt in CVT, is produced without oscillation anywhere (Aaen et al., 2020; Caffier et al., 2018). It is possible that participant 2's GB employs significant airflow in addition to the laryngeal constriction to enable various structures in the vocal tract to be set into movement. Participant 2 had also mentioned that in Florida Style, you are supposed to

[2] Participant 2 deleted onset "m" while performing the Florida style growl several times

“eat the words”, that is, not pronounce, for example, all the consonants in them (see Appendix B). In future studies, focusing even more on the oral and pharyngeal cavities could be of interest regarding this and other types of growl. In addition to this, further studies into airflow velocity and subglottal pressure in growling could deepen our knowledge regarding how growling may be produced and how it relates to vocal fold phonation. This can then be connected to further studies into jaw movement.

Before summarising the current section, we should consider that it is possible that the methodology caused the participants to exaggerate the differences between their growls. Both participants were aware that two (potentially) different growls were recorded and thus could have tried to exaggerate them to provide better, or clearer, data. Additionally, they both expressed that growling while sitting was difficult due to changes to airflow or breath support. It is possible that they compensated for the lack of breath support by, for example, constricting their larynx slightly more than usual.

When considering the visualisation of our data, we can summarise that GA generally exhibits the pattern we expected based on the LAM, namely, that growling involves more tongue retraction and jaw-lowering. We can further summarise that GB also exhibited this pattern, and that GB appeared to involve even more tongue retraction and jaw-lowering than GA. Despite this, GB also exhibited outliers. Conclusively, the version of growl which generally agrees most strongly with the expectations based on LAM was the one which had outliers. We can only speculate about the potential differences between GA and GB, but there are many interesting possibilities here for future studies.

5.3 Phonemic Variation in the Data

As we noted in section 4.3, there are some outliers in our results. If we suppose that the participants were satisfied with every production of growl they performed in the current study (i.e. that they agree that each recording of GA and GB is a good example of GA or GB), then the outlier in Figure 11, namely, the Italian “u” /u/ vowel, which was more fronted in growling than in modal voice, may disagree with the idea that tongue backing contributes to the desired voice quality. It also appears to disagree with the idea within the LAM that the tongue assists in laryngeal constriction. Here, we can recall that Eckers et al. (2009) found that growling could be produced with ventricular phonation. Perhaps this participant’s GB was produced mainly with the ventricular folds, which do not necessarily require tongue retraction since they can operate on their own at valve 2 (see Edmondson & Esling, 2006, p.

161). However, the pattern of the other vowels appears to suggest that the TD, and thus valve 4, is more engaged in GB compared to GA - although to a lesser extent in this participant.

Alternatively, the data from the Italian “u” /u/ vowel suggests that tongue backing, compared with tongue lowering, has more room for variation without causing the perceived voice quality to change, at least in this individual. This, however, goes against the suggested physiology behind the findings that harshness increases the more vowels become lowered and backed (Rees, 1958 as cited by Esling et al., 2019, p. 67). Additionally, the hyoglossus, which is the muscle responsible for tongue retraction (Baer et al., 1988, p. 15; Honda, 1996, p. 43; Takano & Honda, 2007, p. 56), pulls the TD down and back simultaneously. As such, if the hyoglossus is involved to some extent, which is likely, it would be strange to conclude that tongue backing is allowed to be more varied.

Additionally, the suggested conclusions regarding Italian “u” /u/ can only be true supposing that participant 1 produced the stimuli without too much variation. In Appendix E the means and standard variations of each stimulus, with the participants together and separated, can be found. Regarding TD movement at the front/back dimension in “u” /u/ for Participant 1, the modal version had a relatively low standard deviation of 1.037473, but growl had 3.142087 which is relatively high. This indicates that the position of the TD in the front/back dimension varied somewhat. However, at the TD in this dimension, participant 1 exhibited a higher standard variation in growl compared to modal voice in all vowels but one (see Appendix E).

Moving on, regarding H3, “The difference in mean tongue dorsum or jaw movement across the open/close dimension and front/back dimension happens regardless of vowel”, we can say that the LMM models show that the distinction between growl and modal voice is best described if we include “vowel” as an intercept in all models, and as a slope in two models. This indicates that the differences between the vowels affect the result enough to be considered important for describing the data, but it does not tell us how each vowel affects the result. Our visualisations of the English data suggests that all vowels exhibit varying degrees of TD retraction and jaw-lowering, which supports that the expected pattern happens regardless of vowel, and H3. However, the visualisation of the Italian vowels revealed an outlier (/u/ in the front/back dimension) in which one version of growl exhibited a pattern opposite to what we expected. Conclusively, if we take the Italian “u” /u/ vowel into account, we can tentatively reject the null hypothesis in H3 in this dataset regarding these vowels, and

partially also H1. However, we have, of course, not tested every vowel in existence, and only two people who were vocalising in their L2.

Naturally, the results of the current study were limited to two participants who did not share the same L1, but who were used to growling in their L2. Additionally, the participants experienced that sitting down had affected their growling, and attributed this to how it affected their airflow. This might have affected the current study in some capacity. For example, the changes to the airflow might have caused the participants to overcompensate slightly with their laryngeal and oral articulators to achieve the desired sound. They may also have been changing their tongue and jaw positions between the recording of each stimulus to find the best positions for the articulators. This might account for some of the larger standard deviations for the vowels (see Appendix E).

6. Conclusion

In this study, we concluded that tongue retraction is present in, what could be, several types of heavy metal growling to a significant extent. Additionally, jaw-lowering and backing was significantly present. The results agree with the predictions we made based on concepts within the LAM, and suggested several interesting directions for future research. This included research on variations in growl and how those relate to the tongue and jaw, as well as research into airflow velocity and subglottal pressure. There are, of course, many other aspects of growl which there was no space for in the current study, such as the relationship between different types of growl and laryngeal movement. In the future, we hope to see more, and varied, studies into growl.

References

Aaen, M., McGlashan, J., & Sadolin, C. (2020). Laryngostroboscopic Exploration of Rough Vocal Effects in Singing and their Statistical Recognizability: An Anatomical and Physiological Description and Visual Recognizability Study of Distortion, Growl, Rattle, and Grunt using laryngostroboscopic Imaging and Panel Assessment. *Journal of Voice*, *34*(1), 162.e5-162.e14-162.e14. https://doi-org.ludwig.lub.lu.se/10.1016/j.jvoice.2017.12.020

Aaen, M., McGlashan, J., Christoph, N., & Sadolin, C. (2024). Extreme Vocal Effects Distortion, Growl, Grunt, Rattle, and Creaking as Measured by Electroglottography and Acoustics in 32 Healthy Professional Singers. *Journal of Voice*. https://doi-org.ludwig.lub.lu.se/10.1016/j.jvoice.2021.11.010

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh U.P.

Al-Tamimi, J. (2017). Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations. *Laboratory Phonology, 8*(1): 28. doi: https://doi.org/10.5334/labphon.19

Arch Enemy. (2005). My Apocalypse [Song]. On *Doomsday Machine* [Album]. Century Media Records Ltd.

Arch Enemy. (2017). My Apocalypse [Song]. On *As the Stages Burn* [Album]. Savage Messiah Music; Century Media Records Ltd.

