

JOURNAL ARTICLE

# Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations

Jalal Al-Tamimi

Speech and Language Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

Jalal.Al-Tamimi@newcastle.ac.uk

This exploratory study of Jordanian and Moroccan Arabic (JA and MA) aims to evaluate whether pharyngealization is associated with an epilaryngeal constriction which causes ‘retraction’ and tense voice quality in surrounding vowels, following the Laryngeal Articulator Model (LAM) (Esling, 2005). Twenty male speakers (10 per dialect) produced vowels preceded by /d or d<sup>ɣ</sup>/. Thirteen acoustic correlates obtained at the onset and midpoint were used to assess this type of constriction. A predictive modeling approach was used; starting with Bayesian Generalized Linear Mixed Effects modeling followed by Conditional Random Forest for classification. Vowels in the pharyngealized context were more open (higher F1, Z1-Z0), more back (lower F2, higher Z3-Z2), more compact (lower Z2-Z1), and showed spectral divergence (higher Z3-Z2). Voice quality results showed these vowels to be produced with a tense voice. High classification rates of 93.5% for JA and 91.1% for MA were obtained and variable importance score showed formant-based measures outperform voice quality ones. This suggests pharyngealization has ‘retraction,’ with a *back and down* gesture, as a primary correlate followed by [+CONSTRUCTED GLOTTIS]. The implications of these results provide strong support for LAM, the feature [+CET], and the use of the epilarynx to describe pharyngealization.

**Keywords:** acoustics of pharyngealization; tense/pressed voice; constricted epilarynx; Bayesian GLMM; Random Forests; Jordanian and Moroccan Arabic

## 1 Introduction

Pharyngealization (or emphasis) in Arabic is generally assumed to involve retraction of the tongue dorsum towards the upper pharyngeal area, which leads to a lowering of the second formant in the surrounding vowels (e.g., Bin-Muqbil, 2006; Ghazeli, 1977; Watson, 2007; Zawaydeh, 1999; Zawaydeh & de Jong, 2011). Although these characteristics are mostly agreed upon, pharyngealization is also associated with a retracted epiglottis, a raised larynx, a pressed/tense voice quality, and/or a protruded lip posture (see e.g., Al-Tamimi, F. & Heselwood, 2011; Cantineau, 1960; Hess, 1998; Laufer & Baer, 1988; Lehn, 1963; Zeroual & Clements, 2015; Zeroual et al., 2011, among others). Although located near the constriction observed for ‘true’ pharyngeals, authors claimed that the two, i.e., ‘true’ pharyngeals and pharyngealization, share the same place but vary in degree of constriction (e.g., Laufer & Baer, 1988). Hence, and following the Laryngeal Articulator Model (Esling, 2005), both will share an epilaryngeal constriction that may be exhibited differently. An epilaryngeal constriction causes the tongue root and body to be pulled *back and down* in a one combined gesture and causes the vowels to be produced with the ‘retracted’ quality (Esling, 2005; Moisiuk, 2013a; Sylak-Glassman, 2014a, b). A secondary

consequence of an epilaryngeal constriction causes a change in the voice source with an increase in harmonic amplitude especially in high frequencies (Halle & Stevens, 1969; Laver, 1980; Moisik, 2013b; Moisik & Esling, 2010; Stevens, 1977, 1998; Story, 2016). This increase is generally associated with a tense/creaky/harsh voice quality (Edmondson & Esling, 2006; Kuang & Keating, 2012, 2014; Moisik, 2013b). Our aim in this study is to provide a complete analysis of the acoustics of Arabic pharyngealized plosives in order to evaluate the presence of acoustic evidence for epilaryngeal constriction. In addition to the traditional use of absolute formant frequencies, we will investigate Bark-difference metrics as an alternative correlate to the articulatory ‘retracted’ vowel quality that is associated with an epilaryngeal constriction (Esling, 2005). Acoustic correlates of voice quality via harmonic differences will also be investigated to assess the presence of a tense voice quality as a consequence of an epilaryngeal constriction.

The paper is organized as follows. The sections following this introduction provide an overview of the literature on the phonetics of pharyngealization, with special attention on the consequences of epilaryngeal constrictions, before outlining the goals of the current study. Section 2 presents the method used in this study including speakers, dialects, and the corpus from which they were drawn, as well as the acoustic analyses and statistical design. Section 3 presents the results, starting with the most typically used acoustic correlates—absolute formant frequencies—before turning to the Bark-difference metrics, and acoustic correlates of voice quality. Finally, this section presents an exploratory classification technique (random forests) to evaluate primary and secondary correlates in both dialects. The last section ends with a discussion of the results and their implications for the current descriptions of pharyngealization. It is hoped that these accounts will open the door to further research into the role of the laryngeal activity that is associated with pharyngealization and pharyngeals in general.

### 1.1 Correlates of pharyngealization

Pharyngealization is a secondary articulation that involves a constriction located in the pharyngeal area that causes retraction of the body and root of the tongue towards the pharyngeal wall (Laver, 1994). Catford (1977, p. 163) defines this type of articulation as *linguo-pharyngeal* whereby the root of the tongue, including the epiglottis, moves backwards to narrow the pharynx in the front-back dimension. In many, if not all, languages, a narrowing of the pharyngeal passage near the tip of the epiglottis, and a raised larynx are the direct consequences of this constriction (see e.g., Catford, 1977, p. 193; Ladefoged & Maddieson, 1996, p. 307). Although this secondary articulation is pertinent to vowels (Ladefoged & Maddieson, 1996, p. 306), in Arabic and many Semitic languages, however, it is generally associated with consonants directly influencing the surrounding vowels (Laver, 1994).

In Arabic, pharyngealization is associated with consonants that have a primary dental/alveolar place of articulation, e.g., /s<sup>ʕ</sup> t<sup>ʕ</sup> d<sup>ʕ</sup> ð<sup>ʕ</sup>/, although in some dialects this can extend to other categories, such as /q x ɣ ħ ʕ l r m n b/. According to Arabic grammarians (e.g., Sībawayh), three terms can be used to describe these consonants: *muṭbaqa* (IPA: /mut<sup>ʕ</sup>baqa/), *mustaliya* (IPA: /mustaʕlija/), and *mufahhama* (IPA: /mufaxxama/) (Bin-Muqbil, 2006; Cantineau, 1960; Jakobson, 1957/1962; Khattab et al., 2006; Lehn, 1963). The four consonants /s<sup>ʕ</sup> t<sup>ʕ</sup> d<sup>ʕ</sup> ð<sup>ʕ</sup>/ are referred to as *muṭbaqa* (IPA: /mut<sup>ʕ</sup>baqa/), which describes their articulation as being ‘covered’ or ‘lidded’ (Khattab et al., 2006); ‘spread and with a raised tongue’ (Lehn, 1963) and/or with a ‘pressed voice’ (Cantineau, 1960, p. 23). The second term, *mustaliya* (IPA: /mustaʕlija/) ‘with an elevated dorsum’ is used to describe these four consonants /s<sup>ʕ</sup> t<sup>ʕ</sup> d<sup>ʕ</sup> ð<sup>ʕ</sup>/ in addition to /q x ɣ/ (Cantineau, 1960; Jakobson, 1957/1962; Khattab et al., 2006; Lehn, 1963). The difference between

the first and second terms stems from the fact that only /s<sup>ʕ</sup> t<sup>ʕ</sup> d<sup>ʕ</sup> ð<sup>ʕ</sup>/ are considered by Arab grammarians as consonants with a ‘pressed voice’ while the others are not (Cantineau, 1960, pg. 23–24). The last term used by Arab grammarians is *mufahḥama* (IPA: /mufaxxama/) ‘thick, heavy’ consonants, which describes all the consonants with an ‘emphatic’ acoustic impression that seem to block *Imāla*, which is a process by which an /a:/ vowel is raised and produced as [e: or ε:], and concerns /s<sup>ʕ</sup> t<sup>ʕ</sup> d<sup>ʕ</sup> ð<sup>ʕ</sup> q x ɣ h ʕ l r/, some vowels and semi-consonants (Cantineau, 1960; Jakobson, 1957/1962; Khattab et al., 2006; Lehn, 1963). Given the differences between the three categories, Cantineau (1960, p. 23–24) explained that the consonants in the first group, i.e., /s<sup>ʕ</sup> t<sup>ʕ</sup> d<sup>ʕ</sup> ð<sup>ʕ</sup>/ seem to have a distinctive articulation, expressed as ‘pressed voice,’ as they were described separately from other consonants that may share some of their articulatory features. As will be seen in the next section, none of the studies reported below have attempted to explore the nature of this ‘pressed voice’ quality, hence our aim is to evaluate its acoustic consequences.

### 1.1.1 Articulatory-acoustic correlates

Based on perturbation theory and the acoustic theory of speech production (Carré & Mrayati, 1992; Chiba & Kajiyama, 1941; Fant, 1960/1971; Howard & Angus, 2009; Johnson, 2012; Mrayati et al., 1988; Stevens, 1989), articulatory to acoustic mapping can be used to evaluate the acoustic consequences of constriction location. A pharyngeal constriction causes a combined high F1 and low F2 due to the constriction being close to a node for F1 and an antinode for F2, leading to respectively a rising and a lowering of their natural frequencies. This correlates well with the quantal region located in the pharynx (Stevens, 1989). When a pharyngeal constriction is associated with roundness/lip protrusion, as correlates of pharyngealization in general, one would expect a closer proximity between F1 and F2 and a lower F3. This is especially pertinent for front vowels given that F3 is affiliated with the front cavity and roundness affects this formant primarily (Fant, 1960/1971; Lindblom & Sundberg, 1971; Stevens, 1989; Wood, 1986). Back vowels exhibit a different pattern, with a high, low, high configuration for F1, F2, and F3. The cause of this change is formant affiliations with cavities. In back vowels, F2 is affiliated with the back cavity; an increased constriction in the pharynx leads to an increase in the length of the front cavity, and subsequently an increase in F3 (Stevens, 1989). This increase in F3 can be caused by either a narrower constriction in the upper pharynx constriction (Stevens & Keyser, 1989, p. 101) or due to a tighter tongue constriction in the pharyngealized context (see e.g., Fant, 2004, p. 43; Lindblom & Sundberg, 1971, p. 1175). And finally, larynx raising, as a consequence (or cause) of pharyngealization, leads to the high, low, low configuration for F1, F2, and F3 (Moisik, 2013a; Nolan, 1983), although Sundberg and Nordström (1976) reported an increase in F2 and F3.

These predicted consequences are evaluated empirically. The exact location of the constriction responsible for producing pharyngealization in Arabic varies from a (post-)velar to a (mid-)low pharyngeal (Khattab et al., 2006). Many researchers estimate this constriction to be located towards the posterior pharyngeal wall in the vicinity of the upper pharynx near the uvula and designate it only with tongue dorsum retraction (e.g., Bin-Muqbil, 2006; Ghazeli, 1977; Hassan & Esling, 2011; Zawaydeh, 1999; Zawaydeh & de Jong, 2011). This retraction causes lowering of the second formant in the vowels surrounding pharyngealized consonants and this is believed to be the main acoustic correlate to the contrast (e.g., Al-Ani, 1970; Al-Masri & Jongman, 2004; Al-Tamimi, F. & Heselwood, 2011; Al-Tamimi, J. & Barkat-Defradas, 2003; Barkat-Defradas et al., 2003; Ghazeli, 1977; Jongman et al., 2011; Khattab et al., 2006; Laufer & Baer, 1988; Obrecht, 1968; Shahin, 1996, 1997; Zawaydeh, 1999; Zawaydeh & de Jong, 2011; Zeroual et al., 2011, *inter alia*). Given the ‘pharyngeal’ component of pharyngealization, a few studies reported a

relatively higher first formant (F1) in the pharyngealized context (e.g., Al-Tamimi, F. & Heselwood, 2011; Al-Tamimi, J. & Barkat-Defradas, 2003; Barkat-Defradas et al., 2003; Ghazeli, 1977; Jongman et al., 2011; Khattab et al., 2006; Laufer & Baer, 1988; Shahin, 1996, 1997). The observed differences on F1 were not as consistent as those obtained for F2, and this led many researchers to consider only F2 as the main acoustic correlate to pharyngealization in Arabic (e.g., McCarthy, 1994; Watson, 2007). The frequency of the third formant (F3) was reported to be overall higher in the pharyngealized context and sometimes dependent on the vowel quality; a higher F3 was reported with back vowels, e.g., /u:/ potentially reflecting an upper-pharyngeal constriction, and a lower F3 with front vowels, e.g., /i:/ potentially reflecting a mid-pharyngeal constriction, with no differences in /a:/ (e.g., Al-Tamimi, F. & Heselwood, 2011; Al-Tamimi, J., 2007a, b; Jongman et al., 2011; Norlin, 1987; Zeroual et al., 2011). Even though these variable behaviors of F3 are related to differences in constriction location, they are also affected by the slight lip-rounding/protrusion and/or sulcalization of the tongue body reported in the literature (Khattab et al., 2006).

Given the proximity between F1 and F2 observed in 'true' pharyngeals and pharyngealization, it was used as an acoustic correlate of the auditory impression of 'darkness' or 'heaviness' reported in the literature, with pharyngeals obtaining Z2-Z1 (or F2-F1 in Bark) distance below the 3.5 Z threshold for formant merging, whereas pharyngealized consonants were slightly above this threshold (close to 4.5 Z) (e.g., Al-Tamimi, F. & Heselwood, 2011; Heselwood & Al-Tamimi, F., 2011). This proximity between formants was reported in perceptual studies based on synthetic stimuli as a major correlate to the separation between pharyngealized and non-pharyngealized consonants in Arabic, either at the midpoint in Jordanian and Moroccan Arabic (Al-Tamimi, J., 2002; Al-Tamimi, J. & Barkat-Defradas, 2003; Barkat-Defradas et al., 2003) or at the onset in Moroccan Arabic (Yeou, 2001). We will be using this (and other) formant proximity measures as correlates of pharyngealization.

Although not specifically focusing on pharyngealization, work by Alwan (1986, 1989) investigated the acoustic and perceptual correlates responsible for distinguishing between uvular and pharyngeal consonants. At the onset of the vowel, pharyngeals were associated with a higher F1 and lower F2 frequencies, an increased bandwidth of the F2 (through an estimated A2-H1 value), and a smaller distance between F2-F1, whereas uvulars had a relatively stable F1 and slight raising of F2, a larger distance between F2-F1 and an increased bandwidth of F1 (through an estimated A1-H1 value). These same results were confirmed in perception in an /a:/ context, although the F2 onset and the widened F1 bandwidth were not found to contribute as much to the distinction. Bandwidth differences as estimated through harmonic differences seem to be an important correlate to the separation between pharyngeals and uvulars and thus we will be using it as a potential correlate for pharyngealization.

### 1.1.2 Epiglottal-laryngeal articulation

Although it is assumed that pharyngealization is accompanied by retraction of the tongue body, a few studies reported a constriction much lower in the vowel tract. Using various articulatory techniques, researchers have shown pharyngealized consonants to be produced with the epiglottis forming a constriction with the pharyngeal wall in the vicinity of the middle or lower pharynx that is also accompanied by tongue root/epiglottis retraction and larynx raising (e.g., Al-Tamimi, F. & Heselwood, 2011; Laufer & Baer, 1988; Zeroual & Clements, 2015; Zeroual et al., 2011). It is not clear then whether these differences are caused by dialectal and/or individual differences, or whether all

pharyngealized consonants in Arabic have tongue root retraction and larynx raising that leads to a tense/pressed voice quality. In her comparative study of tongue root activity in various languages including Arabic using factor analyses of Ghazeli's (1977) x-ray data, Hess (1998) showed that the pharyngealized set in Arabic is better characterized by an upper pharyngeal constriction. However, her results showed that the pharyngealized set (both consonants and surrounding vowels) is also accompanied by tongue root retraction (through the Laryngopharyngeal Constriction Factor), partial constriction at the pocket of the epiglottis (through the Radical Constriction Factor), raising of the larynx, and widening of the pharyngeal wall (through the Pharynx Shifting Factor) (Hess, 1998, p. 233). In fact, the combined effect of pharyngealization was discussed by Lehn (1963, pg. 29–31), who explained that in addition to the “slight retraction, lateral spreading, and concavity of the tongue and raising of its back (what has been called velarization),” emphatic consonants are associated with “faucal and pharyngeal constriction (pharyngealization),” “slight lip protrusion or rounding (labialization),” and/or “increased tension of the entire oral and pharyngeal musculature resulting in the emphatics being noticeably more fortis than the plain segments.”

More recently, the same patterns of tongue root/epiglottis retraction and larynx raising of pharyngealized consonants were observed by Al-Tamimi, F. and Heselwood (2011) on multiple speakers of Jordanian Arabic using nasoendoscopic, videofluoroscopic, and acoustic data and by Zeroual et al. (2011) and Zeroual and Clements (2015) on 2–3 speakers of Moroccan Arabic using nasoendoscopic, ultrasound, EMA, and acoustic data. However, Hassan and Esling (2011) report observing differences between pharyngeal and pharyngealized consonants while examining emphasis spread in the productions of 1 Iraqi Arabic speaker using fiberoptic nasal laryngoscope. According to the authors, ‘true’ pharyngeal consonants showed aryepiglottic sphinctering, tongue root retraction, and larynx raising as a direct consequence of a laryngeal constriction, while pharyngealized consonants only showed tongue body retraction and raising accompanied by larynx lowering. It is possible then that some dialectal and/or idiosyncratic differences are the reason for these discrepancies. Another possibility is related to the fact that the pharyngealized set is produced with retraction of the tongue dorsum and/or root, whereas pharyngeal sets have a retraction of the tongue root as an enhancement of the epilaryngeal constriction and thus both sets are different in their ‘pharyngeal’ component (see e.g., Esling, 2005; Moisik, 2013a; Sylak-Glassman, 2014b). A slight larynx raising and/or constriction in the pharyngealized vowels and consonants in Arabic compared to pharyngeals may still be observed, and indicates a difference between the two categories (Hess, 1998). The exact degree of larynx raising and/or constriction in pharyngeal and pharyngealized sets is an important point to investigate further both articulatorily and acoustically in order to shed light to the exact nature of the production of these consonants.