Arvaniti, A. (1999). Standard Modern Greek, in *Journal of the International Phonetics Association* 29(2), 167-172

Arvaniti, A. (2007). Greek Phonetics: The State of the Art. *Journal of Greek Linguistics*, *8*, 97–208. https://doi.org/10.1075/jgl.8.08arv

Baer, T., Alfonso, P., J. Honda, K. (1988). Electromyography of the Tongue Muscles During Vowels in / əpVp/ environment. *Ann Bull RILP*, *22*, pp. 7-19 https://www.umin.ac.jp/memorial/rilp-tokyo/R22/R22_007.pdf (2024-10-15)

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), pp. 1-48. doi:10.18637/jss.v067.i01.

Bailly, L., Henrich Bernardoni, N., Müller, F., Rohlfs, A.-K., & Hess, M. (2014). Ventricular-Fold Dynamics in Human Phonation. *Journal of Speech, Language & Hearing*

Research, *57*(4), 1219–1242. https://doi-org.ludwig.lub.lu.se/10.1044/2014_JSLHR-S-12-0418

Bertinetto, P. M., & Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association*, *35*(2), 131–151.

Boersma, P., & Weenink, D. (2024). *Praat: doing phonetics by computer [Computer program]*. Version 6.4.22, retrieved 5 October 2024 from http://www.praat.org

Caffier, P. P., Ibrahim Nasr, A., Ropero Rendon, M. del M., Wienhausen, S., Forbes, E., Seidner, W., & Nawka, T. (2018). Common Vocal Effects and Partial Glottal Vibration in Professional Nonclassical Singers. *Journal of Voice*, *32*(3), 340–346. https://doi-org.ludwig.lub.lu.se/10.1016/j.jvoice.2017.06.009

Colarusso, J. (1985). Pharyngeals and Pharyngealization in Salishan and Wakashan. *International Journal of American Linguistics*, *51*(4), 366–368.

Complete Vocal Institute. (n.d.). *Complete Vocal Technique*. https://completevocalinstitute.com/complete-vocal-technique/ (Retrieved 2024-10-14).

Death. (1995). *Crystal Mountain* [Song]. On *Symbolic* [Album]. The All Blacks B.V.

Eckers, C., Hütz, D., Kob, M., Murphy, P.J., Houben, D., & Lehnert, B. (2009). Voice production in death metal singers [paper]. *NAG/DAGA 2009, International Conference on Acoustics,* Rotterdam, pp. 1747-1750 https://pub.dega-akustik.de/NAG_DAGA_2009/data/articles/000569.pdf (2024-10-16)

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

Erickson, D., Honda, K., & Kawahara, S. (2017). Interaction of jaw displacement and F0 peak in syllables produced with contrastive emphasis. *Acoustical Science and Technology*, *38*(3), 137-146–146. https://doi.org/10.1250/ast.38.137

Esling, J. H. (2005). There Are No Back Vowels: The Laryngeal Articulator Model. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, *50*(1–4), 13. https://doi.org/10.1353/cjl.2007.0007

Esling, J. H., Moisik, S, R., Benner, A., and Crevier-Buchman, L. (2019). *The Laryngeal Articulator Model*. Cambridge: Cambridge university press.

Fant, G. (1960). *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations*. Mouton & Co.

Garellek, M. (2022). Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality. *Journal of Phonetics*, *94*. https://doi-org.ludwig.lub.lu.se/10.1016/j.wocn.2022.101155

Gerratt, B. R. and Kreiman, J. (2001). Toward a taxonomy of nonmodal phonation. *Journal of phonetics*, *29*(4), 365-381. https://doi-org.ludwig.lub.lu.se/10.1006/jpho.2001.0149

Gick, B., Wilson, I., and Derrick, D. (2013). *Articulatory Phonetics*. Oxford: Blackwell.

Giegerich, H. J. (1992). *English phonology: An introduction*. Cambridge University Press.

Gordon, M., & Ladefoged, P. (2001). Phonation Types: A Cross-Linguistic Overview. *Journal of Phonetics*, *29*(4), 383–406. https://doi-org.ludwig.lub.lu.se/10.1006/jpho.2001.0147

Guzman, M., Muñoz, D., Barros, M., Espinoza, F., Herrera, A., Parra, D., & Lloyd, A. (2014). Laryngoscopic, acoustic, perceptual, and functional assessment of voice in rock singers. *Folia Phoniatrica et Logopaedica*, *65*(5), 248-256–256. https://doi-org.ludwig.lub.lu.se/10.1159/000357707

Guzman, M., Acevedo, K., Leiva, F., Ortiz, V., Hormazabal, N., & Quezada, C. (2019). Aerodynamic Characteristics of Growl Voice and Reinforced Falsetto in Metal Singing. *Journal of Voice*, *33*(5), 803. https://doi-org.ludwig.lub.lu.se/10.1016/j.jvoice.2018.04.022

Hagen, R. (2023). On Horseback They Carried Thunder: The Second Lives of Norwegian Black Metal. In Herbst, J., P. (Ed.), *The Cambridge Companion to Metal Music* (pp. 221-236). Cambridge University Press.

Hollien, H. (1974). On Vocal Registers. *Journal of Phonetics*, *2*, 125–143. https://doi-org.ludwig.lub.lu.se/10.1016/s0095-4470(19)31188-x

Honda, K. (1996). Organization of Tongue Articulation for Vowels. *Journal of Phonetics*, *24*(1), 39–52. https://doi-org.ludwig.lub.lu.se/10.1006/jpho.1996.0004

Kato, K., & Ito, A. (2013). Acoustic Features and Auditory Impressions of Death Growl and Screaming Voice. *2013 Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing* [Paper], Beijing, 460–463. https://doi-org.ludwig.lub.lu.se/10.1109/IIH-MSP.2013.120

Kreiman, J. and Sidtis, D. (2011). Introduction. In *Foundations of Voice Studies* (eds Kreiman, J. and Sidtis, D.). https://doi.org/10.1002/9781444395068.ch1

Kreiman, J. (2024). Information Conveyed by Voice Quality. *The Journal of the Acoustical Society of America*, *155*(2), 1264-1271. https://doi.org/10.1121/10.0024609

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. Univ. of Chicago P.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge U.P.

Lengeris, A. (2016). Comparison of perception-production vowel spaces for speakers of Standard Modern Greek and two regional dialects. *Journal of the Acoustical Society of America*, *140*(4), EL314-EL319. https://doi.org/10.1121/1.4964397

Meme. (n.d.) In *Cambridge Dictionary*. https://dictionary.cambridge.org/dictionary/english/meme

Moisik, S. R., Esling, J. (2007). 3D Auditory-Articulatory Modeling of the Laryngeal Constrictor Mechanism [Paper]. *16th International Congress of Phonetic Sciences,* Saarbrücken*,* 373-378.