**Table 1** provides a summary of the main articulatory and acoustics correlates as observed in the literature, with the additional acoustic correlates as investigated in this study. It aims at providing a correlation between articulatory and acoustic correlates in order to highlight the missing acoustic correlates of the laryngeal activity.

The various accounts of the effects of pharyngealization on surrounding vowels in Arabic show that its articulatory correlates form a complex picture with retraction of part(s) of the tongue (dorsum and/or root), retraction of the epiglottis, raising and/or constriction of the larynx leading to a more tense or pressed voice quality, and lip rounding/protrusion. However, the various acoustic accounts of the effects of pharyngealization seem to be restricted to the analyses of the first three formants of the vowels surrounding these consonants with variable outcomes as to their effects. To the best of our knowledge,

**Table 1:** List of articulatory and acoustic correlates of the effects of pharyngealization on the surrounding vowels as highlighted in the literature and the supplementary correlates used in the current study.

Articulatory	Acoustic	Additional correlates
Retraction	↓ F2	↑ F3-F2
Open	↑ F1	↑ F1-F0
Narrow/compact	↓ F2-F1, ↓ F3	↓/↑ F3-F2
Roundness/lip-protrusion	↓ F1 & F2 & F3	↓ F2-F1, ↓/↑ F3-F2
Raised larynx	↑ F1, ↓ F2 & F3	Spectral slope
Pressed/tense voice	–	Spectral slope
Epilaryngeal Constriction	–	Amplitude upper harmonics

there do not seem to be any accounts of the acoustic correlates of the tense or pressed articulation, and/or raised larynx, which are usually measured in terms of articulation (although see our own research Al-Tamimi, J., 2014, 2015). The next section introduces more in-depth account of the consequences of an epilaryngeal constriction following the Laryngeal Articulator Model.

### 1.2 Epilaryngeal constriction

Both classic and more recent articulatory reports of pharyngealization and ‘true’ pharyngeals sharing the same constriction location lead to the conclusion that articulatorily speaking, pharyngealization in Arabic is produced by constricting the epilaryngeal tube. This, in turn, has a direct consequence on retracting the tongue body due to tongue root retraction, with a concomitant larynx raising and/or constriction. This is the view developed in the Laryngeal Articulator Model on the type of constriction seen in the epilarynx (Esling, 2005). This model was extensively modeled by Moisik (2013a) and typologically and formally evaluated in Sylak-Glassman (2014a). According to Moisik (2013a, p. 84), the lower vocal tract “is bounded inferiorly by the glottis and superiorly by the oropharyngeal isthmus and velo-pharyngeal port.” The epilarynx is located within the lower vocal tract above the larynx and has the ventricular folds as its lower part and the rim formed by the epiglottis and aryepiglottic folds as its upper part (Moisik et al., 2012). Pharyngeals in general (and potentially pharyngealized consonants in particular) are produced by sphincterally constricting the epilarynx through constricting the intrinsic or the extrinsic laryngeal muscles (Esling, 2005; Moisik, 2013a; Sylak-Glassman, 2014a, b); tongue retraction is caused by this constriction and is seen as a facilitator and an enhancer of the pharyngeal articulation (Esling, 2005; Moisik, 2013a; Sylak-Glassman, 2014a, b). Constriction of the hyoglossus muscle draws the tongue as a whole backward and downward and leads to retraction of the tongue root and dorsum (Moisik, 2013a, p. 372–373; Sylak-Glassman, 2014b, p. 5). Vowels produced with this configuration are ‘retracted’; they are produced by a *back and down* gesture, which partially matches the consequences we see with pharyngealization.

The constriction as seen in the laryngeal area, and particularly in the epilarynx, has direct consequences on the quality of the sounds produced; ventricular folds couple with the vocal folds and are brought down to the glottis to allow for creaky phonation to occur (Moisik, 2013a; Sylak-Glassman, 2014a, b) and when constricted they are associated with laryngealization, tense and harsh voice quality (Edmondson & Esling, 2006; Stevens, 1977). Aryepiglottic fold constriction is associated with an enhanced and clearer voice quality especially in singing due to an increased energy in the higher frequencies as well as

‘ringing’ seen in specific types of singing (Story, 2016; Titze & Story, 1997). Aryepiglottic fold constriction is observed in the production of ‘true’ pharyngeal consonants (Esling, 2005) and a few studies reported ‘trilling’ as a direct consequence of ‘true’ pharyngeal articulation (Esling, 1999; Moisik et al., 2010). The difference between the two types of constrictions may be seen as a direct correlate of the difference between ‘true’ pharyngeals and pharyngealization; the former is produced by constricting the aryepiglottic folds as a primary feature whereas the latter by constricting the ventricular folds as a secondary feature. Pharyngeals were described as having tongue root retraction and lowering, epiglottal retraction, in addition to larynx raising as one component (Esling, 2005; Hassan & Esling, 2011; Moisik, 2013a)—descriptions that seem to match the few reports of pharyngealization in Arabic (see Section 1.1.2). This seems to suggest that the whole guttural class, including pharyngeal, pharyngealized, and uvular consonants, is produced with an epilaryngeal constriction, albeit to varying degrees of stricture and phonation (Moisik, 2013a; Sylak-Glassman, 2014a, b).

### 1.2.1 Acoustic consequences

From an acoustic point of view, and given that the vowels in the vicinity of pharyngealized consonants are better described as ‘retracted’ (Sylak-Glassman, 2014a, b), we will advocate the use of the proximity between formants as an (psycho-)acoustic correlate. Bark-difference formant frequencies are a well-established method for vowel recognition, perception, and normalization (e.g., Syrdal & Gopal, 1986; Thomas & Kendall, 2007, *inter alia*). The proximity between Z1-Z0 (that is F1-f0 in Bark) correlates well with the openness dimension (i.e., [ $\pm$  HIGH]) (Fahey et al., 1996; Hoemeke & Diehl, 1994; Syrdal & Gopal, 1986; Trautmüller, 1981), with more close ([+ HIGH]) vowels showing a Z value lower than 3 Bark, and more open ([−HIGH]) vowels around 5 Bark (Syrdal & Gopal, 1986, p. 1090). Z2-Z1 correlates well with compactness of the spectrum (Sylak, 2011; Syrdal & Gopal, 1986) and is highest for front vowels, and lowest for (mid-)open back vowels (Syrdal & Gopal, 1986, p. 1090). And finally Z3-Z2 shows the highest difference for (mid-)open back and back vowels and the smallest for front vowels and correlates well with the backness/retraction of vowels (Syrdal & Gopal, 1986, p. 1090). Z3-Z2 correlates well with ‘spectral flatness,’ or divergence of spectral peaks of F2 and F3 as seen in pre-palatal /i/ to distinguish it from /y/ (Wood, 1986). Thus we also hypothesize that a large Z3-Z2 difference between pharyngealized and non-pharyngealized contexts is a by-product of formant merging observed in Z2-Z1 due to the pharyngeal constriction. Hence we expect the vowels in the vicinity of pharyngealized consonants to show a higher Z1-Z0, a lower Z2-Z1, and a higher Z3-Z2.

Voice source changes occur with an epilaryngeal constriction. Constricting the ventricular folds leads to a tense, pressed, or laryngealized voice quality with an overall lowered or flatter spectral tilt, due to the increased energy above the first harmonic (Halle & Stevens, 1969; Hanson et al., 2001; Klatt & Klatt, 1990; Laver, 1980, 1994; Moisik, 2013b; Moisik & Esling, 2010; Stevens, 1977, 1998; Sundberg & Askenfelt, 1981). When an epilaryngeal constriction is caused by an aryepiglottic fold constriction an ‘enhanced’ voice quality, especially in singing, can be seen (Moisik & Esling, 2010; Story, 2016; Titze, 2008; Titze & Story, 1997). Samlan and Kreiman (2014) described the acoustic and perceptual consequences of constricting the epilarynx at either the aryepiglottic or the ventricular folds. Their results suggest no major differences in the increase in energy in the high frequencies, but perceptually the voice qualities were different in both types of constrictions. We will be investigating spectral slope measures that are widely used as an acoustic evaluation of phonation and voice qualities which have been successfully used to distinguish non-modal from modal phonation (e.g., Garellek, 2012; Hanson et al., 2001;

Keating et al., 2015; Klatt & Klatt, 1990; Kuang & Keating, 2012; Ladefoged & Maddieson, 1996, among others).

The various metrics used refer to the differences between the harmonics closest to  $f_0$ ,  $2*f_0$ , F1, F2, and F3 and are expressed as  $H1$ ,  $H2$ ,  $A1$ ,  $A2$ , and  $A3$  respectively. Garellek (2012) and Kuang and Keating (2012) provide a comprehensive summary of the various acoustic correlates of phonation and spectral slope. Given that the tense/pressed voice quality was reported in the context of pharyngealization, we will restrict the explanations below to the creaky/tense/pressed voice quality correlates as compared to breathy voice quality. A tense/pressed voice tends to show a lower  $H1-H2$  as the main acoustic correlate (e.g., Garellek, 2012; Hanson et al., 2001; Keating et al., 2015; Klatt & Klatt, 1990; Kuang & Keating, 2012; Ladefoged & Maddieson, 1996, among others). Spectral tilt measures, i.e.,  $H1-A1$ ,  $H1-A2$ , and  $H1-A3$  seem to be directly correlated with the abruptness of vocal fold closure (Garellek, 2012; Hanson et al., 2001).  $H1-A1$  correlates well with the bandwidth of F1; as the bandwidth of F1 increases, the amplitude of the harmonic closest to F1 decreases and thus a lower  $H1-A1$  indicates a creaky/tense voice (Garellek, 2012; Hanson et al., 2001; Kuang & Keating, 2012).  $H1-A2$  and  $H1-A3$  are also correlated with creaky/tense phonation, showing lower values (Garellek, 2012; Klatt & Klatt, 1990; Kuang & Keating, 2012). Hanson & Chuang (1999) and Hanson et al. (2001) described how an abrupt closure of the glottis can yield a change in the source spectrum; a decrease in spectral tilt around F3 is observed, and hence a lowered  $H1-A3$  would be expected. If the glottis is closed off simultaneously from the front to the back along the length of the vocal folds, then an increase in spectral tilt would be seen and a higher  $H1-A3$  is obtained. This should not be seen as an indication of an increased noise, as an increased noise as seen in breathy voice and/or glottal opening will increase spectral tilt further and  $H1-A3$  will be much higher (for more detail, see e.g., Hanson & Chuang, 1999; Hanson et al., 2001; Klatt & Klatt, 1990; Stevens, 1998). Moving to the higher frequencies, the amplitude difference  $A1-A2$  is directly related to phonation types, and thus a creaky/tense voice tends to have lower  $A1-A2$  (Aralova et al., 2011; Fulop et al., 1998; Guion et al., 2004; Kang & Ko, 2012). And finally, constricting the epilarynx on its own or through retraction of the tongue root leads to an increase in the energy of the harmonics above F1. This increase can either be extensive in that the energy in  $A2$  and  $A3$  are higher than that of  $A1$ ; consequences of an extreme epilaryngeal constriction that leads to an ‘enhanced’ voice (Story, 2016). In this case, a lowered or even negative  $A1-A3$  and  $A2-A3$  is obtained due to the concentration of energy around F3, F4, and F5 as seen in the ‘singer’s’ formant (Story, 2016). However, when the epilaryngeal constriction is minimal and/or is associated with a primary pharyngeal constriction, we would expect a change in the patterns. This would cause an increase in the amplitude differences of  $A1-A3$  and  $A2-A3$ , due to the decrease in energy around F3. This again correlates well with the acoustic consequences of an abrupt closure of the vocal folds (Hanson et al., 2001).

### 1.2.2 Formal representation

As we saw above, an epilaryngeal constriction yields both a ‘lingual’ and a ‘laryngeal’ effect. This combined effect of an epilaryngeal constriction is seen as a fundamental shift from considering lingual or laryngeal constrictions separately and can be used to describe the post-velars as one set of combined articulations (for more detail on the *glottocentric vs linguocentric* view, see Moisik, 2013a; Moisik et al., 2012; Moisik & Esling, 2011; Sylak-Glassman, 2014a, b). This allowed for the introduction of a new set of laryngeal features to account for this combined effect; [ $\pm$ CET] (for [ $\pm$ CONSTRICTED EPILARYNGEAL TUBE]) (Moisik et al., 2012; Moisik & Esling, 2011) and, as an extension, [ $\pm$ CE] (for [ $\pm$ CONSTRICTED EPILARYNX]) (Sylak-Glassman, 2014a). Following the

Lower-Vocal-Tract Phonological Potentials (Moisik, 2013a), [+CET] causes surrounding sounds to be retracted, constricted and to have raised larynx voice and tense phonation; all as a combined unit (Moisik, 2013a; Moisik et al., 2012; Moisik & Esling, 2011). ‘Retracted’ is due to tongue and epiglottis retraction with a *back and down* gesture; ‘constricted’ of certain intralaryngeal muscles and ‘raised larynx’ leading to tense voice quality (Moisik et al., 2012, p. 5). The feature [+CE] is used to distinguish between pharyngeals/epiglottal and pharyngealized/epiglottalized categories with the former being assigned this feature as a primary place feature, whereas the latter receives it as a secondary feature (Sylak-Glassman, 2014a, p. 136–138). Following Moisik et al. (2012) and Moisik and Esling (2011), ‘true’ pharyngeals (and other categories) are assigned this feature as an indication of the primary constriction of the epilarynx; pharyngealized consonants are assigned [+CE] to their primary specification (Sylak-Glassman, 2014a, p. 138). Whether all or parts of these combined articulatory consequences are to be used depends on the category of sounds (Moisik et al., 2012, p. 5–6), hence pharyngealized consonants may show tongue root retraction and/or intralaryngeal muscle constriction but not larynx raising, etc. (for more detail, see e.g., Moisik & Esling, 2011, table 2, p. 1407 on pharyngeal and pharyngealized categories being producing by a glottal source, yielding a raised larynx and a [+CET] specification; Moisik et al., 2012, figure 2, p. 5, where pharyngeal and pharyngealized receive a [+CET] and a [CG] (for [CONSTRUCTED GLOTTIS]) specification).

### 1.3 Aims of the current study

This exploratory study aims at *acoustically* investigating whether an epilaryngeal constriction is associated with pharyngealization in Arabic. We take the views developed in the Laryngeal Articulator Model that an epilaryngeal constriction leads to a ‘retracted’ vowel in the pharyngeal area with both tongue backing *and* lowering that causes a tense/pressed/laryngealized voice quality due to constricting the larynx. Acoustically speaking, we expect the vowels in the pharyngealized context to show a raised F1 and Z1-Z0, a lowered F2 and Z2-Z1 as a direct consequence of ‘retraction’ as defined in the Laryngeal Articulator Model, as well as a raised Z3-Z2 signaling both backness and spectral divergence as an enhancing correlate to the already lowered Z2-Z1. F3 can potentially show a lowered value if pharyngealization is produced in the mid-low pharynx for front vowels or a raised F3 for back vowels due to the tighter constriction. Spectral slope and voice quality correlates are expected to correlate well with the tense/pressed voice quality with an overall lowered H1-H2, H1-A1, H1-A2, H1-A3, A1-A2, and with an increased energy in the high frequencies leading to an overall flatter spectrum and a relatively lowered A1-A3 and A2-A3.

## 2 Method

### 2.1 Material

#### 2.1.1 Speakers

Twenty Jordanian and Moroccan male speakers (10 of each dialect), aged 20 to 30, participated in this experiment. Jordanian Arabic speakers originated from Irbid in the north of Jordan whereas Moroccan Arabic speakers come from Mohammedia (near Casablanca). They all reported no history of articulatory or hearing disorders, and shared the same sociolinguistic background: At the time of recording, they were students at university, and all lived in the city (i.e., spoke an urban variety). Some Jordanian Arabic speakers had some knowledge of French and/or English (beginner to advanced levels), while Moroccan Arabic speakers were non-Berber, and had knowledge of French. Both groups were proficient in the standard and dialectal forms of Arabic.



**Table 2:** List of items used in the current study for both JA and MA.

	/i:/	/ɪ/	/e:/	/a:/	/e/	/ʊ/	/u:/	
JA	/d/	'di:nak	'dɪjja	'de:r	'da:r	'dɛm	'dʊbb	'du:d
		(your) religion	compensation	monastery	home	blood	bear	worms
	/dˤ/	ma'dˤi:q	'dˤɪdˤdˤ	'dˤe:f	'dˤa:q	'dˤɛbtˤ	'dˤʊħa	maˤdˤu:dˤ
		strait	against	guest	shrunk	exactly	before noon	bitten
	/i:/		/a:/	/ə/	/ʊ/	/u:/		
MA	/d/	'di:b		'da:b	dəm'li:ʒ	'dʊll	'du:da	
		wolf		melted	bracelet	humiliation	worm	
	/dˤ/	'dˤi:f		'dˤa:q	'dˤəbtˤ	'dˤʊlma	'dˤu:sˤ	
		guest		shrunk	exactly	darkness	the second	

the recordings took place via a computer assisted user interface developed specifically for this task. After a training phase, and for purposes of other experiments using this dataset, speakers were asked to produce each item as realized in the target word, in the target CV syllable and then in the target isolated vowel, while trying to keep the production of each vowel constant across realizations and having at least 0.5 sec gap between each sequence. This particular task is aimed at evaluating the role of contextual information on the degree of vowel hypo- vs hyper-articulation and in vowel perception (for more details see Al-Tamimi, J., 2007a). The words were randomly presented with five repetitions in an adapted carrier sentence using Modern Standard Arabic (MSA) script without vocalization in order not to influence the production of each speaker, and to obtain the real dialectal realization of each word. The carrier sentence had an important role here as it was used as a way to convey the meaning of the unvocalized word, and hence reduce the influence of the MSA script. In less than 2% of the time, some speakers produced a non-dialectal form; in these cases, the experimenter clicked on a button (N for no) that enabled the word to be added to the end of the set and the speaker was asked to reproduce it. In this second round of production, most speakers reproduced the words in a dialectal form. The quality of the productions was assessed by the author and in the case of Moroccan Arabic by a native speaker of the variety. Speakers were asked to produce each word, while not moving or modifying the distance to the microphone, the loudness, intonation, or their speech rate. Each sound file was then digitized directly on the same computer, with a sampling frequency of 22.05 kHz, 16-bit quantization, in mono channel using a Sony MS 907 microphone (distance 15–20 cm from the speakers' mouths). Given that the corpus used here is part of a larger study that investigated the role of dynamic specification in production and perception of vowels, the length of all the experiments (production and perception) was about 2 hours per speaker (for more details see Al-Tamimi, J., 2007a). For the current study, vowels produced in the word condition are used to evaluate the differences observed in vowel realizations as a function of the pharyngealized vs non-pharyngealized environments; the total number of words produced by the speakers for this study was 700 for Jordanian and 500 for Moroccan Arabic (henceforth JA and MA).