Moisik, S. R., Esling, J. H., & Crevier-Buchman, L. (2010). A high-speed laryngoscopic investigation of aryepiglottic trilling. *Journal of the Acoustical Society of America*, *127*(3), 1548–1558. https://doi-org.ludwig.lub.lu.se/10.1121/1.3299203

Moisik, S. R., Esling, J. H., Crevier-Buchman, L., Amelot, A., Halimi, P. (2015). Multimodal Imaging of Glottal Stop and Creaky Voice: Evaluating the Role of Epilaryngeal Constriction. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (paper 247). Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Retrieved at: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0247.pdf

Pariente, I. (2015). The Interaction of Vowel Quality and Pharyngeals in Sephardic Modern Hebrew. *Folia Linguistica: Acta Societatis Linguisticae Europaeae*, *49*(2), 421–438. https://doi-org.ludwig.lub.lu.se/10.1515/flin-2015-0015

Posit team. (2024). *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

Pring, J. T. (1982). αλεπού. *The Oxford dictionary of modern Greek : Greek-English and English-Greek* (2nd ed., p. 7). Clarendon P.

R Core Team (2024). _R: A Language and Environment for Statistical Computing_. *R Foundation for Statistical Computing*. Vienna, Austria. https://www.R-project.org/

Ruge, H. (1974). *Nygrekisk fonetik* (2. uppl.). Stockholms univ., Inst. för klassiska språk.

Sadolin, C. (2021). *Complete Vocal Technique*. Denmark: Tarm Bogtryk A/S

Sakakibara, K.-I., Fuks, L. Imagawa, H., and Tayama, N. (2004). Growl voice in ethnic and pop styles. In *Proceedings of the Intl. Symposium on Musical Acoustics (ISMA)*, Nara, Japan.

Schötz, S., Frid, J., Gustafsson, L., & Löfqvist, A. (2013). Functional Data Analysis of Tongue Articulation in Palatal Vowels: Gothenburg and Malmöhus Swedish /i:, y:, u-:/. *Interspeech 2013,* 1326-1330, doi: 10.21437/Interspeech.2013-352

Smith, T. S. (1977). *A phonetic study of the function of the extrinsic tongue muscles*.

Stuart-Smith, J. (1999). Voice Quality in Glaswegian. In *ICPhS-14*, 2553-2556.

Svensson Lundmark, M. (2020). *Articulation in time : some word-initial segments in Swedish*. Centre for Languages and Literature, Lund University.

Takano, S., & Honda, K. (2007). An MRI analysis of the extrinsic tongue muscles during vowel production. *Speech Communication*, *49*(1), 49–58. https://doi-org.ludwig.lub.lu.se/10.1016/j.specom.2006.09.004

Vietti, A. & Mereu, D. (2023). Mid vowels at the crossroads between standard and regional Italian. *Sociolinguistica*, *37*(1), 17-39. https://doi-org.ludwig.lub.lu.se/10.1515/soci-2022-0033

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). _*dplyr: A Grammar of Data Manipulation*_. R package version 1.1.4. https://CRAN.R-project.org/package=dplyr

Wieling, M., & Tiede, M. (2017). Quantitative identification of dialect-specific articulatory settings. *Journal of the Acoustical Society of America*, *142*(1), 389–394. https://doi-org.ludwig.lub.lu.se/10.1121/1.4990951

Winter, B. (2020). *Statistics for linguists. an introduction using R*. Routledge, Taylor & Francis Group.

Appendix A: Interview Questions

Interview Questions: First interview (Both Participants)

Have you received any voice training?

> (If yes) When you received voice training, was it geared towards a specific kind or in a specific paradigm?

How did you learn how to growl?

Did you find it difficult to start growling?

For how long have you been growling?

Are you currently still growling?

Are there any voice techniques you can do other than growling?

Do you have some idea what happens when you're preforming these techniques?

Are you inspired by any particular vocalists?

When you are growling, do you have preference for how to move?

Is it difficult to growl if you're sitting down?

Do you ever use growl in contexts other than performance?

How is your voice doing overall?

Do you have any questions for me?

Do you need some time to warm up before recording with the articulograph?

Interview Questions: Second interview (Participant 1)

Background Qs

How old are you?

Do you speak any languages other than Italian and English?

About the EMA recording

During the articulography recording, was there anything that was difficult to growl?

Do you think that the fact that you were sitting down during the articulography recording impacted your growling?

Clarifications about things that came up during the first interview

In the last interview, you used the term *compression*. Would you elaborate on what you mean by *compression*?

You also mentioned something about a pitched growl. What is a pitched growl?

Is it possible to growl with and without voice?

In the recording of our last interview, right after I asked if there were any specific vocalists that influence you, there is a bit where I couldn't quite tell what you were saying, but it might've been the name of a vocalist. Would you mind repeating the names, and affiliated bands, of the vocalists that have inspired you?

In the previous interview you mentioned sometimes using *fry* vocals. What are fry vocals?

Interview Questions: Second interview (Participant 2)

Background Qs

How old are you?

Do you speak any other languages?

About the EMA recording

During the articulography recording, was there anything that was difficult to growl?

Do you think that the fact that you were sitting down during the articulography recording impacted your growling?

Clarifications about things that came up during the first interview

As an example of Florida style growling, you mentioned the band Cannibal Corpse. Have all vocalists of this band performed in this style?

When you told me about the types of growling that you can do, you mentioned sometimes trying a 'out of the box' growl. The example you gave was trying to imitate Chuck from Death. Specifically, his growling. Could you give an example of a song or part of a song that you would try to imitate?

In the previous interview, you mentioned that in the Gothenburg style of growling, you scream a bit more. What is a scream to you?

Last time we talked about growling standing up versus sitting down and you mentioned that sitting down was more difficult because you're more relaxed sitting down. Is being in a relaxed position difficult because of the emotional aspects of growl, because of physiological aspects, or some other aspect?

Appendix B: Full Interview Results

Participant 1

This, male (age 32), participant had Italian as their L1 and English as their L2. He has received some voice training for clean singing. This voice training included basics like breathing and was not within any specific paradigm. He has been growling for 15 years and learned from YouTube videos and experimenting, and cited Randy Blythe of the band *Lamb of God* as his main inspiration. He would also attempt to imitate Mikael Åkerfelt (Opeth). Participant 1 is still growling regularly, although more so when he is in Italy. Learning to growl was especially difficult in the beginning, and the participant thinks that this was the case because he had not received any formal training in singing yet and thus had not learned how to breathe properly. The participant described that in the beginning, "I felt that I was kind of ruining something like pushing myself too much […] I was just screaming loud in some random way that was kind of painful in the beginning".

Participant 1 considered himself to use a single growl technique, and when asked if he could produce any other similar techniques or multiple types of growl, he expressed that he could modify his standard growl into what might be a different version of his growl, for example to make it sound lower, but that he did not know if that version was actually different. He also mentioned something which had not shown up in the literature review for the current study, namely, that he did not know the *fry technique*, but that he did use *fry vocals* and *false vocals*, and tried to do some *blends* sometimes. Based on the current study's literature review, it is unclear what the *fry technique*, *fry vocals*, or *blends* might refer to. He also mentioned that he never uses *squeals* or *inhales*. Finally, he expressed that growling was more difficult to do while sitting down compared to standing up, and in the second interview, he confirmed that sitting down affected his growl.