**2.2 Data processing and acoustic analyses**

Data were segmented manually using Praat (Boersma & Weenink, 2009). The onset/offset of a vowel corresponded to the first/last periodic pulse in the waveform and coincided with the vertical homogeneity of the first four formants in the 0 to 5000 Hz display of a

wide-band spectrogram. In the cases where a sonorant followed the vowel, intensity drop and visual inspections were used to determine the boundary position of the vowels.

### 2.2.1 Acoustic analyses

Acoustic analyses were performed automatically using a Praat script designed by the author and adapted from Al-Tamimi, J. and Khattab (2015). Before performing the analyses, measurement frame positions of the onset and midpoint of a vowel were estimated following Al-Tamimi, J. (2004, 2007b) and Al-Tamimi, J. and Khattab (2015) in order to obtain accurate measurements and reduce errors from the automatic analyses performed by Praat. The acoustic periodicity of voiced frames in vowels was first estimated by computing  $f_0$  (see below for more details) for each sound file and by performing a PointProcess (cross-correlation) analysis. Following this, the average length of a complete glottal cycle was estimated for each speaker and each sound file; this glottal cycle ranged over 8–10 ms for our male speakers. Starting off from the initial estimates of measurement times as obtained from the TextGrids (following the segmentation as described above), these were then adjusted to match the time of maximum intensity occurring within the length of an average complete glottal cycle. Following this, they were left-aligned to the original onset estimate, and centered at the original midpoint estimate. The intensity values, computed every 5 ms, were interpolated before computing the maximum; the adjustments that resulted from this process were up to 2–3 ms around the original positions. All the reported measurements are obtained at the estimated positions.

**Formant frequencies:** Formant frequencies (F1, F2, and F3) were measured at the vowel onset and midpoint of each vowel. These were obtained from a 25-ms Kaiser2 (Gaussian-like) window with a 5-ms time step and interpolation. A maximum of five formants was requested in the formant analysis using the default Burg algorithm for formant estimation with a maximum frequency of 5 kHz for male speakers. Following formant estimation, Praat's Formant track function was used to reduce the errors in automatic formant estimation. Formant frequencies were then verified manually to prevent potential errors obtained from automatic extraction. When formant tracks obtained through the initial analysis were weakened, as in the case of /*u u:*/ vowels, a second analysis was run using an LPC smoothed curve obtained from a 256-point zero-padded DFT spectrum computed from a 10-ms Kaiser2 window left-aligned at the onset or centered at the midpoint of the vowel, after down-sampling the sound file to 10 kHz, low-pass filtering with an anti-aliasing filter with a cut-off frequency of 5 kHz, and pre-emphasizing it by a factor of 0.98. Both the FFT and the smoothed LPC displays were used to estimate the position of a particularly weakened formant.

**Bark-difference formant frequencies:** Once the first three formant frequencies were estimated, these were converted to the psychoacoustic Bark scale following Traunmüller's (1990) formula 1, where  $Z_n$  is the Z value (i.e., critical bandwidth) of the formant<sub>*n*</sub>, and  $f_n$  is the frequency in Hz of the formant<sub>*n*</sub> (including  $f_0$  in both cases), and any Z values lower than 2 Bark, were corrected using formula 2, where Z reflects the original Z value and  $Z_c$  is the corrected one:

$$Z_n = \left\{ 26.81 / (1 + 1960 / f_n) \right\} - 0.53 \quad (1)$$

$$Z < 2.0 \text{Bark} : Z_c = Z + 0.15(2 - Z) \quad (2)$$

The fundamental frequency  $f_0$  was also estimated and only used in Bark-difference results. Overall, the results showed non-significant changes according to the consonant and vowel

quality; there was however a pattern of lower  $f_0$  in JA and higher in MA. The estimation followed the procedure in Al-Tamimi, J. and Khattab (2015) and used the two-pass method following (Hirst, 2011). For the first pass, Praat's default settings were used: 5-ms step, 40-ms Kaiser2 window, floor and ceiling = 75–500 Hz respectively. This allowed to obtain the pitch contour for each speaker. The first and third quartiles (i.e., 25–75% values) were obtained for each speaker and these were then multiplied by a coefficient; 0.75 and 1.5 respectively. These were then used as the new floor and ceiling in the second pass and the actual  $f_0$  computations was done using a 5-ms step and an effective Gaussian window length of 30 ms. The floor and ceiling values for speakers ranged between 75–100 and 150–200 Hz respectively. The frequencies obtained in Hz were converted to the Bark scale using formulas 1 and 2. Once the Bark transformation applied on F1, F2, F3, and  $f_0$  in both positions of the vowel, the Bark-difference Z values were then computed to reflect Z1-Z0, Z2-Z1, and Z3-Z2 at the onset and at the midpoint of the vowel.

**Voice quality:** To obtain voice quality measures at both onset and midpoint of the vowel, the sound files were low-pass filtered with an anti-aliasing filter which had a cut-off frequency of 5 kHz, down-sampled to 10 kHz, and pre-emphasized by a factor of 0.98. Intervals 40 ms long were defined to allow for at least 4 to 5 complete glottal cycles to be used in estimating spectra of vowels. From each sound file, one interval was left-aligned at the onset and a second centered at the midpoint of the vowel; these were then windowed using a Kaiser2 window function. From each windowed interval, a 256-point zero-padded DFT spectrum was computed and the logarithmic power spectral density, with a bin size of 19 Hz, was computed. Following Al-Tamimi, J. and Khattab (2015), the amplitudes of the first and second harmonics and of the first to the third formants were automatically obtained by detecting the highest peaks for a particular harmonic; maximum amplitude was obtained from  $f_0 \cdot 0.9$  to  $f_0 \cdot 1.1$  and from  $2 \cdot f_0 \cdot 0.95$  to  $2 \cdot f_0 \cdot 1.05$  for  $H1$  and  $H2$ , respectively. For the amplitude of the harmonics closest to the three first formants, we estimated the bandwidth frequencies using the formula proposed by Hawks and Miller (1995) instead of using the automatically estimated ones in Praat due to many errors that prevented manual correction. Then maximum amplitudes were obtained in the region from  $F1 - 0.5 \cdot \text{Bandwidth1}$  to  $F1 + 0.5 \cdot \text{Bandwidth1}$  for  $A1$ . The same procedure was applied for  $A2$  and  $A3$  (and using Bandwidths 2 and 3 respectively). The automatic detection of formant frequencies, and highest peaks were manually checked to prevent errors. Instead of using inverse-filtering to estimate these harmonics, we relied on the normalization procedure as developed by Iseli et al. (2007) and implemented in our Praat script to obtain the 'corrected' versions of these harmonics that removes the boosting effects of surrounding formants.  $H1$ ,  $H2$ ,  $A1$ , and  $A2$  were normalized by correcting for the effects of  $F1$  and  $F2$ , whereas  $A3$  was normalized by correcting for the effects of  $F1$ ,  $F2$ , and  $F3$ . Then amplitude differences were obtained for  $H1 \cdot H2^*$ ,  $H1 \cdot A1^*$ ,  $H1 \cdot A2^*$ ,  $H1 \cdot A3^*$ ,  $A1 \cdot A2^*$ ,  $A1 \cdot A3^*$ , and  $A2 \cdot A3^*$  at both onset and midpoint of the vowel.

### 2.2.2 Statistical analyses

A total of 30,966 measurements (17,992 in JA and 12,974 in MA) were obtained from the 13 different acoustic correlates summarized above (Section 2.2.1), measured at both onset and midpoint (i.e., a total of 26 measures). Our main aim in this study is to evaluate the degree to which a particular acoustic correlate can be used to successfully predict the difference between the pharyngealized and non-pharyngealized consonants, hence, we adopted a predictive approach (Baguley, 2012; Hastie et al., 2009; Kuhn & Johnson, 2013). First, we started by using Generalized Linear Mixed-Effect Modeling (GLMM) (Quené

& Van den Bergh, 2008) in a confirmatory way that was followed by an exploratory classification technique, namely Conditional Random Forests that allowed us to evaluate which acoustic correlate(s) are the most predictive of the two consonant categories. All analyses were run using the statistical software R (version 3.3.1) (R Core Team, 2016).

#### *Generalized Linear Mixed-Effects Modeling (GLMM)*

Before running the GLMMs, we started by examining the correlation levels in our data. We used correlation matrices ordered by hierarchical clustering obtained with the function `hclust` from the package `reshape2` (Wickham, 2007), and generated via a modified R script (Melike, 2016). These correlation matrices are presented in Appendix 1 for JA and in Appendix 2 for MA. The correlation matrices showed that out of the total combinations of 650 correlations in each dialect, 138 turned out to be significant in JA and 172 in MA. Correlation coefficients are judged as statistically significant at  $p < 0.01$ , i.e., any absolute  $r^2$  value  $> 0.16$  was significant (correlation and  $p$  values obtained with the function `rcorr` from the package `Hmisc` (Harrell Jr, 2016)). In both dialects, voice quality and formant-based measures are negatively correlated with each other, and all formant measures (being absolute or Bark-difference) are positively correlated with each other (except in JA, where Z3-Z2 is negatively correlated with Z2 and Z2-Z1). Given the recommendations in Baayen (2008, p. 181–183), it was not advisable to use all these measures together in a regression analysis as they will be giving either the same outcome (in the case of positively correlated ones) or cancel each other out (for negatively correlated cases). Hence we decided to use separate regression analyses on each of the individual acoustic correlates and compared these via predicted probabilities. In addition, due to the high predictive outcome of some acoustic correlates (e.g., F2, Z3-Z2, see below), it was not possible to use one GLMM model combining all the acoustic correlates, as these were canceling each other out, resulting in model non-convergence.

Due to these constraints (multicollinearity and high predictive power), we used an alternative approach. Normally distributed acoustic correlates were used (by transforming the absolute formant frequencies to the Bark scale; Bark-difference and spectral slope measures are already on logarithmic scales). We obtained the descriptive statistics using the package `psycholing` (Fraundorf, 2015), with the means and standard deviations for the original and  $z$ -scored values. The  $z$ -scored values were computed to a mean of 0 and a standard deviation of 1, separately for each dialect (using the means and standard deviations presented in column “All” in Table A3 in Appendix 3 that provides a summary of the combined contexts in JA and MA and for each of /d/ and /d<sup>f</sup>/ contexts, with the mean and standard deviations in the original and the  $z$ -scored scale. We also included the differences in each of the acoustic cues between /d<sup>f</sup>/ and /d/ to allow for comparability of the outcome with the results of the GLMMs).

Following this, and to avoid multicollinearity, we ran separate GLMMs on the individual  $z$ -scored acoustic correlates following Schielzeth (2010) as a simple and meaningful way to compare individual models and to assess their effect sizes. We started by running GLMMs with the ‘consonant’ as a binomial response category (dummy coded with /d/ = 0 and /d<sup>f</sup>/ = 1), and the separate acoustic correlates as predictors (i.e., a Mixed-Effect Logistic Regression). The dummy coding of consonant is a requirement of the GLMM and allows for a meaningful interpretation of the results with the  $\beta$  of the `Intercept` representing the average in the /d/ environment whereas the predictor’s  $\beta$  is that in the /d<sup>f</sup>/ environment. Vowels (contrast coded) and repetitions (centered) were added as additional predictors; however, they did not improve the fit of the GLMM (through likelihood ratio comparison), thus these were dropped from the final models.

For the random part of the model, we used speakers and items as crossed random factors (Baayen et al., 2008). Within item variation with respect to repetitions was included to allow for this random variation to be taken into account. Both by-speaker and by-item random slopes were used following a maximal specification model (Barr et al., 2013). These allowed both speakers and words to vary with respect to the within variation that is due to the acoustic correlates. By-item random slopes were necessary given that we only had one item per consonant and vowel combination; this allowed the acoustic correlates to vary within each word, otherwise the model would inaccurately overestimate this variation. In many occasions, random intercepts for words were not successful at providing a clear picture of the results, and hence were dropped, i.e., only a by-item random slope was used.

To prevent quasi- and complete separation of the data that was obtained with `lme4` package (Bates et al., 2015), we used a Bayesian Generalized Linear Mixed-Effects Models as implemented in the package `blme` (Chung et al., 2013). The function `bglmer` requires defining a prior for both the fixed and the random effects; we used a multivariate normal distribution with a diagonal variance-covariance matrix for the fixed effect prior, i.e., `fixef.prior = normal(cov = diag(2.5, 2))`, with 2 being the 2 fixed-effect parameters and 2.5 representing a variance of 2.5 which is equivalent to a 1.58 SD that is close to our actual SD for the z-scored data (being 1SD) (following the recommendations of Gelman et al., 2008, p. 1366 for numerical predictors). The random effect prior was kept at its default, (i.e., `cov.prior = wishart`, for more detail, see Chung et al., 2013).

#### *Random Forests via Conditional Inference Trees*

Following the individual GLMMs, we wanted to use a classification technique as a predictive model (Hastie et al., 2009; Kuhn & Johnson, 2013). Random Forests are one of the most versatile machine learning algorithms as they do not require many tunings of their settings. They have been applied to sociolinguistic data (for a detailed description of these methods, see Tagliamonte & Baayen, 2012), and also to acoustic cue weighting in perception (Brown et al., 2014). Random Forests were originally proposed by Breiman (2001) as an ensemble learning algorithm that uses *independent* classification and regression trees in growing a forest. The independence stems from the randomness of the selection process in the algorithm: By randomly selecting a subset of the data and of the predictors, a decision tree is constructed. This is then repeated several times (for the total number of trees in a forest), and a majority of votes is then taken for prediction (Liaw & Wiener, 2002). After the forest is grown, one can estimate the prediction accuracy as well as the ranking of the most important predictors. This model can also be used to predict the outcome on new unseen data, e.g., newly collected production/perception data or on the testing set (*out-of-bag* set). The algorithm splits the data into learning and testing sets; the former constitute around 66.7% of the data and named as an *in-bag* set, while the latter constitutes the remaining data of 33.3% and named as an *out-of-bag* set. Then subsampling without replacement is used in growing the forest (for more details, see Strobl et al., 2009).

Instead of using the original implementation of Random Forests available in the R package `randomForest` (Liaw & Wiener, 2002), we used Random Forests grown from Conditional Inference Trees as implemented in the package `party` (Hothorn et al., 2006; Strobl et al., 2008, 2007). Strobl et al. (2008, 2007) found that the original `randomForest` provided biased estimates of Random Forests' variable importance (see below) as it was biased towards variables with multiple categories and multiple cut-points, and also overestimated variable importance measures when correlated data is used (as is the case in our study). To guard against this bias, they developed an unbiased selection

process (subsampling without replacement). This form of random forests, i.e., based on conditional inference trees, are well suited to deal with collinear variables (Strobl et al., 2008, 2007; Tagliamonte & Baayen, 2012), with “small  $n$  large  $p$ ” case, where the number of predictor variables  $p$  is greater than the number of subjects  $n$ ” (Strobl et al., 2009, p 323) and they allow to evaluate the importance of the predictors used in classification by ranking them after controlling for interactions and collinearity.

Random Forests were run using the `party` package with the function `cforest` on the combined 26 different acoustic correlates for each dialect separately and by using the recommended `cforest_unbiased` control with `mtry = 5` (rounded square root of 26 predictors). Number of trees `ntree` can be tuned to allow for less computation time, and we followed the density-based metrics developed by (Oshiro et al., 2012) to estimate the density of our dataset using the formula 3, with  $a$  = number of predictors,  $n$  = number of observations and  $c$  = number of classes.

$$Density = \log_a \{n / c\} \quad (3)$$

For JA and MA, the density-based metrics were equal to 1.79 and 1.69 respectively. According to Oshiro et al. (2012), these values constitute a low density database. The authors found that a large number of trees close to and above 2000 is only needed with high density datasets that exceed a density-based value of 3. A low number of trees (`ntree = 128` trees) was needed for the majority of their datasets to achieve a high predictive accuracy. Hence, we implemented the same procedure as that suggested by Oshiro et al. (2012). We started by building 15 different random forests for each dialect by using `ntree` from 100 to 1500 in a 100 trees increment. Then for all generated random forests, we checked their predictive power, by using the function `predict` using the *out-of-bag* set as a cross-validation (using `OOB = TRUE`). Then we used an AUC (for Area Under the Curve) based comparison using the package `pROC` (Robin et al., 2011), by generating an ROC curve (for Receiver Operating Characteristics) and then by performing a non-parametric Z test of significance on the correlated ROC curves using the function `roc.test`, following DeLong et al. (1988). The results of this comparison showed that for JA, 400 trees were enough to reach the highest predictive accuracy, whereas for MA, 300 trees were enough. Hence we ran random forests with these `ntree` values.

Then we used an AUC-based estimation of the variable importance, `varimpAUC` as it takes into account both accuracy and error of estimation and used conditional permutation tests with `conditional = T` (for more detail, see specification in Strobl et al., 2008, 2009). We then evaluated how well correlated are the random forest results with the actual data, and the percentage correct classification by using an *out-of-bag* cross-validation.

After this initial random forest using all of the acoustic correlates, we ran six additional exploratory random forest analyses that will be used for their predictive accuracy. These aimed at evaluating whether the predictive accuracy would be different depending on the type of acoustic correlates used. We compared formant-based metrics only, i.e., absolute + Bark-difference formants, absolute formants, or Bark-difference formants only). The aim here is to evaluate the strength of these metrics at separating the two categories. This was followed by the combination of absolute formants or Bark-difference formants in addition to voice quality, in order to assess whether there is an increase/decrease in predictive power when voice quality is used. And finally, we assessed the predictive power of the voice quality metrics on their own, in order to evaluate whether any observed statistical difference are important enough to allow for discriminating the two consonantal categories. In each case, we used the above procedure to estimate the optimal number of

trees needed to obtain the highest predictive accuracy, and adapted `mtry` to each case. The next section presents the results of this study.