Participant 2

Participant 2 (age 44) has the L1 Greek. He considered himself to know primarily one type of growling, which he called *melodic* or *clearly articulated growling*. He likened this to the Gothenburg death metal style, but also said that he was capable of preforming another style (Florida style). He explained that what he refers so as *melodic* or *clearly articulated* is like the Gothenburg style because it is a "more melodic clearly articulated style". He mentioned that "you can still kind of make out the words" and that the "growling follows at least partially the melodic line of the guitar or some kind of melodic pattern". In the stimuli of the

current study, this participant's growls were called *Growl G* (for Gothenburg) and *Growl F* (for Florida).

Regarding the Florida Style, participant 2 said it is "more base heavy (.) articulation isn't really a concern if anything you know you're supposed to what I call eat the words so that you don't pronounce (.) a lot of the consonants for example". As an example of this style, participant 2 mentions the band Cannibal Corpse (both the previous and current vocalist). As a follow-up question, he was asked if he ever growled with a very specific tongue position such as in a retroflex position, he mentions that he typically does that with the Florida style. "what I typically do is put a lot of air under my jaw so I feel like this gives the base, the resonance you know, of the voice needs and sing almost from like the top of my mouth where as for the Gothenburg style is more you feel it more in your throat cause you scream a bit more so I feel like the sides are doing more work than the than the bottom of the of the jaw".

Additionally, he also mentioned that he would attempt to imitate "out of the box" styles. The example he gave here is Chuck Schuldiner (of the band *Death*). He gave the song *Crystal Mountain* as an example. This song is found on Death's (1995) album *Symbolic*. However, he found Chuck's growling to a bit difficult to imitate so he would end up producing a growl more like the Gothenburg style.

Another interesting thing to note is that, when asked about what he thinks he might be doing when performing growl, he mentions that the first advice he would give to young growlers is that they need to be willing to shout. This is because "you can be a timid person in your everyday life but in front of the microphone you have to you know be willing to shout […] I've felt I've always felt as I was approaching the mic I need to now shout I need to be able to be heard I mean in in a rehearsal situation where we had two guitars base and the drums that was (.) drums that was playing blast beats you need to be able to cut over that so my first thing is always okay I'm gonna' shout as much as I can (.) and then the other thing the second thing that was happening I felt was that now I need to also shout but also to growl."

Finally, like participant 1, he expressed that growling is difficult to do while sitting down. In the second interview, he also expressed that his growling was affected by having to sit down.

Appendix C: Stimuli and Recording Paradigms

Recording 1, Participant 1

1. Warm up phrases (to practice speaking with the sensors)
2. Nonce words and English lyric in random order (x5)
3. Nonce words and English lyric in random order in Growl 1 (x5)
4. Nonce words and Italian lyric in random order (x5)
5. Nonce words and Italian lyric in random order in Growl 1 (x5)
6. Modal-Growl-Modal Growl 1 (x5)
7. Nonce words and English lyric in random order in Growl 2 (x5)
8. Nonce words and Italian lyric in random order in Growl 2 (x5)

Note that *growl 2* was placed last. This is because (1) the participant only viewed it as a modification of their standard growl and (2) because the participant expressed during the interview that they did not use this technique as much, and we wanted to collect the more secure data first without tiring the participant out too much.

Recording 2, Participant 2

1. Warm up phrases (to practice speaking with the sensors)
2. English word – Nonce word (x1) **Step 2 and 3 repeated x2**
3. Nonce words in random order (x3)
4. Nonce words in random order in Growl A (x6)
5. Modal-Growl-Modal Growl A (x6)
6. Nonce words in random order in Growl B (x6)
7. English lyric in Growl A
8. Greek lyric in Growl A

Similarly to participant 1, participant 2 expressed that growl A was a bit more tiring to perform which is why, similarly to the paradigm for participant 1, growl B was not included in every task. Furthermore, because this experiment requires that the participants intend to produce the same vowel, a practice set was introduced for participant 2. This task was woven together with the data collection of modal voice nonce words. Each nonce word was introduced together with an English or Greek word. The participant' pronunciation of the vowel in the English/Greek-based word was the vowel they were instructed to produce in the nonce word: Bee-Meem, Gym – Myym, Moon – Moom, Car – Maam. For the real Greek

words, the name of the letter representing the vowel was used (e.g. όμικρο ’omikro’). For the [u] vowel, which is spelled *ου* (Ruge 1974, p. 4), this was not possible. Therefore, a dictionary was consulted and a neutral noun which contained the vowel was chosen: αλεπού ‘alepu’, which means fox (Pring, 1984). Additionally, the number of repetitions per stimuli were increased to 6. The time to run this experiment was calculated to be ca 45 minutes. The following words and were used for the Greek practice and modal stimuli:

άλφα – μααμ	‘alpha’	/mɐm/
έψιλο – μεεμ	‘epsilon’	/mɛm/
ιώτα – μιιμ	‘iota’	/mim/
όμικρο – μοομ	‘omikro’	/mom/ /mɔm/
αλεπού – μουμ	‘alepu’ (fox)	/mum/

Appendix D: Praat Script Used for Retrieving EMA Data

This script was created by Dr. Johan Frid (researcher at Lund University Humanities Lab)

```
path_to_data$ = "C:/Users/Name/data/"

sweep$ = "0103"

channel = 24 # JWz


Read from file: path_to_data$+"Folder_name/pos/"+sweep$+".txt"

Down to Matrix

Transpose

To Sound (slice): channel

Scale times to: 0, 1

Override sampling frequency: 250


Read from file: path_to_data$+"Folder_name/wav/"+sweep$+".wav"
```

Appendix E: Means and standard deviations

Table 2: Mean and Standard Deviation of Vowels at the TD, up/down dimension, time-point C in both participants, and in each participant (P1 and P2). Data is from English stimuli only.

Both	**Vowel**	**Modal**		**Growl**	
		Mean	Sd	Mean	Sd
	/ɑː/	-45.23697	11.720700	-47.73656	13.83328
	/iː/	-21.70000	8.092715	-30.93138	13.75945
	/uː, ʊː/	-38.49273	13.558608	-45.90138	14.65634
	/ɪ/	-29.10867	10.257686	-37.02069	14.30089
P1	**Vowel**	**Modal**		**Growl**	
		Mean	Sd	Mean	Sd
	/ɑː/ - Eng	-32.96533	3.526098	-33.57533	2.439713
	/iː/ - Eng	-13.34000	1.001965	-17.76143	2.304467
	/uː, ʊː/ - Eng	-25.29267	3.796130	-29.09500	3.513275
	/ɪ/ - Eng	-19.33214	2.841419	-22.12385	3.177039
	/a/ - It	-36.0780	0.9158439	-36.720	1.762334
	/ɛ/ - It	-20.0960	2.6962715	-30.162	1.719159
	/i/ - It	-13.21250	1.025878	-19.35900	1.015310
	/u/ - It	-23.4180	1.230374	-30.49000	1.940842
P2	**Vowel**	**Modal**		**Growl**	
		Mean	Sd	Mean	Sd
	/ɑː/ - Eng	-55.46333	2.177659	-60.23176	3.414096
	/iː/ - Eng	-28.94533	2.074508	-43.22333	6.195296
	/uː, ʊː/ - Eng	-49.49278	7.232908	-57.76471	2.460290
	/ɪ/ - Eng	-37.66313	5.400247	-49.12438	5.009431
	/ɐ/ - Gr	-55.244	2.8287064	-62.45583	2.158731
	/ɛ/- Gr	-39.266	2.2386223	-54.01167	3.571383
	/i/- Gr	-27.20000	1.2283526	-43.52250	7.676333
	/o, ɔ/ - Gr	-54.502	2.8894757	-63.01583	2.903651
	/u/ - Gr	-44.61200	0.6263944	-56.51583	2.885158

Table 3: Mean and Standard Deviation of Vowels at the TD, front/back dimension, time-point C, in both participants, and in each participant (P1 and P2). Data is from English stimuli only.