### 3 Results

#### 3.1 GLMM

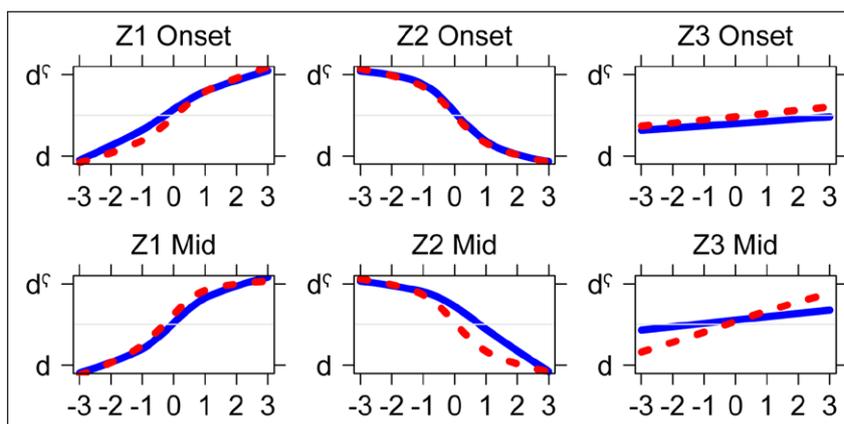
We started by running two GLMM analyses for each acoustic correlate: One including the acoustic correlate as a predictor (i.e., a Full model) and a second without the acoustic correlate (i.e., a Null model). Then we proceeded with model comparisons using likelihood ratio tests to derive all  $p$  values (Barr et al., 2013) and to test whether a particular acoustic correlate is significantly contributing to the difference between the two consonants. Out of the 26 different acoustic correlates used in this study, 16 made a significant contribution to the difference in JA, compared to 20 in MA. These are highlighted in bold in **Table 3** (as a reference, complete descriptive statistics of the results are presented in the Table A3 in Appendix 3). **Figures 1, 2, 5, and 6** were generated using the predicted probabilities of the individual acoustic correlates as obtained from the Full GLMM. These predicted probabilities were obtained from the `predict` function in `blme` based on a new dataset with the range between  $-3$  and  $3$   $z$ -score with a  $0.1$  increment (using a modified version of an R script developed by Winter, 2016). These were then plotted the following packages: `lattice` (Sarkar, 2008), `latticeExtra` (Sarkar & Andrews, 2016) and `gridExtra` (Auguie, 2016). Using this range allows for a meaningful comparison between all the measures. When looking at the predicted probability curves, we will refer to a sigmoidal shape or a flat shape; the former represents the logistic function curve and the latter a linear shape. A sigmoidal curve indicates a high level of separation between the groups and hence will have a high predictive function. To interpret the effect size associated with each acoustic correlate, the absolute  $\beta$  value will be used to report the standardized effect size; the negative or positive signs are indicative of falling or raising probability curves from  $/d^{\text{f}}/$  to  $/d/$ . As an example, **Figure 1** shows the predicted probabilities for Z1, Z2, and Z3 at both onset and midpoint of the vowel in JA (blue solid) and MA (red dashed). In both dialects, Z1 at the onset of the vowel shows raising curves from  $/d/$  (values starting from  $-3$   $z$ -score) to  $/d^{\text{f}}/$  (values ending with  $+3$   $z$ -score); Z2 shows the reverse pattern, i.e., falling curves from  $/d^{\text{f}}/$  (values starting from  $-3$   $z$ -score) to  $/d/$  (values ending with  $+3$   $z$ -score); and Z3 shows an overall flat to raising curves from  $/d/$  (values starting from  $-3$   $z$ -score) to  $/d^{\text{f}}/$  (values ending with  $+3$   $z$ -score), although the shape is not as sigmoidal as that seen for Z1 nor Z2. We have also provided the percent correct classification of each GLMM based on the  $\beta$  value by using the formula:  $\% = \exp|\beta| / (1 + \exp|\beta|)$ .

##### 3.1.1 Absolute Formants

From model comparison results (see **Table 3**), it can be seen that F1 and F2 at the onset and midpoint are, as expected, reliably associated with the differences between pharyngealized and non-pharyngealized consonants. In MA, F3 at the midpoint only showed a tendency towards being a correlate to the difference. GLMM results confirm the outcome as observed in the literature; F1 shows an increased  $\beta$  values whereas F2 shows a lowering in  $/d^{\text{f}}/$ . F3 showed an overall higher values in  $/d^{\text{f}}/$ , albeit non-significant. The predicted probabilities as displayed in **Figure 1** show a clear pattern of rise, fall, rise patterns for F1, F2, and F3; with F3 showing a slight rise rather than a flat line. Given that formant frequencies were  $z$ -scored, it is possible to compare the strength of the effect as these are standardized effect size measures (Schielzeth, 2010). In JA, F2 at the onset has the highest absolute  $\beta$  value and shows more of a sigmoidal curve (see Z2 Onset, **Figure 1**). In MA, however, F1 mid and F2 at both onset and mid have the highest absolute  $\beta$  values and show a near complete sigmoidal curve. When converting the  $\beta$  values to percent correct, all four acoustic correlates (in JA

**Table 3:** Summary of statistical results, with model comparison between Null and Full models, with  $\chi^2$  (1 df), and *p* values, followed by the GLMM results of the Full model with the estimates of the Fixed effect of each acoustic measure (with  $\beta$ , SE (Standard Error), Wald's *z* value, associated *p* value, and % correct classification based on  $\beta$  values); significant results are in bold.

	MA														
	JA					GLMM results									
	Model comparison		GLMM results			Model comparison		GLMM results							
	$\chi^2$ (1)	<i>p</i>	$\beta$	SE	<i>z</i>	Pr (>  <i>z</i>  )	%	$\chi^2$ (1)	<i>p</i>	$\beta$	SE	<i>z</i>	Pr (>  <i>z</i>  )	%	
Z1	Onset	8.07	<0.005	1.28	0.47	2.71	<0.007	78.26	24.2	<0.00001	2.05	0.32	6.51	<0.00001	88.63
	Mid	8.18	<0.005	2.39	0.97	2.46	<0.02	91.62	25.51	<0.00001	4.54	1.04	4.38	<0.00001	98.95
Z2	Onset	27.38	<0.00001	-6.92	0.92	-7.51	<0.00001	99.90	30.88	<0.00001	-4.59	0.64	-7.23	<0.00001	99.00
	Mid	10.303	<0.001	-1.55	0.77	-2.00	<0.05	82.43	2.9403	=0.086	-4.79	0.67	-7.17	<0.00001	99.17
Z3	Onset	0.27	=0.61	0.12	0.29	0.40	=0.70	52.91	0.75	=0.39	0.16	0.22	0.74	=0.47	53.96
	Mid	0.46	=0.50	0.17	0.24	0.69	=0.49	54.19	3.26	=0.071	0.56	0.32	1.74	=0.082	63.66
Z1-Z0	Onset	12.38	<0.0001	1.36	0.38	3.59	<0.0001	79.55	20.58	<0.00001	1.85	0.31	6.03	<0.00001	86.38
	Mid	6.65	<0.01	2.07	0.94	2.20	<0.05	88.78	27.61	<0.00001	4.50	0.97	4.63	<0.00001	98.90
Z2-Z1	Onset	26.15	<0.00001	-6.09	0.90	-6.74	<0.00001	99.77	30.1	<0.00001	-4.65	0.56	-8.23	<0.00001	99.05
	Mid	39.622	<0.00001	-3.77	1.70	-2.22	<0.05	97.75	13.327	<0.0005	-2.58	0.85	-3.04	<0.002	92.95
Z3-Z2	Onset	22.94	<0.00001	6.35	1.03	6.15	<0.00001	99.83	28.37	<0.00001	4.38	0.58	7.58	<0.00001	98.76
	Mid	8.5688	<0.005	1.55	0.56	2.76	<0.006	82.46	22.109	<0.00001	4.75	1.07	4.45	<0.00001	99.15
H1*-H2*	Onset	0.29	=0.60	0.09	0.19	0.49	=0.62	52.32	0.06	=0.81	0.05	0.18	0.26	=0.42	51.15
	Mid	0.03	=0.87	0.05	0.19	0.28	=0.78	51.34	0.88	=0.35	0.14	0.17	0.81	=0.80	53.52
H1*-A1*	Onset	2.56	=0.11	-0.41	0.27	-1.49	=0.14	60.02	22.46	<0.00001	-1.51	0.23	-6.41	<0.00001	81.83
	Mid	8.27	<0.005	-0.58	0.21	-2.74	<0.006	64.07	13.68	<0.0001	-1.33	0.34	-3.93	<0.0001	79.10
H1*-A2*	Onset	10.35	<0.001	-0.76	0.23	-3.34	<0.001	68.24	19.74	<0.00001	-1.98	0.33	-6.05	<0.00001	87.82
	Mid	5.76	<0.02	-0.61	0.26	-2.29	<0.05	64.73	17.63	<0.00001	-1.46	0.32	-4.60	<0.00001	81.14
H1*-A3*	Onset	3.18	=0.074	0.25	0.15	1.65	=0.098	56.31	4.03	<0.05	-0.32	0.18	-1.82	=0.068	58.00
	Mid	2.07	=0.15	0.05	0.19	0.28	=0.17	51.34	1.93	=0.17	-0.27	0.22	-1.25	=0.21	56.69
A1*-A2*	Onset	2.54	=0.11	-0.42	0.28	-1.51	=0.13	60.32	5.74	<0.02	-0.69	0.28	-2.44	<0.02	66.56
	Mid	0.79	=0.37	-0.16	0.19	-0.80	=0.42	53.88	13.18	<0.0001	-0.80	0.21	-3.77	<0.0002	68.97
A1*-A3*	Onset	11.04	<0.001	0.77	0.22	3.55	<0.0005	68.36	2.23	=0.14	0.33	0.24	1.39	=0.17	58.24
	Mid	0.17	=0.68	0.09	0.21	0.43	=0.67	52.29	1.16	=0.28	0.30	0.29	1.04	=0.30	57.37
A2*-A3*	Onset	16.66	<0.00001	1.23	0.24	5.07	<0.00001	77.40	11.39	<0.001	1.01	0.27	3.74	<0.0002	73.37
	Mid	1.72	=0.19	0.41	0.34	1.20	=0.23	60.17	4.01	<0.05	0.96	0.52	1.85	=0.065	72.28



**Figure 1:** Predicted probability curves spanning  $-3$  to  $+3$  z-score for absolute formant frequencies in both JA (blue solid) and MA (red dashed). Falling curves indicate a lowering in the predicted probabilities from  $/dʰ/$  to  $/d/$  (as in Z2) and vice versa. The reference line (in gray) indicates the non difference line.

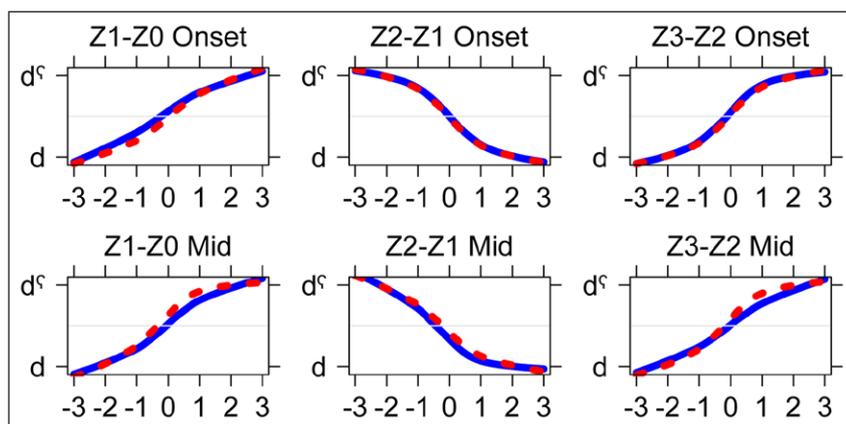
and MA) show rates between 78–99%, with Z2 showing highest values. Overall, the results obtained for the absolute formant frequencies are comparable in direction and range to the previous literature (see Table A3 in Appendix 3 and Section 1.1.1).

Following perturbation theory and the acoustic theory of speech production (Carré & Mrayati, 1992; Chiba & Kajiyama, 1941; Fant, 1960/1971; Howard & Angus, 2009; Johnson, 2012; Mrayati et al., 1988; Stevens, 1989), the location of the constriction responsible for producing pharyngealization is predicted to be located in the pharyngeal area, potentially between the upper to mid pharynx, as is generally reported in the literature. F3 frequencies, albeit non-significant, show a tendency for an overall raising pattern from  $/d/$  to  $/dʰ/$ ; however, our results are consistent with a lowered F3 in front vowels (by around 70 Hz) and a raised F3 in back vowels (by around 200 Hz) from  $/d/$  to  $/dʰ/$ ; patterns that are observed in our previous studies (Al-Tamimi, J., 2007a, b, 2009) (see also Section 2.2.1).

### 3.1.2 Bark-difference formants

Bark-difference formant frequencies are used as a psychoacoustic representation of the patterns observed in formants and is claimed to reflect the openness, retraction, and spectral divergence of vowels (Fahey et al., 1996; Hoemeke & Diehl, 1994; Sylak, 2011; Syrdal & Gopal, 1986; Traunmüller, 1981; Wood, 1986). Model comparison results summarized in **Table 3** show that all Bark-difference metrics are significantly associated with the difference between the two categories. Specifically, in the  $/dʰ/$  context, Z1-Z0 and Z3-Z2 show increased  $\beta$  values, whereas Z2-Z1 shows lowered ones. The predicted probabilities presented in **Figure 2** show a near perfect sigmoidal curve for all Bark-difference metrics, with variable degrees of separation. Generally, MA shows higher  $\beta$  values for Z1-Z0 and an increased sigmoidal curve than in JA, reflecting a more open production for the vowels (see absolute formant frequencies results in **Figure 1** above).

Z2-Z1 reflects compaction of the spectrum with more back and open vowel productions being associated with a lowered Z2-Z1. A low pharyngeal constriction will show Z2-Z1 values below 3 Bark values whereas a mid to upper pharyngeal constriction will show higher Z2-Z1 values around 4.6 Bark (Al-Tamimi, F. & Heselwood, 2011; Heselwood & Al-Tamimi, F., 2011). Our results show a lowering of Z2-Z1 values that is close to the latter figure, with JA being close to 5.5 Bark and MA around 6.2 Bark in the  $/dʰ/$  context



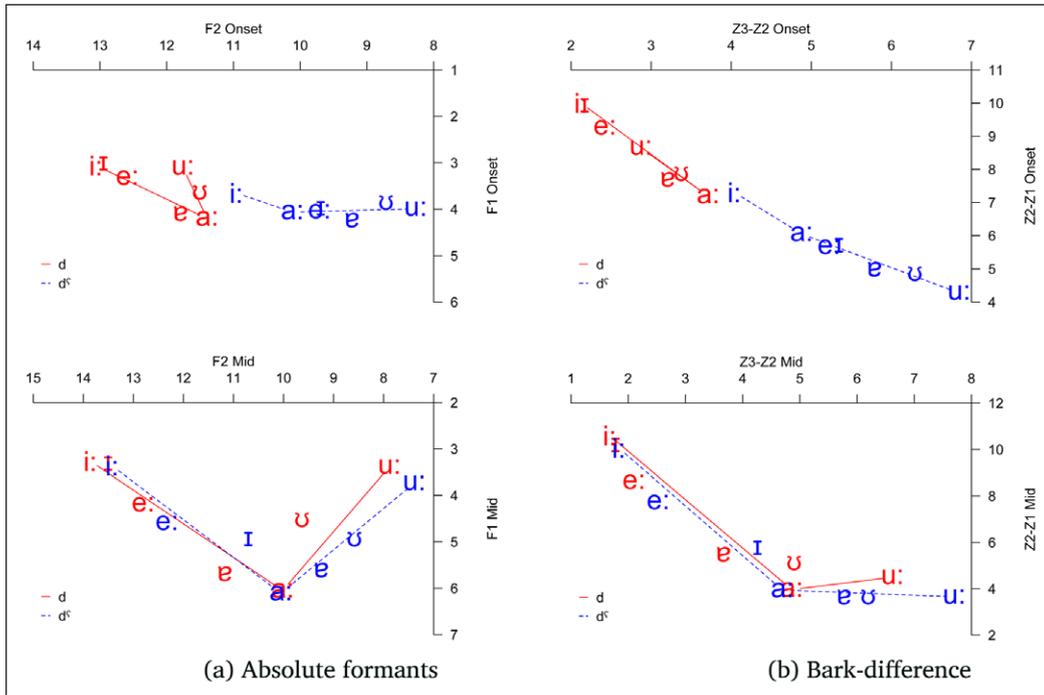
**Figure 2:** Predicted probability curves spanning  $-3$  to  $+3$  z-score for Bark-difference formant frequencies in both JA (blue solid) and MA (red dashed). Falling curves indicate a lowering in the predicted probabilities from  $/d^s/$  to  $/d/$  (as in Z2-Z1) and vice versa. The reference line (in gray) indicates the non difference line.

(see Table A3 in Appendix 3). And finally, Z3-Z2 values show large  $\beta$  values at the Onset for JA and at both positions in MA. They show a more retracted vowel quality that is close to that of the  $/o \text{ } \text{ɔ}/$  vowels (when compared with the results in Syrdal & Gopal, 1986, p. 1090). This large difference in the pharyngealized contexts is mostly due to the differences observed in the F2; an increased distance between Z3-Z2 is caused by the lowered F2 in the  $/d^s/$  rather than that of F3. This increased distance can also be correlated with spectral-flatness or divergence (Wood, 1986), and hence serves as an enhancing correlate to the already lowered Z2-Z1. The predictive strength of these correlates is variable with rates between 78–98% (see Table 3 for more detail).