Both	**Vowel**	**Modal**		**Growl**	
		Mean	Sd	Mean	Sd
	/ɑː/	10.138303	11.123461	7.057500	10.906753
	/iː/	21.617143	6.163954	15.662276	8.624214
	/uː, ʊː/	9.646061	9.841311	4.772241	10.777944
	/ɪ/	17.548333	8.171328	12.927862	10.133618
P1	**Vowel**	**Modal**		**Growl**	
		Mean	Sd	Mean	Sd
	/ɑː/ - Eng	19.99000	5.333720	17.02733	3.873764
	/iː/ - Eng	27.57385	1.275451	22.98643	3.147877
	/uː, ʊː/ - Eng	16.64400	4.252389	15.36833	4.615548
	/ɪ/ - Eng	22.79643	1.784388	22.57385	2.773883
	/a/ - It	20.7460	1.008876	10.8242	2.217059
	/ɛ/ - It	28.4340	1.471234	23.5980	2.427966
	/i/ - It	27.9825	1.070370	22.8530	3.474194
	/u/ - It	15.6700	1.037473	14.9600	3.142087
P2	**Vowel**	**Modal**		**Growl**	
		Mean	Sd	Mean	Sd
	/ɑː/ - Eng	1.928556	7.244007	-1.739412	6.476258
	/iː/ - Eng	16.454667	3.221446	8.826400	5.986869
	/uː, ʊː/ - Eng	3.814444	9.391419	-2.707353	6.707636
	/ɪ/ - Eng	12.956250	8.839172	5.090250	6.274767
	/ɐ/ - Gr	4.206	5.301352	-6.2545000	4.531999
	/ɛ/- Gr	8.794	4.243257	-0.5133333	4.423000
	/i/- Gr	16.760	3.224081	6.3150000	4.127452
	/o, ɔ/ - Gr	3.986	4.763516	-7.7550000	3.485981
	/u/ - Gr	5.632	4.315214	-3.6050000	3.634070

Table 4: Mean and Standard Deviation of Vowels at the Jaw, up/down dimension, time-point C, in both participants, and in each participant (P1 and P2). Data is from English stimuli only.

Both	**Vowel**	**Modal**		**Growl**	
		Mean	Sd	Mean	Sd
	/ɑː/	-65.50697	7.877160	-69.86906	7.442765
	/iː/	-51.66250	7.682754	-60.17103	7.813089
	/uː, ʊː/	-55.72848	12.192848	-64.03241	9.878693
	/ɪ/	-51.38033	9.582698	-61.47931	9.662058
P1	Vowel	**Modal**		**Growl**	
		Mean	Sd	Mean	Sd
	/ɑː/ - Eng	-58.07133	4.139189	-63.94667	4.989710
	/iː/ - Eng	-43.94692	2.723241	-54.37571	7.172289
	/uː, ʊː/ - Eng	-44.02667	2.558302	-53.72000	5.971096
	/ɪ/ - Eng	-41.56500	1.240625	-52.78615	7.764426
	/a/ - It	-61.246	2.453065	-70.457	1.867619
	/ɛ/ - It	-53.502	3.729185	-64.949	1.553043
	/i/ - It	-45.76500	1.811307	-58.03800	1.829613
	/u/ - It	-42.12200	1.629669	-59.14200	2.871224
P2	Vowel	Modal		Growl	
		Mean	Sd	Mean	Sd
	/ɑː/ - Eng	-71.70333	3.640375	-75.09471	4.862674
	/iː/ - Eng	-58.34933	2.071523	-65.58000	3.064141
	/uː, ʊː/ - Eng	-65.48000	7.276164	-71.31176	3.194954
	/ɪ/ - Eng	-59.96875	2.757547	-68.54250	2.706458
	/ɐ/ - Gr	-72.438	5.0162755	-82.54833	3.408601
	/ɛ/- Gr	-67.170	1.7408044	-78.32250	6.011890
	/i/- Gr	-57.848	0.8761963	-68.85417	4.189408
	/o, ɔ/ - Gr	-65.286	3.2846887	-79.77500	4.971492
	/u/ - Gr	-57.662	0.9607133	-69.96250	4.279040

Table 5: Mean and Standard Deviation of Vowels at the Jaw, front/back dimension, time-point C, in both participants, and in each participant (P1 and P2). Data is from English stimuli only.

Both	**Vowel**	**Modal**		**Growl**	
		Mean	Sd	Mean	Sd
	/ɑː/	47.25333	1.8483856	45.33125	2.365735
	/iː/	52.64286	0.9670798	51.00793	1.691925
	/uː, ʊː/	49.97545	2.8792470	48.20414	2.484018
	/ɪ/	51.34567	0.9653682	49.01828	1.945838
P1	Vowel	**Modal**		**Growl**	
		Mean	Sd	Mean	Sd
	/ɑː/ - Eng	48.57200	1.1233191	46.28867	2.0346599
	/iː/ - Eng	52.81538	1.1066588	51.92786	1.2316542
	/uː, ʊː/ - Eng	51.63733	0.8861108	50.53417	0.7922633
	/ɪ/ - Eng	51.42286	0.9383326	50.64231	1.0741055
	/a/ - It	48.2820	1.0334747	42.734	1.4488248
	/ɛ/ - It	50.2200	0.7084137	47.246	0.8932861
	/i/ - It	52.72750	0.6425146	51.99400	1.1686573
	/u/ - It	52.70200	0.2937176	50.51200	1.1458602
P2	Vowel	**Modal**		**Growl**	
		Mean	Sd	Mean	Sd
	/ɑː/ - Eng	46.15444	1.6059420	44.48647	2.367200
	/iː/ - Eng	52.49333	0.8380647	50.14933	1.637221
	/uː, ʊː/ - Eng	48.59056	3.2389059	46.55941	1.849488
	/ɪ/ - Eng	51.27813	1.0140231	47.69875	1.414826
	/ɐ/ - Gr	45.384	1.9668071	40.72250	1.378287
	/ɛ/- Gr	47.482	0.6701269	43.84417	1.982562
	/i/- Gr	52.194	0.5300283	49.39417	1.421558
	/o, ɔ/ - Gr	48.278	1.7544999	41.03833	2.805582
	/u/ - Gr	51.798	0.2631919	47.02167	2.086209

Appendix F: Additional Figures

Figure F1: Participant 1, Jaw, English, Up/Down

Figure F2: Participant 1, Jaw, Italian, Up/Down

Figure F3: Participant 1, Jaw, Italian, Front/Back

Figure F4: Participant 2, Jaw, English, Up/Down

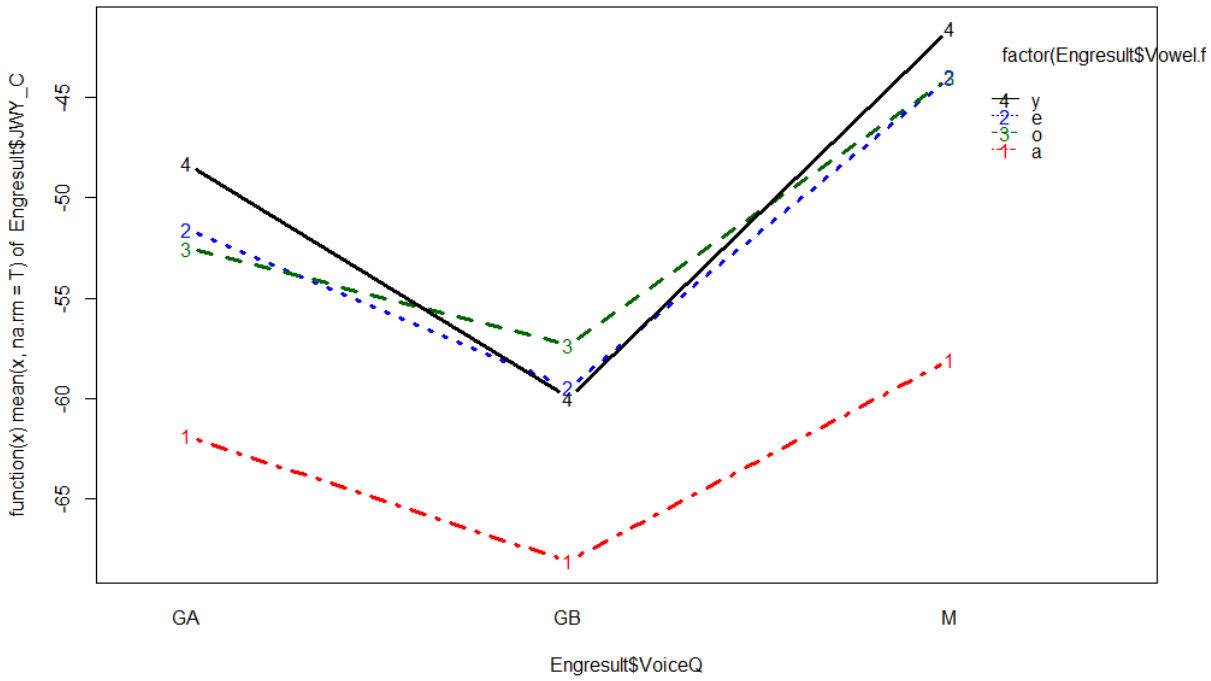
Figure F5: Participant 2, Jaw, English, Front Back

Figure F6: Participant 2, Jaw, Greek, Up/Down

Figure F7: Participant 2, Jaw, Greek, Front Back

Figure F1

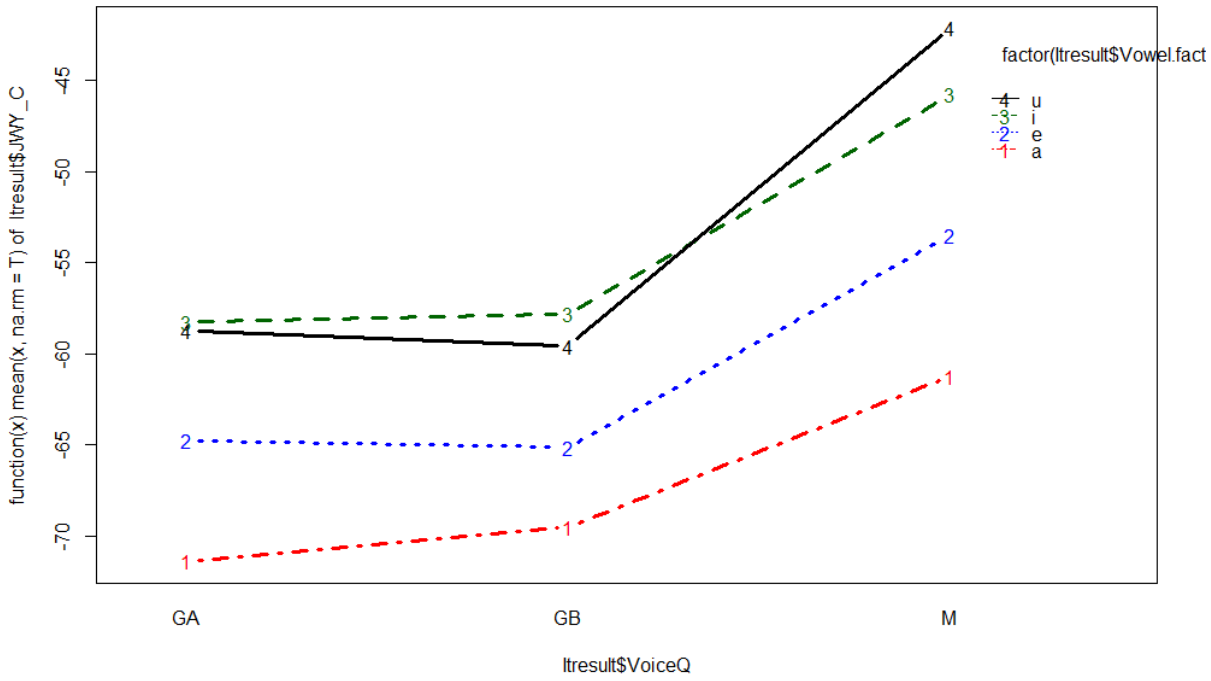
Participant 1, Jaw, English, Up/Down



Note: Mean coordinate at time point C in every vowel separated, in each voice quality, at the jaw. Data is from participant 1, in English. Lower values represent a more lowered jaw.