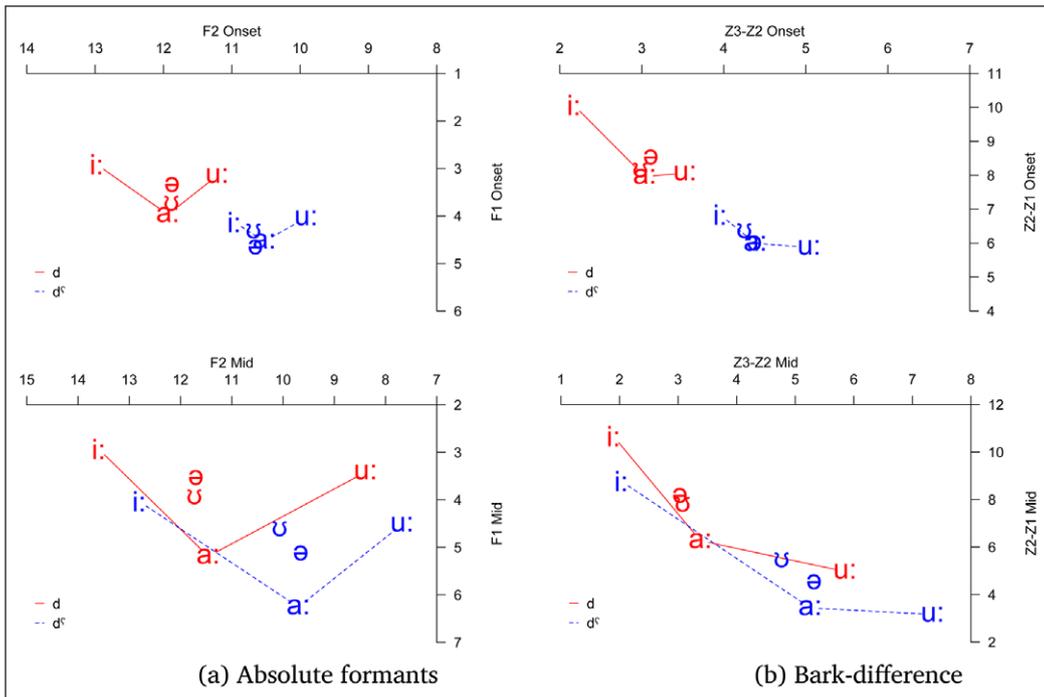
### 3.1.3 Vowel spaces

We used relational formant frequencies to display vowel spaces; using the traditional F2\*F1 (in Bark) vowel spaces and that of Z3-Z2\*Z2-Z1 at both onset and midpoint of the vowels in JA and MA for both of  $/d/$  (in red solid) and  $/d^s/$  (in blue dashed) contexts. Figures 3 and 4 show both Z2\*Z1 (a) and Z3-Z2\*Z2-Z1 (b) at onset (top) and midpoint (bottom) in both JA and MA respectively (charts generated with the package phonR, McCloy (2016)). Both figures are an attempt to provide a better understanding of how Bark-difference representation is well suited to account for the ‘retracted’ vowel quality as defined in the Laryngeal Articulator Model (Esling, 2005; Moisik et al., 2012; Sylak-Glassman, 2014b).

Vowel spaces at the onset provide the clearest separation between the two categories compared with that at the midpoint, although in MA, the midpoint results lend support to a more open configuration (see Figure 4). At the onset, JA shows a clear back vowel quality as represented with absolute F2, whereas MA shows a clear open and back vowel articulation through absolute F1 and F2 respectively (see Figures 3a and 4a). Moving on to the Bark-difference results (see Figures 3b and 4b), and particularly at the onset (top), both JA and MA show a clear separation between the two categories, with vowels in the pharyngealized context showing a ‘compact’ (with lower Z2-Z1) and ‘backed’ (with higher Z3-Z2) production; they are ‘retracted’ in the sense of the Laryngeal Articulator Model (Esling, 2005; Moisik et al., 2012; Sylak-Glassman, 2014b). By using absolute formants only (Figures 3a and 4a), vowels in JA are not represented with an open configuration as all vowels seem to range around 4 Bark. Our claim is then if one is to acoustically evaluate



**Figure 3:** Absolute formant (Z1\*Z2; **a**) Versus Bark-difference (Z2-Z1\*Z3-Z2; **b**) Vowel spaces at onset (top) and midpoint (bottom) of /d/ (red solid) and /dʕ/ (blue dashed) in JA.



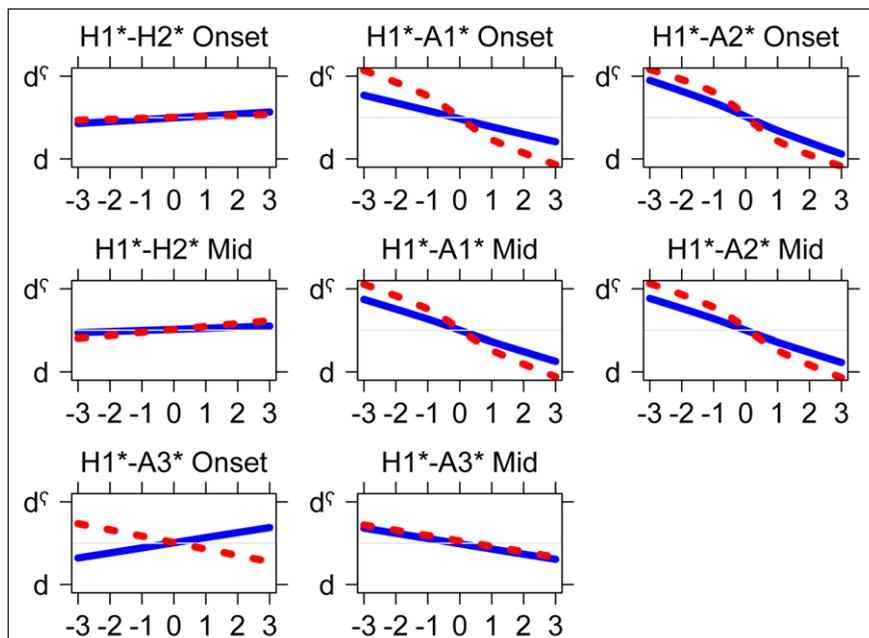
**Figure 4:** Absolute formant (Z1\*Z2; **a**) Versus Bark-difference (Z2-Z1\*Z3-Z2; **b**) Vowel spaces at onset (top) and midpoint (bottom) of /d/ (red solid) and /dʕ/ (blue dashed) in MA.

the predictions of the Laryngeal Articulator Model, absolute formant frequencies alone are not useful at showing the ‘retraction’ that has a combined *back and down* gesture; Bark-difference vowel spaces are better suited.

3.1.4 Voice quality

To our knowledge, this is the first attempt to evaluate any voice quality differences as associated with pharyngealization in Arabic (although see Alwan, 1986, 1989, on the estimated bandwidth in pharyngeals). Our voice quality results are separated into spectral slope and high frequency energy (Figures 5 and 6 respectively). Starting with spectral slope, model comparison results suggest that  $H1^*-A1^*$ ,  $H1^*-A2^*$ , and  $H1^*-A3^*$ , in either both or one position are significantly contributing to the difference between the two consonants in both dialects (Table 3). GLMM results show an overall decrease in  $\beta$  values in all significant metrics (except for a tendency to observe an increase in  $H1^*-A3^*$  Onset in JA). MA results show a stronger contribution of these metrics to the distinction between the consonants compared to JA. This is shown further in the curve shape in Figure 5 where a stronger declination from /d<sup>ɕ</sup>/ to /d/ can be seen in MA for all significant correlates;  $H1^*-A1^*$ ,  $H1^*-A2^*$  at both onset and midpoint show a near sigmoidal curve, although this is not comparable to those seen for absolute and Bark-difference formants (see above). In JA, the curves seem more linear in significant correlates, with lower  $\beta$  and a predictive strength varying between 56–82% (Table 3). These spectral slope measures seem to act as secondary correlates to pharyngealization as they do not show the same predictive power as formant frequencies.

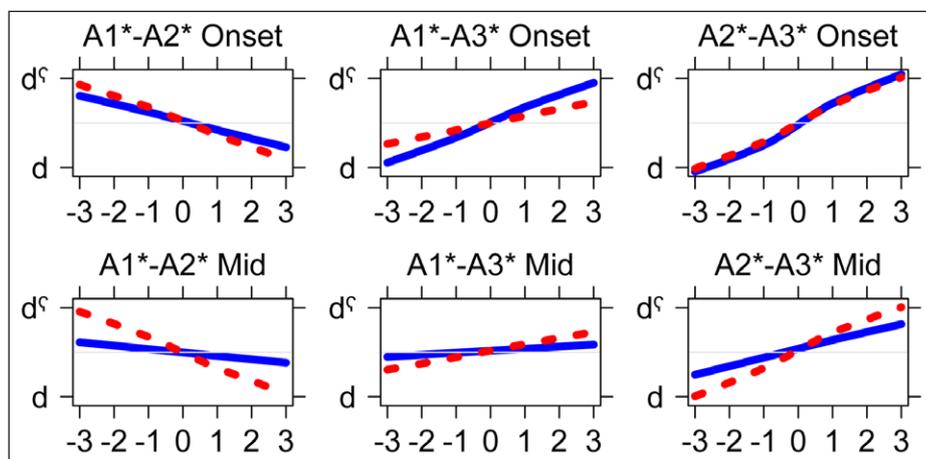
In both dialects, this decrease from /d<sup>ɕ</sup>/ to /d/ is a clear indication of a tense/pressed voice quality that is related to a greater glottal constriction (Keating et al., 2015). This leads to variation in the bandwidths of F1 and F2 (through  $H1^*-A1^*$  and  $H1^*-A2^*$  respectively) that accompanies tense/pressed voice quality; an increased bandwidth of F1 and F2 causes a decrease in the amplitude of the harmonic closest to F1 and F2, respectively (Garellek, 2012; Hanson et al., 2001). And finally, a decrease in  $H1^*-A3^*$  is indicative of tense phonation with a flatter spectrum (Garellek, 2012; Klatt & Klatt, 1990; Kuang & Keating, 2012). This seems to be the case with MA as a more abruptly constricted glottis can lead to an increase in energy around F3 and hence a flatter spectrum and a lowered  $H1^*-A3^*$  (Hanson & Chuang, 1999; Hanson et al., 2001). For JA however, the increased



**Figure 5:** Predicted probability curves spanning -3 to +3 z-score for Spectral slope results in both JA (blue solid) and MA (red dashed). Falling curves indicate a lowering in the predicted probabilities from /d<sup>ɕ</sup>/ to /d/ (as in  $H1^*-A1^*$ ) and vice versa. The reference line (in gray) indicates the non difference line.

$H1^*-A3^*$  is indicative of a lowered energy around F3 in the pharyngealized context, which seems to correlate well with the predictions of a constricted glottis with a simultaneous front to back closure along the length of the vocal folds (for more detail, see e.g., Hanson & Chuang, 1999; Hanson et al., 2001; Klatt & Klatt, 1990; Stevens, 1998). Both dialects display tense voice quality that is associated with pharyngealization though spectral slope measures, but seem to display different patterns in terms of how constricted the glottis is; abrupt in MA and with a simultaneous front to back closure of the vocal folds in JA.

Moving on to high frequency components, and through model comparison (Table 3), not all acoustic correlates show a significant contribution to the difference between the two consonants. In JA, only  $A1^*-A3^*$  and  $A2^*-A3^*$  at the onset are significant, whereas in MA, it is  $A1^*-A2^*$  and  $A2^*-A3^*$  at the onset and midpoint. This can also be seen from the predicted probabilities as shown in Figure 6 where some curves are near sigmoidal, while others are linear. The observed differences (through  $\beta$  and percent correct values) are small compared to formant based measures, and these high frequency components seem to act as secondary correlates. More specifically, a lower  $A1^*-A2^*$  is correlated with a constriction closer to the tongue root (i.e., mid to lower pharynx) as is the case in languages with [-ATR] vowel specification (see e.g., Aralova et al., 2011; Fulop et al., 1998; Guion et al., 2004; Kang & Ko, 2012, among others). This suggests that in MA, a lower constriction location for pharyngealization may be in operation, although a non-significantly lowered  $A1^*-A2^*$  at the onset in JA is obtained. With respect to the two other metrics, an increase in  $A1^*-A3^*$  and  $A2^*-A3^*$  is observed in both JA and MA with raised predicted probabilities (see Figure 6). This increase is indicative of a relatively decreased energy around F3 with respect to that of F1 or F2. In fact, A1 and A2 are already high in energy due to the pharyngeal constriction, and the change observed in A3 energy would be indicative of a change due to the epilaryngeal constriction that leads to a less/flatter spectral tilt (Halle & Stevens, 1969). The combination of spectral slope and the high frequency components are indicative of a constricted epilarynx; the former shows acoustic correlates of a tense voice quality, and the latter a constricted epilarynx with a relative increase in  $A1^*-A3^*$  and  $A2^*-A3^*$  due to decreased energy around F3. This combination suggests that pharyngealization in Arabic is associated with constricted ventricular folds (Moisik et al., 2012; Moisik & Esling, 2011; Sylak-Glassman, 2014a). This constriction of the glottis leads to a tense voice quality that is potentially associated with



**Figure 6:** Predicted probability curves spanning -3 to +3 z-score for the High frequency components in both JA (blue solid) and MA (red dashed). Falling curves indicate a lowering in the predicted probabilities from /d<sup>s</sup>/ to /d/ (as in  $A1^*-A2^*$ ) and vice versa. The reference line (in gray) indicates the non difference line.

a constricted and/or raised larynx posture (Klatt & Klatt, 1990; Laver, 1994; Sundberg & Askenfelt, 1981). These novel acoustic consequences require further investigation from an articulatory point of view to shed light into this secondary correlate of pharyngealization in Arabic.

### 3.1.5 Summary of results

The results presented above showed the vowels in the vicinity of pharyngealized consonants in Arabic to be ‘retracted’ (Esling, 2005; Sylak-Glassman, 2014a, b). They are produced as more open (higher F1, Z1-Z0) and more back (lower F2, higher Z3-Z2), with a constriction in the pharyngeal area that causes compaction of the spectrum (lower Z2-Z1), and spectral divergence as an enhancing correlate to the compacted spectrum (higher Z3-Z2). Z2-Z1 seems to provide a better combined correlate to the compaction of the spectrum rather than the separate F1 and F2, although both are related to each other (see Section 3.1.3). Our novel results with respect to spectral slope and the high frequency components seem to suggest that pharyngealization in Arabic is associated with a secondary constricted ventricular folds and hence a constricted epilarynx; this induces a tense voice quality with lower  $H1^*-A1^*$ ,  $H1^*-A2^*$ ,  $H1^*-A3^*$ , and  $A1^*-A2^*$ , and an increased  $A1^*-A3^*$  and  $A2^*-A3^*$  due to the decreased energy around F3. This is caused by a secondary constriction of the epilarynx, and more specifically the ventricular folds.

Given that we used individual GLMM analyses due to the constraints of our data (high collinearity), the individual  $\beta$  and percent correct values provided some insights into the discriminatory power of each of these acoustic correlates. However, it is not clear how these behave together and which acoustic correlates are more informative than the others. The next section presents an exploratory classification technique, namely random forests, that will shed light into the discriminatory power of the combined acoustic correlates.

## 3.2 Random Forests

The results presented in the previous section showed that none of the acoustic correlates exhibited a completely flat line (i.e.,  $\beta = 0$  and less than 50% correct), thus we included all 26 acoustic correlates (the 13 acoustic correlates at both onset and midpoint) in the random forest analysis for each dialect separately (a separate analysis on the significant only results showed the same overall percent correct classification, and correlation with actual data). Overall, the random forest analysis correlated well with our actual data (JA:  $r^2 = 0.87$ , mean squared error: 0.065; MA:  $r^2 = 0.82$ , mean squared error: 0.088). After running the random forest analysis, we used the `predict` function from the `party` package to provide predictions based on the *out-of-bag* data for cross-classification by specifying `OOB=TRUE`. This allows the algorithm to train itself on two-thirds of the data (the *in-bag* set), and then to use the remaining third of the data (the *out-of-bag* set) for classification (see Hothorn et al., 2006; James et al., 2013, for more detail). The overall percent correct classification with *out-of-bag* cross-classification was 93.5% for JA and 91.1% for MA, indicating a high discriminatory power of all of our acoustic correlates. Following this, we ran conditional permutations variable importance to measure the strength of each of the acoustic correlates *conditional* on each other and by using `varimpAUC` and `conditional=TRUE`. **Figure 7** provides the ranking of the different acoustic correlates using the mean decrease in accuracy in the overall classification when a particular acoustic correlate is removed.

The results show a clear separation between formant-based and voice quality-based measures with the former being highly predictive of the two consonant category (**Figure 7**). Looking at the results in detail, the ranking of the predictors in JA (**Figure 7a**) shows that F2 Onset, Z3-Z2 Onset, followed by Z2-Z1 Onset, are the most important

correlates, whereas in MA (Figure 7b), Z2-Z1 Onset followed by F2 Onset, and Z3-Z2 Onset are the most important. All the other correlates have lower values and thus can be regarded as secondary. It can be claimed that these differences in order are indicative of a difference in the phonetic implementation of pharyngealization across these dialects; when compared with the GLMM results above, MA potentially shows more mid-lower pharyngeal constriction, while JA shows mid-upper pharyngeal constriction. More articulatory data is needed to shed light on this difference. It should be noted that the differences between the dialects in the range of mean decrease in accuracy is large, with JA showing a score of 0.027 of the conditional mean decrease in accuracy when F2 Onset is removed, whereas in MA it is a 0.015 decrease for F2 Onset. These results correlate well with the  $\beta$  values presented in the GLMM results (see Section 3.1).

Following the initial random forest classification results, we ran six additional ones that are only used for their predictive accuracy. Table 4 provides a summary of the predictive accuracy of each of these additional random forests and as a comparison, the results of the initial classification are provided (first column).

The results point toward formant-based metrics as the most predictive for the difference between the two contexts. This is to be expected as pharyngealization is primarily reported as affecting formant frequencies. When both absolute and Bark-difference metrics are used together (i.e., Form + BkDiff), they show a non-significant lower classification rate than that of the full model (by 0.1–0.3%). This indicates that formant-based metrics have the most explanatory power, and may be suggestive of a

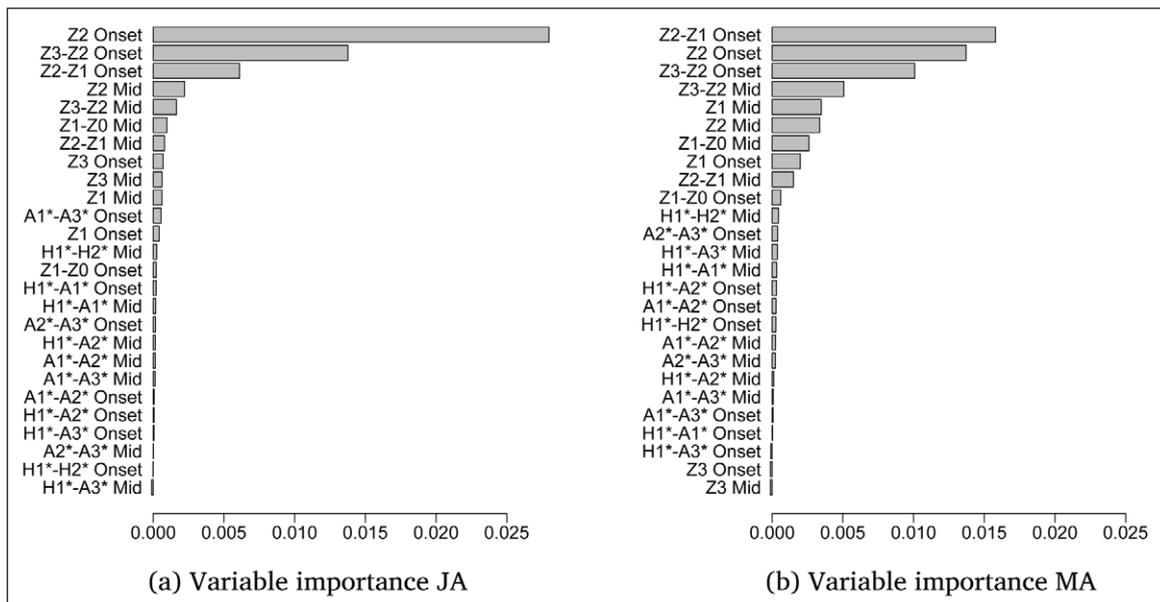


Figure 7: Mean decrease in accuracy importance scores in JA (a) and MA (b).

Table 4: Summary of predictive accuracy as classification rate for each of the random forests in JA and MA.

	Form + BkDiff + VQ	Form + BkDiff	Form	BkDiff	Form + VQ	BkDiff + VQ	VQ
JA	93.5%	93.2%	92.1%	92.9%	92.2%	93.1%	70.2%
MA	91.1%	91.0%	90.5%	90.8%	90.6%	91.0%	75.8%

Form = Absolute Formants; BkDiff = Bark-Difference formants; VQ = Voice Quality.

minimal role of voice quality metrics in this contrast. When comparing Bark-difference metrics to absolute ones (either on their own or combined with voice quality), the results show a minor non-significant improvement to the classification rates (by 0.5–1%). This minor improvement is indicative of both absolute formant and Bark-difference metrics as showing the same patterns; they are both indicative of [+OPEN] and [+BACK]. However, Bark-difference metrics show a minor advantage in representing pharyngealization, as the GLMM results summarized above (see **Figures 3b** and **4b**) and the current classification results show that they correlate well with the ‘retracted’ quality of the vowels in the pharyngealized context, following LAM (Esling, 2005; Sylak-Glassman, 2014a, b). Overall, the ranking of the various metrics was comparable to those reported in the full model (see **Figure 7**). However when absolute formants were used on their own, F2 at the Onset was the main correlate in both dialects. As expected, these results point to the fact that formant-based metrics are the primary correlates to pharyngealization in JA and MA.

Finally, the predictive accuracy of voice quality metrics on their own was assessed. The results presented in the last column of **Table 4** suggest that in both JA and MA, voice quality metrics are relatively important as they provide a classification accuracy of 70.2% for JA and 75.8% for MA. These rates are lower than those obtained for formant-based metrics, but provide support for voice quality metrics to act as a secondary correlates to pharyngealization. The ranking of correlates (not presented here) matches what we already saw in **Figure 7**, in that  $A1^*-A3^*$ ,  $A2^*-A3^*$ ,  $H1^*-A1^*$ ,  $H1^*-A2^*$  are the most predictive acoustic correlates in both dialects. However, in MA, spectral slope correlates are used more than in JA, and JA tends to use more the high-frequency components more. In both dialects, voice quality measures are indicative of an epilaryngeal constriction used as a secondary articulatory setting with formant-based metrics being the best predictors.

#### 4 Discussion and Conclusion

This exploratory study is aimed at investigating whether pharyngealization in Arabic is associated with an epilaryngeal constriction from an acoustic perspective. Traditionally, pharyngealization in Arabic is generally assumed to involve tongue body (dorsum) retraction towards the upper-pharyngeal areas that leads to a lowering of the second formant in the surrounding vowels (e.g., Bin-Muqbil, 2006; Ghazeli, 1977; Watson, 2007; Zawaydeh, 1999; Zawaydeh & de Jong, 2011). However, both classic and more recent articulatory evidence show this constriction to be located much lower in the pharynx, with epiglottis retraction, and raising of the larynx that leads to a pressed/tense voice quality (see e.g., Al-Tamimi, F. & Heselwood, 2011; Cantineau, 1960; Hess, 1998; Laufer & Baer, 1988; Lehn, 1963; Zeroual & Clements, 2015; Zeroual et al., 2011, among others). Following the Laryngeal Articulator Model (Esling, 2005), ‘true’ pharyngeals seem to show this type of configuration, with the whole epilarynx being constricted around the aryepiglottic folds. Pharyngealization on the other hand seems to show this type of constriction albeit as a secondary one with a primary pharyngeal constriction (Esling, 2005; Moisik, 2013a; Sylak-Glassman, 2014a, b). Acoustically, only the first three formants are usually investigated (see Section 1.1.1 above). However, these are only indicative of a pharyngeal constriction and do not *directly* explain any voice quality correlates that are a by-product of an epilaryngeal constriction. An epilaryngeal constriction is expected to show a combined effect of the *back and down* movement of the tongue as well as laryngeal muscles constriction (Esling, 2005; Moisik, 2013a; Sylak-Glassman, 2014a, b).

This study investigates the absolute formant frequencies, which is typical of such studies, but also complements this by looking at Bark-difference metrics and voice quality

correlates. This is the first study that combines these acoustic correlates in investigating pharyngealization in Arabic. Our results with respect to absolute formants are in agreement with previous literature, with vowels in the pharyngealized context showing decrease in F2 and an increase in F1 regardless of the vowel quality. In addition, F3 shows an overall increase in both dialects, but seems to show lowering when the vowel is front, and rising when the vowel is back (see Section 3.1.1 and **Figure 1**). Using the Distinctive Regions Model (DRM) (Carré & Mrayati, 1992; Mrayati et al., 1988) that is based on the principles of Perturbation Theory (Chiba & Kajiyama, 1941), it is possible to predict the constriction location based on the acoustic output. When F1 is rising, F2 is falling, and F3 is rising as in back and central vowels, the location of the constriction is close to the DRM region R4 in the upper pharynx. A combination of rising F1, and falling F2 and F3 seems to be close to the DRM region R3 in the mid-lower pharynx (see e.g., Carré & Mrayati, 1992, Figure 8, p. 150 and Figure 12, p. 156). These acoustic results are compatible with the articulatory accounts presented in Ghazeli (1977, p. 174, as cited in Laufer & Baer, 1988, p. 55) that ‘emphatic’ consonants have a secondary tongue retraction located midway between uvulars and pharyngeals.

Following the predictions of the Laryngeal Articulator Model (Esling, 2005; Sylak-Glassman, 2014a, b) pharyngealization is associated with a retracted production that has a combined *back and down* gesture. When considering this articulatory configuration, it is expected to observe both a lowering of F2 and a rising of F1 as a combined acoustic consequence. Thus the distance between these two formants can be seen as an alternative metric to signal the effects of pharyngealization, i.e., Z2-Z1. We investigated other formant metrics, e.g., Z1-Z0 and Z3-Z2 as correlates of a more open and a more back articulation and these correlated well with the traditional F1 and F2 dimensions separately. These have provided a clearer picture to the difference between the two consonants as reflected by the acoustic vowels spaces (see **Figures 3 and 4**).

The figures obtained for Z2-Z1 (see Section 3.1.2) can also be correlated with an upper-mid pharyngeal constriction. Compared to /d/, Z2-Z1 in the pharyngealized context was significantly lower and averaged around 5 to 6 Bark in both dialects (see Table A3 in Appendix 3), which is close to the range reported in other studies (see e.g., Al-Tamimi, F. & Heselwood, 2011; Heselwood & Al-Tamimi, F., 2011). These figures are of course indicative of a much higher constriction location than ‘true’ pharyngeals that have a smaller distance of about 3 Bark. Z1-Z0 and F1 results in both dialects are indicative of a more open production of the vowels in the vicinity of /d<sup>ɣ</sup>/ (Fahey et al., 1996; Hoemeke & Diehl, 1994; Syrdal & Gopal, 1986; Traunmüller, 1981). A 0.5 to 1 Bark difference (78 to 118 Hz) that correlates well with a one-degree change on the openness dimension agrees with the similarity scales between vowels and pharyngeals or uvulars as reported in Sylak-Glassman (2014b). When in contact with either category, vowels tend to be produced with one supplementary opening degree. In fact, the figures reported for Z2-Z1 are correlated with spectral integration (Fahey et al., 1996; Traunmüller, 1983, 1984). When dealing with ‘phonetically’ similar vowels, i.e., allophones changing in quality due to (non-) pharyngealization, a Z2-Z1 distance close to or below 6 Bark signals spectral integration (Traunmüller, 1983, p. 5–6). This as a whole leads to the conclusions that Z2-Z1 seems to be *the* major correlate to distinguish the two categories as the closeness of formants leads to them being perceived as one. And finally, Z3-Z2 provided another important acoustic correlate to the /d/-/d<sup>ɣ</sup>/ distinction: The pharyngealized category showed higher values with a difference of 1 to 2 Bark. In the non-pharyngealized context, all Z3-Z2 values were below the 3.5 Bark threshold for spectral integration (Chistovich & Lublinskaya, 1979; Syrdal & Gopal, 1986). We use the term ‘spectral flatness’ or divergence (Wood, 1986)

to explain the results in the pharyngealized context; vowels in the /d<sup>ɣ</sup>/ environment show a clear separation between the F2 and F3 due to the extremely lowered F2 and the relatively ‘stable’ F3. Thus the spectral divergence through the larger Z3-Z2 difference in the pharyngealized context is acting as an enhancement feature to the already spectrally integrated peak, i.e., Z2-Z1.

The combination of these three psychoacoustic metrics at the onset of the vowels in the pharyngealized context seems to show a direct psychoacoustic manifestation of the [+FLAT] feature as initially suggested by (Jakobson, 1957/1962; Jakobson et al., 1952/1976) as an acoustic correlate to pharyngealization. Z2-Z1 seems to be the main psychoacoustic correlate as it is close and below the threshold for spectral integration, Z3-Z2 seems to play a role as an enhancement feature and Z1-Z0 seems to correlate well with the more open articulation.

A second articulatory consequence of an epilaryngeal constriction is voice quality changes. Ventricular folds constriction leads to a tense, harsh, laryngealized voice quality, whereas an aryepiglottic constriction leads to ‘trilling’ as seen in ‘true’ pharyngeals and an ‘enhanced’ voice quality (Edmondson & Esling, 2006; Esling, 1996; Moisik, 2013a, b; Moisik et al., 2010; Story, 2016). Our results with respect to voice quality correlates—spectral slope and tilt—provide a secondary description to pharyngealization in Arabic. Pharyngealization is associated with a tense voice quality that is caused by constricting the intrinsic muscles of the larynx—the ventricular folds (see e.g., Garellek, 2012; Hanson & Chuang, 1999; Hanson et al., 2001; Keating et al., 2015; Kuang & Keating, 2012). Our estimated F1 and F2 bandwidths correlate well with a tense/pressed voice quality through a decrease in  $H1^*-A1^*$  and  $H1^*-A2^*$  metrics (Garellek, 2012; Hanson et al., 2001). Our estimated bandwidth through harmonic amplitude difference shows a clear difference from the results obtained by Alwan (1986, 1989), as an increased F2 bandwidth was found in pharyngeals but an increased F1 bandwidth in uvulars. Our results suggest pharyngealized consonants to be different from both pharyngeals *and* uvulars through the combined increase in the bandwidths of F1 and F2. Finally, a lowered  $H1^*-A3^*$  is also suggestive of a tense voice quality through an abrupt constriction of the glottis, specifically in MA (Garellek, 2012; Hanson & Chuang, 1999; Hanson et al., 2001; Klatt & Klatt, 1990). In JA, an increase in  $H1^*-A3^*$  is indicative of an increased spectral tilt and partially reduced energy around F3; consequences of a constricted glottis with a simultaneous front to back closure of the vocal folds (Hanson & Chuang, 1999; Hanson et al., 2001). This also seems to correlate well with a tense/pressed voice quality (Garellek, 2012; Hanson et al., 2001; Keating et al., 2015; Kuang & Keating, 2012).

The high frequency components provide an additional correlate to the glottal constriction through a constricted epilarynx and this constriction provides additional energy to the upper formants when comparing the two categories (Halle & Stevens, 1969; Stevens, 1977; Story, 2016). Our results showed that in both JA and MA an increased  $A1^*-A3^*$  and  $A2^*-A3^*$  is found. This is caused by a decrease in the energy around F3 that is potentially caused by the type of constriction in the glottis. Constricting the epilarynx on its own would lead to a decrease in the upper harmonic differences as this would lead to a concentration of the harmonics closest to F3, F4, and F5. This is seen in the singer’s formant—a primary technique used in signing to enhance the voice quality (Story, 2016; Titze & Story, 1997). In our case, constricting the epilarynx is secondary; pharyngeal constriction leads to a boost in the energy around F1 and F2, and the changes seen with respect to F3 will be minimal. This causes a decrease in the energy around F3 leading to the increased harmonic difference observed in this study. Liénard and Di Benedetto (1999) found that higher levels of vocal effort seem to increase the amplitudes of A2 and A3 more than what is seen in A1; the latter is relatively higher in amplitude than the former. Their harmonic

differences, based on our estimations, seem to show a similar pattern to the results reported here; a relatively increased  $A1^*-A3^*$  and  $A2^*-A3^*$  in the high effort is observed compared to the normal setting (Liénard & Di Benedetto, 1999, Figure 3, p. 416). In this particular case, the high effort was measured in the vowels; our results are mostly seen at the onset of the vowel and are indicative of this secondary constriction that shows a high effort comparable to what is seen in tense voice. Finally, and according to Story (2016, Figure 7, p. 10), a constricted epilarynx would lead to an increase in the frequency *and* amplitude of F3 to make it move closer to F4. This closeness correlates with the singer's formant and changes the 'timbre' or quality of voice of a particular speaker. F1 and F2 are responsible for the phonetic quality of a vowel, whereas voice quality is manifested through F3 to F5 frequencies and amplitudes. Hence if we postulate that pharyngealization is associated with an epilaryngeal constriction, F3 to F5 amplitude and frequencies should be investigated as well. This future work is currently planned on a forthcoming dataset.

The Random forest technique was used as a predictive model to test for classification rates and to provide a fine-tuned ordering of the various correlates used in this study. The results showed a clear separation between formant-based and voice quality-based measures. Voice quality results are secondary in the sense that they allow for a finer description of pharyngealization in Arabic as involving a constricted glottis through constriction of the epilarynx. The primary acoustic correlates are formant-based and reflect the changes in the vowel quality into a more open (high F1 and Z1-Z0), a more compact (low Z2-Z1), a more retracted (low F2, and high Z3-Z2) with spectral integration (low Z2-Z1) and divergence (high Z3-Z2). The primary acoustic correlate is different in the two dialects if taking into account the psychoacoustic correlates: In JA it is F2-based with retraction and spectral divergence, while in MA it is a combined F1 and F2-based through a more compact spectrum with retraction. When considering absolute formant frequencies alone, both dialects display the same pattern i.e., F2-based followed by F1-based; results that are largely reported in the literature (see Section 1.1.1). Bark-difference metrics highlight the fact that the supposed differences between Arabic dialects in how pharyngealization is implemented (with some dialects having a more 'guttural' quality) can be evaluated from an acoustic point of view: It is possible that MA has a more 'pharyngeal' quality than JA (see e.g., Bellem, 2007; Embarki et al., 2011), though it is not clear whether these differences are only based on formant patterns or voice quality ('timbre' of the voice). More detailed articulatory and acoustic description is required to shed light into these differences.