Figure F2

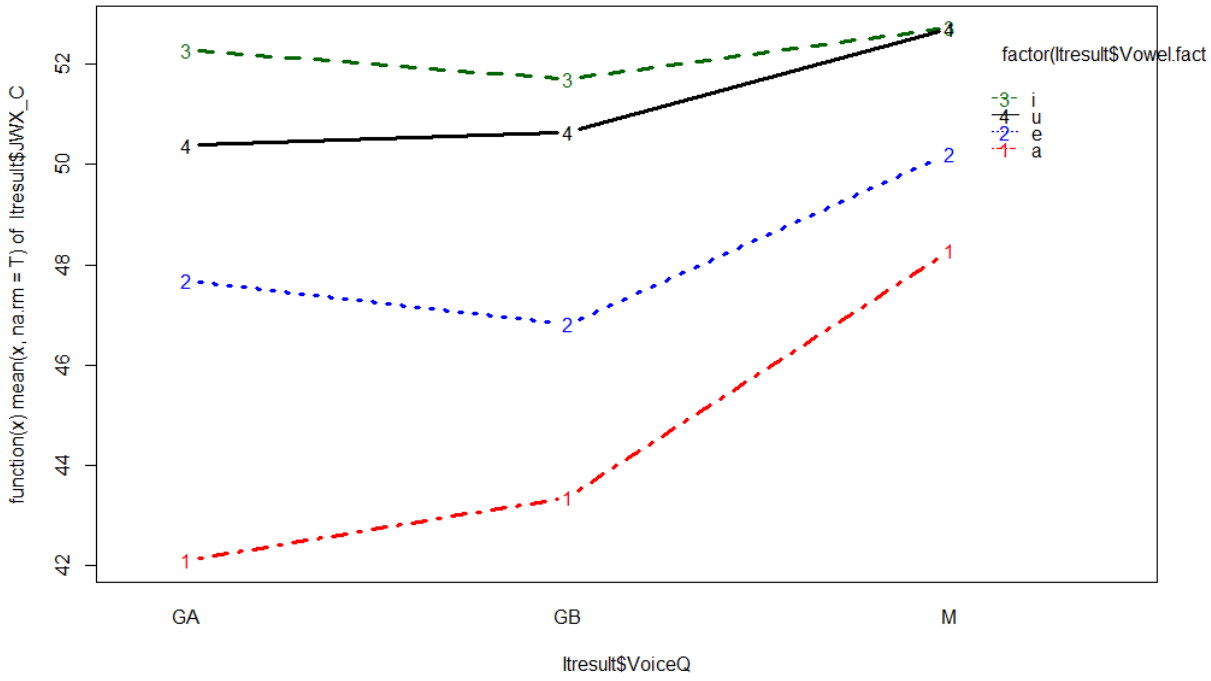
Participant 1, Jaw, Italian, Up/Down



Note: Mean coordinate at time point C in every vowel separated, in each voice quality, at the jaw. Data is from participant 1, in Italian. Lower values represent a more lowered jaw.

Figure F3

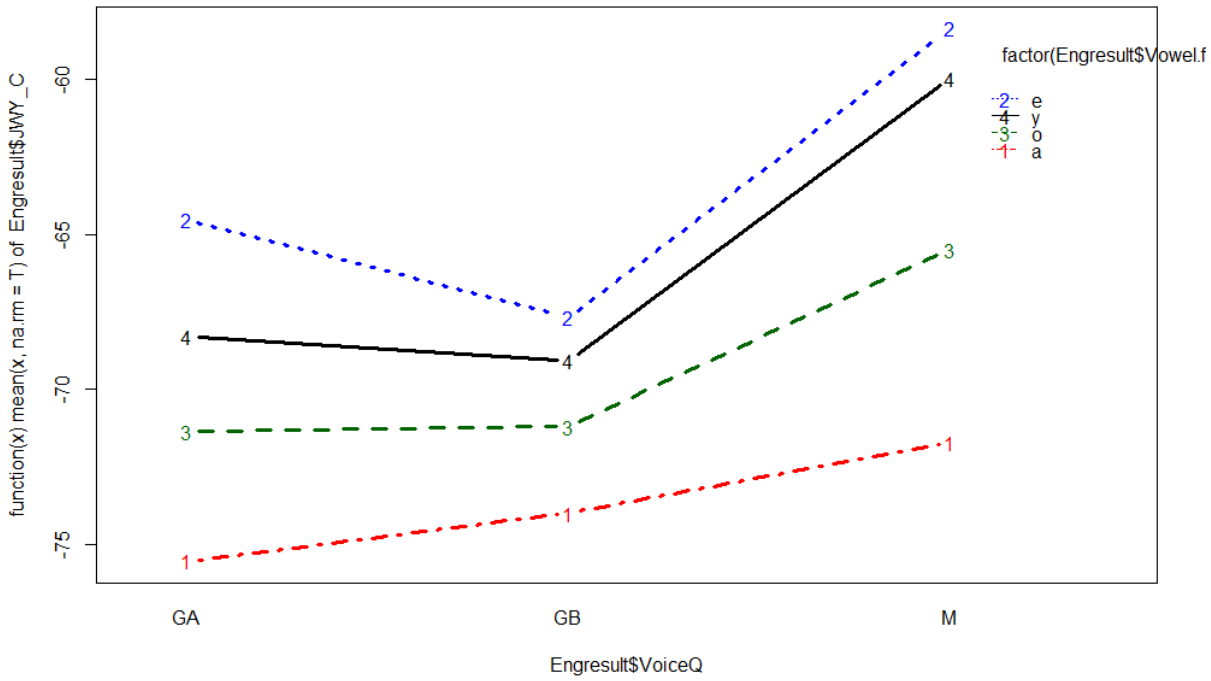
Participant 1, Jaw, Italian, Front/Back



Note: Mean coordinate at time point C in every vowel separated, in each voice quality, at the jaw. Data is from participant 1, in Italian. Lower values represent a more backed jaw.

Figure F4

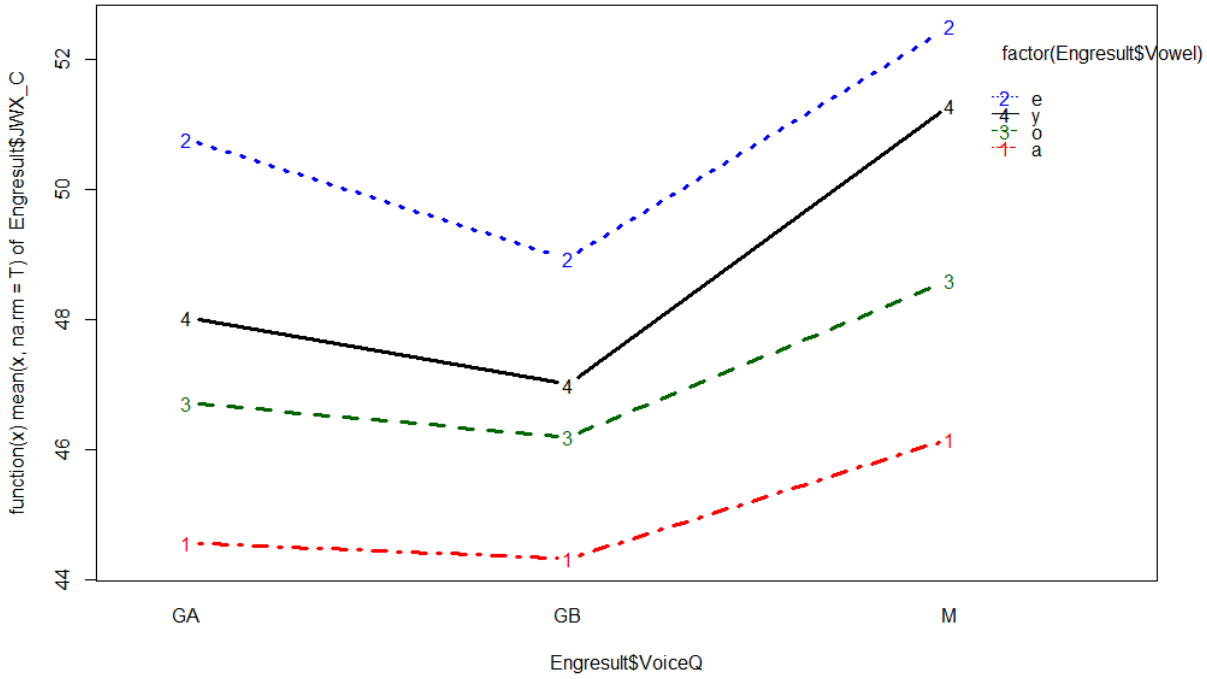
Participant 2, Jaw, English, Up/Down



Note: Mean coordinate at time point C in every vowel separated, in each voice quality, at the jaw. Data is from participant 2, in English. Lower values represent a lower jaw.