#### **4.1 Implications for formal representations**

Various features were proposed in the literature to describe the effects of pharyngealization, and these include: [CONSTRUCTED PHARYNX] (Hoberman, 1987); [CONSTRUCTED TONGUE ROOT] (Stevens, 1998); [CONSTRUCTED] (Edmondson & Esling, 2006); [CONSTRUCTED EPILARYNGEAL TUBE] (Moisik, 2013a; Moisik et al., 2012; Moisik & Esling, 2011); [CONSTRUCTED EPILARYNX] (Sylak-Glassman, 2014a); [EXPANDED PHARYNX] (Lindau, 1975, 1978); [FLAT] (Jakobson, 1957/1962; Jakobson et al., 1952/1976); [GUTTURAL] (Hayward & Hayward, 1989; Watson, 2007); [LOWER/UPPER PHARYNX] (Czaykowska-Higgins, 1987); [PHARYNGEAL] (McCarthy, 1994; Zeroual & Clements, 2015); [RETRACTED] (Esling, 2005; Sylak-Glassman, 2014b); [RETRACTED TONGUE ROOT] (Shahin, 1996, 1997, 2011; Watson, 2007). This list is not exhaustive, however, it highlights the need for a unified set of feature(s) to account for pharyngealization. Currently, [+RTR] is the widely used feature specification for pharyngealization at least in Arabic that reflects the activity of the retraction of the root of the tongue. It is not clear however whether this retraction causes only lowering of F2 frequencies (i.e., Z2 and Z3-Z2) or whether it

accounts for the more open production with rising of F1 (i.e., Z1, Z1-Z0). In addition, the more compacted spectrum through the lower Z2-Z1 does not seem to be reflected by [+RTR]. Finally, it is not clear whether constricting the epilarynx can be accounted for by [+RTR]. It is of course possible to add an additional laryngeal specification, i.e., [+CONSTRICTED GLOTTIS] that can account for our secondary correlates. However, given that the tongue retraction as seen here is a by-product of constricting the epilarynx (e.g., Esling, 2005; Moisik, 2013a; Sylak-Glassman, 2014a, b, and Section 1.2), we follow Moisik et al. (2012), Moisik and Esling (2011), and Sylak-Glassman (2014a) and postulate that pharyngealization in Arabic is produced by an epilaryngeal constriction that causes the tongue root and body to be retracted by a *back and down* gesture. This induces a laryngeal constriction leading to a raised larynx posture and a tense/pressed voice quality. Based on these descriptions and on our results, pharyngealization in Arabic will then be associated with a [+CET] that has ‘retraction’ and ‘constricted glottis’ as one component. The feature [+CE] (Sylak-Glassman, 2014a) that follows [+CET], can account for this combined effect, though it is not clear whether [+CONSTRICTED GLOTTIS] is inherent to [+CE] or should be added as an additional feature to describe pharyngealization (see Sylak-Glassman, 2014a, p. 136–140). The inclusion of voice quality metrics allows for pharyngealization in Arabic to be described with [+CET]; if formant metrics only were used here, our conclusion would be that [+RTD] (for ‘retraction,’ following Moisik et al., 2012; Moisik & Esling, 2011) is the only correlate of pharyngealization. Of course this analysis is only based on the outcomes of this study and does not look at the patterns in the language. However, highlighting this laryngeal activity as associated with pharyngealization in Arabic can potentially shed light into what a ‘guttural’ quality is from a phonetic point of view and whether it is a ‘voice quality’ in the sense of ‘timbre’ of the vowel (following Story, 2016), a ‘darker’ or ‘heavier’ auditory quality from formant merging, or both.

#### 4.2 Limitations

This exploratory study highlights some novel acoustic evidence in describing pharyngealization in Arabic. However, it should be noted that only coronal voiced stops were investigated. Our formant-based measures are congruent with those reported in the literature in the direction of the effect, however, there are some differences with the amount of distance between these consonants (on F1, F2, and Z2-Z1 values). It is not clear whether some of the results obtained are restricted to this category of sounds or whether they can also be extended to other coronal consonants. We expect the acoustic correlates of pharyngealization reported here to be reflected in all pharyngealized consonants to different degrees. Our voice quality results should not be seen as restricted to this category of sounds, as it is associated with a tense/pressed voice quality through an epilaryngeal constriction in general, thus we expect to obtain the same degree of high frequency components regardless of the consonant category. An additional limitation to our study is the use of stimuli varying in location of the pharyngealized consonants according to syllable structure and to the surrounding sounds, i.e., other gutturals. This was due to finding real words fitting this consonant/vowel combination. Thus some of the outcomes may have been different if the corpus was more balanced; some dialectal differences may be due to variable syllable structures, or voice quality results at the midpoint would have been different. Given the exploratory nature of this study, we wanted to evaluate the results on such a corpus to allow for testing this particular hypothesis. Our exploratory study did not investigate the acoustic correlates of ‘true’ pharyngeals or uvulars as our aim was to evaluate which acoustic correlates are mostly associated with pharyngealization to allow for an extension to other categories. Finally, our study is acoustic in nature, and any conclusions on exact articulatory consequences should be evaluated cautiously. A planned subsequent project will investigate further this idea of

an epilaryngeal constriction as associated with back consonants and will specifically use a combined articulatory (e.g., ultrasound, and electoglottography) and advanced acoustic data of all front and back consonants in Arabic to shed light into the laryngeal constriction and the exact constriction location of each consonant; this will allow for a meaningful articulatory-to-acoustic mapping.

## 5 Conclusion

This exploratory study of Jordanian and Moroccan Arabic looked at the combined effect of multiple acoustic correlates in describing pharyngealization. By following the predictions of the Laryngeal Articulator Model (Esling, 2005) and its subsequent developments (Moisik, 2013a; Sylak-Glassman, 2014a), we showed how pharyngealization in Arabic is associated with a ‘retracted’ production, with a combined *back and down* gesture through an epilaryngeal constriction leading to a constricted glottis. By using 13 different acoustic correlates obtained at each of onset and midpoint of the vowel, including formant-based and voice quality-based measures, it was possible to evaluate the ‘retraction,’ compaction, and epilaryngeal constriction in our data. Formant distance measures (through Bark-difference) allowed for a better evaluation of the spectral convergence (Z2-Z1) and divergence (Z3-Z2) and showed differences in how pharyngealization is implemented between the two dialects. In summary, the results suggest that MA is associated with a more mid-low pharyngeal constriction and JA with an upper-mid pharyngeal constriction. Voice quality measures from spectral slope and high frequency components correlated well with a tense/pressed voice quality and an epilaryngeal constriction respectively. These novel results were assessed through GLMM and exploratory random forest analyses. ‘Retraction’ (i.e., combined *back and down* gesture) is the primary acoustic correlate of pharyngealization in Arabic with an epilaryngeal constriction leading to a [+ CONstricted GLOTTIS] as a secondary correlate through an epilaryngeal constriction. If formant-based measures were used alone, a simple ‘retraction’ would be the main and only correlate to pharyngealization. Voice quality correlates allowed for the epilaryngeal constriction to be included. A combined articulatory and acoustic investigation of the state of the epilarynx in Arabic pharyngeals and pharyngealization is worth pursuing in order to evaluate the role of the epilarynx as an active ‘articulator.’ A subsequent perceptual study will shed light into which acoustic correlates are the most prominent in identifying pharyngeals and pharyngealization in Arabic. All these can also be the basis for dialectal classification and discrimination as presented in the current study.

## Additional Files

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

- **Appendix 1.** Figure A1: Correlation matrix for JA. DOI: <https://doi.org/10.5334/labphon.19.s1>
- **Appendix 2.** Figure A2: Correlation matrix for MA. DOI: <https://doi.org/10.5334/labphon.19.s2>
- **Appendix 3.** Table A3: Descriptive statistics for JA and MA. DOI: <https://doi.org/10.5334/labphon.19.s3>

## Acknowledgements

The author would like to thank Lisa Davidson, Associate Editor of Laboratory Phonology, Kip Wilson, Editorial Assistant of Laboratory Phonology, three anonymous reviewers, and Ghada Khattab, Danielle Turton, Dan McCarthy, Scott Moisik, John Esling, Bodo Winter, François Pellegrino, Thami BenKirane, the audiences of BAAP 2014, the 18<sup>th</sup> ICPhS, the 2<sup>nd</sup> Arabic Linguistics Forum, the 2<sup>nd</sup> PaPE, and the Phonetics and Phonology group at Newcastle University for fruitful discussions on various parts

of this study. This work was partially supported by the French Research Ministry (Programme Cognitive: Projet Langage et Cognition n°. 41, P.I. R. Carré), the French Agence Nationale de la Recherche (project NT05-3-43182 CL<sup>2</sup>, P.I. F. Pellegrino), and by three funds to attend BAAP 2014, ICPHS 2015, and PaPE 2017 conferences from Newcastle University.

### Competing Interests

The author has no competing interests to declare.

### References

- Al-Ani, S. 1970. *Arabic Phonology: An Acoustical and Physiological Investigation*. Paris: The Hague Mouton. DOI: <https://doi.org/10.1515/9783110878769>
- Al-Masri, M., & Jongman, A. 2004. Acoustic correlates of emphasis in Jordanian Arabic: Preliminary results. In: Agwuele, A., Warren, W., & Park, S.-H. (Eds.), *Proceedings of the 2003 Texas Linguistics Society conference*, 96–106. Somerville, MA: Cascadilla Proceedings Project.
- Al-Tamimi, F., & Heselwood, B. 2011. Nasoendoscopic, videofluoroscopic and acoustic study of plain and emphatic coronals in Jordanian Arabic. In: Heselwood, B., & Hassan, Z. (Eds.), *Instrumental studies in Arabic phonetics*, 165–191. John Benjamins. DOI: <https://doi.org/10.1075/cilt.319>
- Al-Tamimi, J. 2002. *Variabilité phonétique en production et en perception de la parole: Le cas de l'arabe jordano-palestinien* (Unpublished master's thesis). Université Lyon 2, Lyon, France.
- Al-Tamimi, J. 2004. L'équation du locus comme mesure de la coarticulation VC et CV : Étude préliminaire en Arabe Dialectal Jordanien. In: *Actes des 25èmes Journées d'Études sur la Parole (JEP)*, Dinard, France, 9–12.
- Al-Tamimi, J. 2007a. *Indices dynamiques et perception des voyelles : Étude translinguistique en arabe dialectal et en français* (Unpublished doctoral dissertation). Lyon: Université Lyon 2.
- Al-Tamimi, J. 2007b. Static and Dynamic cues in vowel production: A cross dialectal study in Jordanian and Moroccan Arabic. In: *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, 541–544. Saarbrücken, Germany.
- Al-Tamimi, J. 2009. Effect of pharyngealisation on vowels revisited: Static and Dynamic analyses of vowels in Moroccan and Jordanian Arabic. In: *Workshop "Pharyngeals and Pharyngealisation"*.
- Al-Tamimi, J. 2014. An acoustic investigation of pharyngealisation effects on vowels: A crossdialectal investigation in Jordanian and Moroccan Arabic. In: *Proceedings of the British Association of Academic Phoneticians (BAAP) Colloquium*.
- Al-Tamimi, J. 2015. Spectral tilt as an acoustic correlate to pharyngealisation in Jordanian and Moroccan Arabic. In: *The Scottish Consortium for ICPHS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0436. Retrieved from: <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0436.pdf>.
- Al-Tamimi, J., & Barkat-Defradas, M. 2003. Inter-dialectal and inter-individual variability in production and perception: A preliminary study in Jordanian and Moroccan Arabic. In: Ferrando, I., & Sanchez Sandoval, J. J. (Eds.), *5th AIDA Proceedings (Association Internationale de Dialectologie Arabe)*, 171–186. Cádiz, Spain.
- Al-Tamimi, J., & Ferragne, E. 2005. Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. In: *Interspeech*, 2465–2468.

- Al-Tamimi, J., & Khattab, G. 2015. Acoustic cue weighting in singleton vs geminate contrast in Lebanese Arabic: The case of fricative consonants. *The Journal of the Acoustical Society of America*, 138(1), 344–360. DOI: <https://doi.org/10.1121/1.4922514>
- Alwan, A. 1986. *Acoustic and perceptual correlates of pharyngeal and uvular consonants* (Doctoral dissertation). Cambridge, MA: Massachusetts Institute of Technology. Retrieved from: <http://hdl.handle.net/1721.1/34302>.
- Alwan, A. 1989. Perceptual cues for place of articulation for the voiced pharyngeal and uvular consonants. *The Journal of the Acoustical Society of America*, 86(2), 549–556. DOI: <https://doi.org/10.1121/1.398234>
- Aralova, N., Grawunder, S., & Winter, B. 2011. The Acoustic Correlates of Tongue Root Vowel Harmony in Even (Tungusic). In: *The 17th International Congress of Phonetic Sciences*, 240–243. Hong Kong.
- Auguie, B. 2016. gridExtra: Miscellaneous Functions for “Grid” Graphics [Computer software manual]. Retrieved from: <https://CRAN.R-project.org/package=gridExtra> (R package version 2.2.1).
- Baayen, R. H. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511801686>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. DOI: <https://doi.org/10.1016/j.jml.2007.12.005>
- Baguley, T. 2012. *Serious stats: A guide to advanced statistics for the behavioral sciences*. Palgrave Macmillan.
- Barkat, M. 2000. *Détermination d'indices acoustiques robustes pour l'identification automatique des parlers arabes* (Unpublished doctoral dissertation). Lyon: Université Lyon 2.
- Barkat-Defradas, M., Al-Tamimi, J., & Benkirane, T. 2003. Phonetic variation in production and perception of speech: A comparative study of two Arabic dialects. In: *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, 857–860. Barcelona, Spain.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. DOI: <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Bellem, A. 2007. *Towards a Comparative Typology of Emphatics: Across Semitic and Into Arabic Dialect Phonology*. (Unpublished doctoral dissertation). London: School of Oriental and African Studies, University of London.
- Bin-Muqbil, M. S. 2006. *Phonetic and phonological aspects of Arabic emphatics and gutturals* (Unpublished doctoral dissertation). Madison: University of Wisconsin-Madison.
- Boersma, P., & Weenink, D. 2009. *Praat. Doing Phonetics by Computer*. V. 5.1.0: Retrieved 22 July 2009 from: <http://www.praat.org/>.
- Breiman, L. 2001. Random forests. *Machine learning*, 45(1), 5–32. DOI: <https://doi.org/10.1023/A:1010933404324>
- Brown, L., Winter, B., Idemaru, K., & Grawunder, S. 2014. Phonetics and politeness: Perceiving Korean honorific and non-honorific speech through phonetic cues. *Journal of Pragmatics*, 66, 45–60. DOI: <https://doi.org/10.1016/j.pragma.2014.02.011>
- Cantineau, J. 1960. *Etudes de linguistique arabe*. Paris: Klincksieck.
- Carré, R., & Mrayati, M. 1992. Distinctive regions in acoustic tubes. Speech production modeling. *Journal d'Acoustique*, 5, 141–159.

- Catford, J. C. 1977. *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.
- Chiba, T., & Kajiyama, M. 1941. *The Vowel: Its Nature and Structure*. Tokyo-Kaiseikan Pub. Co., Ltd., Tokyo.
- Chistovich, L. A., & Lublinskaya, V. V. 1979. The ‘center of gravity’ effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research*, 1(3), 185–195. DOI: [https://doi.org/10.1016/0378-5955\(79\)90012-1](https://doi.org/10.1016/0378-5955(79)90012-1)
- Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. 2013. A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, 78(4), 685–709. DOI: <https://doi.org/10.1007/s11336-013-9328-2>
- Cohen, D. 1962. Koinè, langues communes et dialectes arabes. *Arabica*, 9, 119–144. DOI: <https://doi.org/10.1163/157005862X00637>
- Czaykowska-Higgins, E. 1987. *Characterizing tongue root behavior*. Manuscript, MIT, Cambridge, Massachusetts.
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 837–845. DOI: <https://doi.org/10.2307/2531595>
- Edmondson, J. A., & Esling, J. H. 2006. The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology*, 23(2), 157–191. DOI: <https://doi.org/10.1017/S095267570600087X>
- Embarki, M., Ouni, S., Yeou, M., Guilleminot, C., & Al Maqtari, S. 2011. Acoustic and electromagnetic articulographic study of pharyngealisation: Coarticulatory effects as an index of stylistic and regional variation in Arabic. In: Heselwood, B., & Hassan, Z. (Eds.), *Instrumental studies in Arabic phonetics*, 193–215. John Benjamins. DOI: <https://doi.org/10.1075/cilt.319>
- Esling, J. 1996. Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Association*, 26(2), 65–88. DOI: <https://doi.org/10.1017/S0025100300006125>
- Esling, J. 1999. The IPA Categories “Pharyngeal” and “Epiglottal” Laryngoscopic Observations of Pharyngeal Articulations and Larynx Height. *Language and Speech*, 42(4), 349–372. DOI: <https://doi.org/10.1177/00238309990420040101>
- Esling, J. 2005. There Are No Back Vowels: The Laryngeal Articulator Model. *The Canadian Journal of Linguistics/La revue canadienne de linguistique*, 50(1), 13–44. DOI: <https://doi.org/10.1353/cjl.2007.0007>
- Fahey, R. P., Diehl, R. L., & Traunmüller, H. 1996. Perception of back vowels: Effects of varying  $F_1 - F_0$  Bark distance. *The Journal of the Acoustical Society of America*, 99(4), 2350–2357. DOI: <https://doi.org/10.1121/1.415422>
- Fant, G. 1960/1971. *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations*, second edition. Walter de Gruyter.
- Fant, G. 2004. *Speech Acoustics and Phonetics: Selected Writings*, 24. Springer Science & Business Media. DOI: <https://doi.org/10.1007/978-1-4020-5746-5>
- Fraundorf, S. 2015. psycholing: R Functions for Common Psycholinguistic and Cognitive Designs [Computer software manual]. (R package version 0.5).
- Fulop, S., Kari, E., & Ladefoged, P. 1998. An Acoustic Study of the Tongue Root Contrast in Degema Vowels. *Phonetica*, 55(1–2), 80–98. DOI: <https://doi.org/10.1159/000028425>
- Garellek, M. 2012. The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics*, 40(1), 152–161. DOI: <https://doi.org/10.1016/j.wocn.2011.10.003>

- Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. 2008. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 1360–1383. DOI: <https://doi.org/10.1214/08-AOAS191>
- Ghazali, S., Hamdi, R., & Barkat, M. 2002. Speech Rhythm Variation in Arabic Dialects. In: *Proceedings of the 1st International conference on Speech Prosody*, 331–334. Aix-en-Provence, France.
- Ghazali, S., Hamdi, R., & Knis, K. 2007. Intonation and rhythmic patterns across the Arabic dialects continuum. In: Benmamoun, E. (Ed.), *Papers from the annual symposium on Arabic Linguistics*, 97–122. Volume XIX: Urbana, Illinois, USA, Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/cilt.289.10gha>
- Ghazali, S. 1977. *Back Consonants and Backing Coarticulation in Arabic* (Doctoral dissertation). Austin: University of Texas at Austin.
- Guion, S. G., Post, M. W., & Payne, D. L. 2004. Phonetic correlates of tongue root vowel contrasts in Maa. *Journal of Phonetics*, 32(4), 517–542. DOI: <https://doi.org/10.1016/j.wocn.2004.04.002>
- Halle, M., & Stevens, K. 1969. On the Feature ‘Advanced Tongue Root’. In: *Quarterly Progress Report No. 94*, Research Laboratory of Electronics, M.I.T, 209–215.
- Hanson, H. M., & Chuang, E. S. 1999. Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *The Journal of the Acoustical Society of America*, 106(2), 1064–1077. DOI: <https://doi.org/10.1121/1.427116>
- Hanson, H. M., Stevens, K. N., Kuo, H.-K. J., Chen, M. Y., & Slifka, J. 2001. Towards models of phonation. *Journal of Phonetics*, 29, 451–480. DOI: <https://doi.org/10.1006/jpho.2001.0146>
- Harrell, F. E., Jr. 2016. Hmisc: Harrell Miscellaneous [Computer software manual]. Retrieved from: <https://CRAN.R-project.org/package=Hmisc> (R package version 3.17-4).
- Hassan, Z., & Esling, J. 2011. Investigating the emphatic feature in Iraqi Arabic: Acoustic and articulatory evidence of coarticulation. In: Heselwood, B., & Hassan, Z. (Eds.), *Instrumental studies in Arabic phonetics*, 218–234. John Benjamins. DOI: <https://doi.org/10.1075/cilt.319>
- Hastie, T., Tibshirani, R., & Friedman, J. 2009. *The elements of statistical learning*. Springer series in statistics Springer, Berlin. DOI: <https://doi.org/10.1007/978-0-387-84858-7>
- Hawks, J. W., & Miller, J. D. 1995. A formant bandwidth estimation procedure for vowel synthesis. *The Journal of the Acoustical Society of America*, 97(2), 1343–1344. DOI: <https://doi.org/10.1121/1.412986>
- Hayward, K. M., & Hayward, R. J. 1989. ‘Guttural’: Arguments for a new distinctive feature. *Transactions of the Philological Society*, 87(2), 179–193. DOI: <https://doi.org/10.1111/j.1467-968X.1989.tb00626.x>
- Hellmuth, S. 2013. Phonology. In: Owens, J. (Ed.), *The Oxford handbook of Arabic linguistics*, 45–70. Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199764136.013.0003>
- Heselwood, B., & Al-Tamimi, F. 2011. A study of the laryngeal and pharyngeal consonants in Jordanian Arabic using nasoendoscopy, videofluoroscopy and spectrography. In: Heselwood, B., & Hassan, Z. (Eds.), *Instrumental studies in Arabic phonetics*, 101–128. John Benjamins. DOI: <https://doi.org/10.1075/cilt.319>
- Hess, S. A. 1998. *Pharyngeal Articulations* (Unpublished doctoral dissertation). Los Angeles, CA: UCLA.
- Hirst, D. 2011. The analysis by synthesis of speech melody: From data to models. *Journal of Speech Sciences*, 1, 55–83. Retrieved from: <http://www.journalofspeechsciences.org/index.php/journalofspeechsciences/article/view/21>.