Figure F5

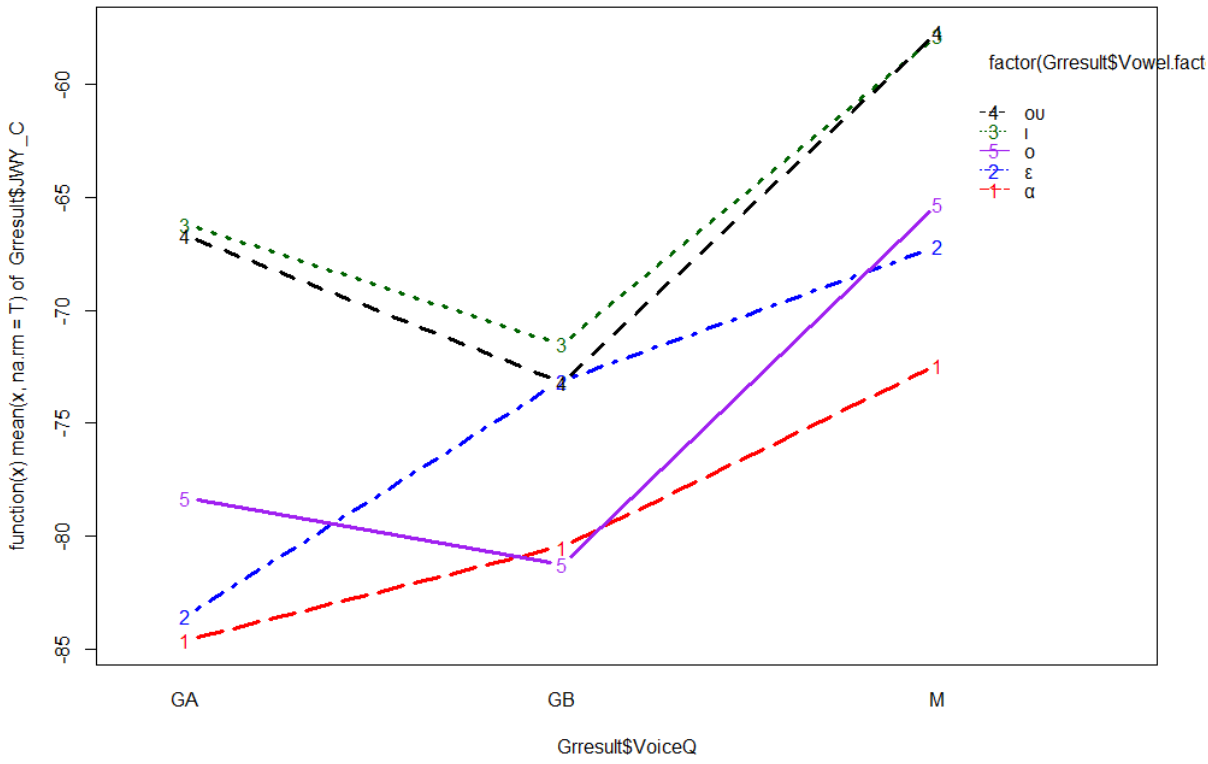
Participant 2, Jaw, English, Front Back



Note: Mean coordinate at time point C in every vowel separated, in each voice quality, at the jaw. Data is from participant 2, in English. Lower values represent a more backed jaw.

Figure F6

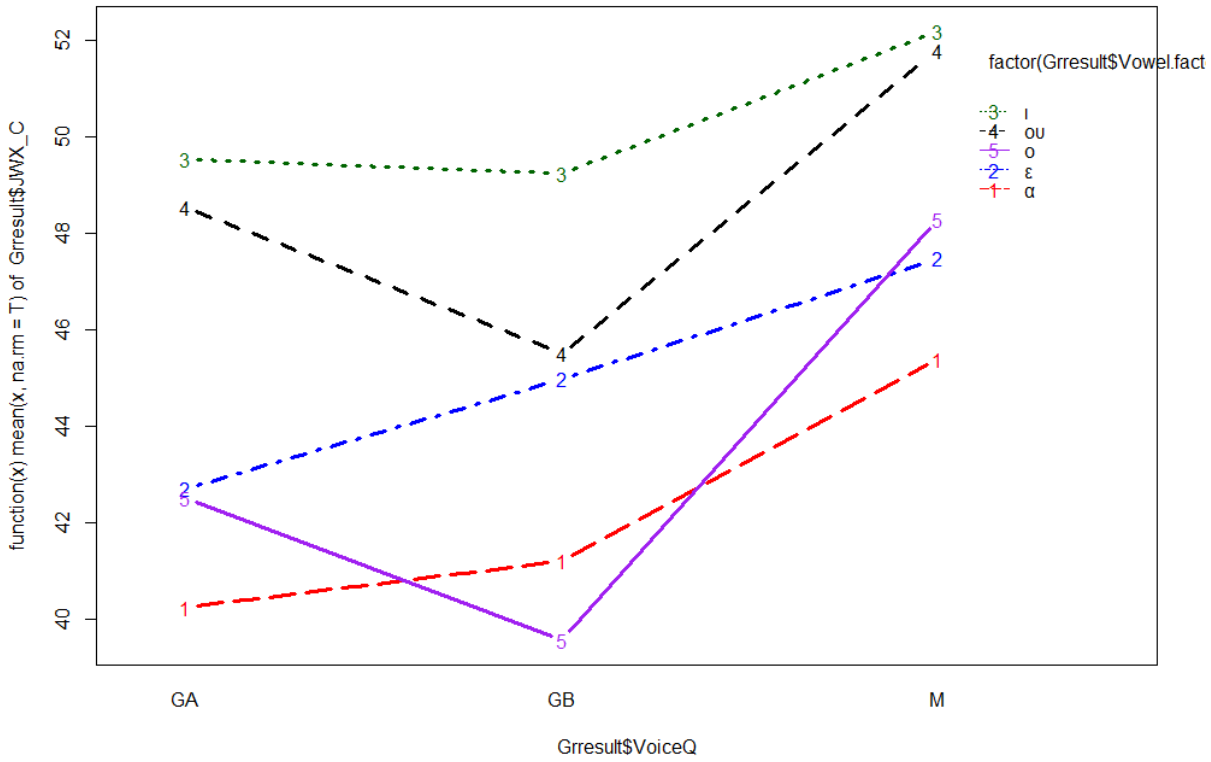
Participant 2, Jaw, Greek, Up/Down



Note: Mean coordinate at time point C in every vowel separated, in each voice quality, at the jaw. Data is from participant 2, in Greek. Lower values represent a lower jaw.

Figure F7

Participant 2, Jaw, Greek, Front Back



Note: Mean coordinate at time point C in every vowel separated, in each voice quality, at the jaw. Data is from participant 2, in Greek. Lower values represent a more backed jaw.