- Hoberman, R. D. 1987. Emphasis (pharyngealization) as an autosegmental harmony feature. *Parasession on Autosegmental and Metrical Phonology*, 23, 167–181.
- Hoemeke, K. A., & Diehl, R. L. 1994. Perception of vowel height: The role of  $F_1$ - $F_0$  distance. *The Journal of the Acoustical Society of America*, 96(2), 661–674. DOI: <https://doi.org/10.1121/1.410305>
- Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. J. 2006. Survival ensembles. *Biostatistics*, 7(3), 355–373. DOI: <https://doi.org/10.1093/biostatistics/kxj011>
- Howard, D. M., & Angus, J. 2009. *Acoustics and psychoacoustics*. Taylor & Francis. DOI: <https://doi.org/10.4324/9780080468488>
- Iseli, M., Shue, Y.-L., & Alwan, A. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America*, 121(4), 2283–2295. DOI: <https://doi.org/10.1121/1.2697522>
- Jakobson, R. 1957/1962. Mufaxxama – The ‘emphatic’ phonemes in Arabic. In: Jakobson, R. (Ed.), *Selected Writings I: Phonological Studies (originally published in 1957 in E. Pulgram (Ed.): Studies presented to Joshua Whatmough)*, 510–522. The Hague: Mouton.
- Jakobson, R., Fant, G., & Halle, M. 1952/1976. *Preliminaries to Speech Analysis: The distinctive features and their correlates*. Cambridge, Massachusetts: MIT Press.
- James, G., Witten, D., & Hastie, T. 2013. *An Introduction to Statistical Learning: With Applications in R*. Springer. DOI: <https://doi.org/10.1007/978-1-4614-7138-7>
- Johnson, K. 2012. *Acoustic and Auditory Phonetics*. Wiley-Blackwell, A John Wiley & Sons, Ltd., Publication.
- Jongman, A., Herd, W., Al-Masri, M., Sereno, J., & Combest, S. 2011. Acoustics and perception of emphasis in Urban Jordanian Arabic. *Journal of Phonetics*, 39(1), 85–95. DOI: <https://doi.org/10.1016/j.wocn.2010.11.007>
- Kang, H., & Ko, S. 2012. In search of the acoustic correlates of tongue root retraction in three Altaic languages: Western Buriat, Tsongol Buriat, and Ewen. *Altai Hakpo*, 22, 179–203. Retrieved from: [http://qcpages.qc.cuny.edu/~sko/papers/Kang%26Ko\\_2012\\_AH22\\_paper.pdf](http://qcpages.qc.cuny.edu/~sko/papers/Kang%26Ko_2012_AH22_paper.pdf).
- Keating, P., Garellek, M., & Kreiman, J. 2015. Acoustic properties of different kinds of creaky voice. In: The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0821. Retrieved from: <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0821.pdf>.
- Khatab, G., Al-Tamimi, F., & Heselwood, B. 2006. Acoustic and auditory differences in the /t/-/t̤/ opposition in male and female speakers of Jordanian Arabic. In: *Perspectives on Arabic Linguistics XVI: Papers from the sixteenth annual symposium on Arabic linguistics*, 266, 131–160. John Benjamins.
- Klatt, D., & Klatt, L. 1990. Analysis, synthesis and perception of voice quality variations among male and female talkers. *The Journal of the Acoustical Society of America*, 87, 820–856. DOI: <https://doi.org/10.1121/1.398894>
- Kuang, J., & Keating, P. 2012. Glottal articulations of phonation contrasts and their acoustic and perceptual consequences. *UCLA Working Papers in Phonetics*, 111, 123–161. Retrieved from: <http://escholarship.org/uc/item/0n76t3hn>.
- Kuang, J., & Keating, P. 2014. Vocal fold vibratory patterns in tense versus lax phonation contrasts. *The Journal of the Acoustical Society of America*, 136(5), 2784–2797. DOI: <https://doi.org/10.1121/1.4896462>
- Kuhn, M., & Johnson, K. 2013. *Applied predictive modeling*, 26. Springer. DOI: <https://doi.org/10.1007/978-1-4614-6849-3>
- Ladefoged, P., & Maddieson, I. 1996. *The Sounds of the World's Languages*. Cambridge, MA: Blackwell.

- Laufer, A., & Baer, T. 1988. The emphatic and pharyngeal sounds in Hebrew and in Arabic. *Language and Speech*, 31, 181–205. DOI: <https://doi.org/10.1177/002383098803100205>
- Laver, J. 1980. *The Phonetic Description of Voice Quality*. Cambridge University Press.
- Laver, J. 1994. *Principles of Phonetics*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139166621>
- Lehn, W. 1963. Emphasis in Cairo Arabic. *Language*, 39(1), 29–39. DOI: <https://doi.org/10.2307/410760>
- Liaw, A., & Wiener, M. 2002. Classification and regression by randomforest. *R News*, 2(3), 18–22.
- Liénard, J.-S., & Di Benedetto, M.-G. 1999. Effect of vocal effort on spectral properties of vowels. *The Journal of the Acoustical Society of America*, 106(1), 411–422. DOI: <https://doi.org/10.1121/1.428140>
- Lindau, M. 1975. [Features] for Vowels. *UCLA Working Papers in Phonetics*, 30, 1–163.
- Lindau, M. 1978. Vowel Features. *Language*, 54(3), 541–563. DOI: <https://doi.org/10.1353/lan.1978.0066>
- Lindblom, B., & Sundberg, J. 1971. Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*, 50(4B), 1166–1179. DOI: <https://doi.org/10.1121/1.1912750>
- Marçais, P. 1977. *Esquisse grammaticale de l'arabe maghrébin*. Paris: Adrien-Maisonneuve.
- McCarthy, J. J. 1994. The phonetics and phonology of Semitic pharyngeals. In: Keating, P. (Ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, 191–233. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511659461.012>
- McCloy, D. R. 2016. phonr: tools for phoneticians and phonologists [Computer software manual]. (R package version 1.0-7)
- Melike. 2016. *Heatmap Table Using ggplot2*. Retrieved from: <http://rpubs.com/melike/heatmapTable>.
- Moisik, S. 2013a. *The epilarynx in speech* (Unpublished doctoral dissertation). Victoria: University of Victoria.
- Moisik, S. 2013b. Harsh voice quality and its association with blackness in popular American media. *Phonetica*, 69(4), 193–215. DOI: <https://doi.org/10.1159/000351059>
- Moisik, S., Czaykowska-Higgins, E., & Esling, J. 2012. The epilaryngeal articulator: A new conceptual tool for understanding lingual-laryngeal contrasts. *McGill Working Papers in Linguistics*, 22(1). Retrieved from: <https://www.mcgill.ca/mcgwpl/files/mcgwpl/moisik2012.pdf>.
- Moisik, S., & Esling, J. 2010. Examining the Acoustic Contributions of the Epilaryngeal Tube to the Voice Source and Vocal Tract Resonance. *Canadian Acoustics*, 38(3), 138–139. Retrieved from: <http://jaa.caa-aca.ca/index.php/jaa/article/view/2283>.
- Moisik, S., & Esling, J. 2011. The ‘whole larynx’ approach to laryngeal features. In: *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, 1406–1409. Hong Kong.
- Moisik, S., Esling, J., & Crevier-Buchman, L. 2010. A high-speed laryngoscopic investigation of aryepiglottic trilling. *The Journal of the Acoustical Society of America*, 127(3), 1548–1558. DOI: <https://doi.org/10.1121/1.3299203>
- Mrayati, M., Carré, R., & Guérin, B. 1988. Distinctive regions and modes: A new theory of speech production. *Speech Communication*, 7(3), 257–286. DOI: [https://doi.org/10.1016/0167-6393\(88\)90073-8](https://doi.org/10.1016/0167-6393(88)90073-8)
- Nolan, F. 1983. *The phonetic bases of speaker recognition. Cambridge studies in speech science and communication*. Cambridge University Press, Cambridge.
- Norlin, K. 1987. *A phonetic study of emphasis and vowels in Egyptian Arabic* (Unpublished doctoral dissertation). Lund: Lund University.

- Obrecht, D. H. 1968. *Effects of the Second Formant on the Perception of Velarisation Consonants in Arabic*. The Hague: Mouton.
- Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. 2012. How many trees in a random forest? In: *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, 154–168. DOI: [https://doi.org/10.1007/978-3-642-31537-4\\_13](https://doi.org/10.1007/978-3-642-31537-4_13)
- Quené, H., & Van den Bergh, H. 2008. Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413–425. DOI: <https://doi.org/10.1016/j.jml.2008.02.002>
- R Core Team. 2016. R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from: <https://www.R-project.org/>.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, 77. DOI: <https://doi.org/10.1186/1471-2105-12-77>
- Samlan, R. A., & Kreiman, J. 2014. Perceptual consequences of changes in epilaryngeal area and shape. *The Journal of the Acoustical Society of America*, 136(5), 2798–2806. DOI: <https://doi.org/10.1121/1.4896459>
- Sarkar, D. 2008. *Lattice: Multivariate Data Visualization with R*. New York: Springer. Retrieved from: <http://lmdvr.r-forge.r-project.org> ISBN: 978-0-387-75968-5. DOI: <https://doi.org/10.1007/978-0-387-75969-2>
- Sarkar, D., & Andrews, F. 2016. latticeExtra: Extra Graphical Utilities Based on Lattice [Computer software manual]. Retrieved from: <https://CRAN.R-project.org/package=latticeExtra> (R package version 0.6-28).
- Schielzeth, H. 2010. Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, 1(2), 103–113. DOI: <https://doi.org/10.1111/j.2041-210X.2010.00012.x>
- Shahin, K. N. 1996. Accessing pharyngeal place in Palestinian Arabic. In: Parkinson, M. E. D. (Ed.), *Perspectives on Arabic Linguistics IX: Papers from the Ninth Annual Symposium on Arabic Linguistics*, 131–150. Amsterdam/Philadelphia: John Benjamins. DOI: <https://doi.org/10.1075/cilt.141.10sha>
- Shahin, K. N. 1997. Acoustics of pharyngealization vs. uvularization harmony. In: *Perspectives on Arabic Linguistics: Papers from the Annual Symposium on Arabic Linguistics*, 10, 215–237. John Benjamins. DOI: <https://doi.org/10.1075/cilt.153.14sha>
- Shahin, K. N. 2011. Pharyngeals. In: van Oostendorp, M., Ewen, C. J., Hume, E., & Rice, K. (Eds.), *The Blackwell Companion to Phonology*, 604–627. Blackwell Publishing.
- Stevens, K. N. 1977. Physics of laryngeal behavior and larynx modes. *Phonetica*, 34(4), 264–279. DOI: <https://doi.org/10.1159/000259885>
- Stevens, K. N. 1989. On the quantal nature of speech. *Journal of Phonetics*, 17, 3–45. Retrieved from: <http://www.talkingneanderthal.com/Stevens%281989%29.pdf>.
- Stevens, K. N. 1998. *Acoustic Phonetics*. MIT Press.
- Stevens, K. N., & Keyser, S. J. 1989. Primary Features and Their Enhancement in Consonants. *Language*, 65(1), 81–106. DOI: <https://doi.org/10.2307/414843>
- Story, B. H. 2016. The Vocal Tract in Singing. In: Welch, G., Howard, D., & Nix, J. (Eds.), *The Oxford Handbook of Singing*. DOI: <https://doi.org/10.1093/oxfordhb/9780199660773.013.012>
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. 2008. Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1), 1–11. DOI: <https://doi.org/10.1186/1471-2105-9-307>
- Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 1–25. DOI: <https://doi.org/10.1186/1471-2105-8-25>

- Strobl, C., Malley, J., & Tutz, G. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14(4), 323–348. DOI: <https://doi.org/10.1037/a0016973>
- Sundberg, J., & Askenfelt, A. 1981. Larynx height and voice source. A relationship? In: *Quarterly Progress and Status Report, Speech Transmission Laboratory*, Royal Institute of Technology, 22(2–3), 23–36.
- Sundberg, J., & Nordström, P. E. 1976. Raised and lowered larynx – the effect on vowel formant frequencies. In: *Quarterly Progress and Status Report, Speech Transmission Laboratory*, Royal Institute of Technology, 17, 35–39.
- Sylak, J. 2011. Pharyngealization in Chechen is gutturalization. In: *Annual Meeting of the Berkeley Linguistics Society*, 37(2), 81–95. Retrieved from: [journals.linguisticsociety.org/proceedings/index.php/BLS/article/download/812/595](http://journals.linguisticsociety.org/proceedings/index.php/BLS/article/download/812/595).
- Sylak-Glassman, J. 2014a. *Deriving Natural Classes: The Phonology and Typology of Post-Velar Consonants* (Unpublished doctoral dissertation). Berkeley, CA: University of California, Berkeley.
- Sylak-Glassman, J. 2014b. An Emergent Approach to the Guttural Natural Class. *Proceedings of the Annual Meetings on Phonology*, 1–12. DOI: <https://doi.org/10.3765/amp.v1i1.44>
- Syrdal, A. K., & Gopal, H. S. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *The Journal of the Acoustical Society of America*, 79(4), 1086–1100. DOI: <https://doi.org/10.1121/1.393381>
- Tagliamonte, S. A., & Baayen, R. H. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2), 135–178. DOI: <https://doi.org/10.1017/S0954394512000129>
- Thomas, E., & Kendall, T. 2007. *NORM: The vowel normalization and plotting suite*. [Online Resource: <http://ncslaap.lib.ncsu.edu/tools/norm/>].
- Titze, I. R. 2008. Nonlinear source–filter coupling in phonation: Theory. *The Journal of the Acoustical Society of America*, 123(5), 2733–2749. DOI: <https://doi.org/10.1121/1.2832337>
- Titze, I. R., & Story, B. H. 1997. Acoustic interactions of the voice source with the lower vocal tract. *The Journal of the Acoustical Society of America*, 101(4), 2234–2243. DOI: <https://doi.org/10.1121/1.418246>
- Trautmüller, H. 1981. Perceptual dimension of openness in vowels. *The Journal of the Acoustical Society of America*, 69(5), 1465–1475. DOI: <https://doi.org/10.1121/1.385780>
- Trautmüller, H. 1983. *On Vowels: Perception of Spectral Features, Related Aspects of Production, and Sociophonetic Dimensions*. (Doctoral dissertation, Summary of six papers, final versions highlighted in this list/Sammanläggningsavhandling). Stockholm: University of Stockholm. DOI: [https://doi.org/10.1016/0167-6393\(84\)90008-6](https://doi.org/10.1016/0167-6393(84)90008-6)
- Trautmüller, H. 1984. Articulatory and perceptual factors controlling the age- and sexconditioned variability in formant frequencies of vowels. *Speech Communication*, 3(1), 49–61. DOI: [https://doi.org/10.1016/0167-6393\(84\)90008-6](https://doi.org/10.1016/0167-6393(84)90008-6)
- Trautmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, 88(1), 97–100. DOI: <https://doi.org/10.1121/1.399849>
- Versteegh, K. 2001. *The Arabic Language*. Edinburgh University Press.
- Watson, J. C. E. 2007. *The Phonology and Morphology of Arabic*. Oxford University Press.
- Wickham, H. 2007. Reshaping Data with the reshape Package. *Journal of Statistical Software*, 21(12), 1–20. Retrieved from: <http://www.jstatsoft.org/v21/i12/>.
- Winter, B. 2016. *Analysis of rapid prosody transcription experiment*. Retrieved from: [https://github.com/bodowinter/rapid\\_prosody\\_transcription\\_analysis](https://github.com/bodowinter/rapid_prosody_transcription_analysis).

- Wood, S. 1986. The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels. *The Journal of the Acoustical Society of America*, 80(2), 391–401. DOI: <https://doi.org/10.1121/1.394090>
- Yeou, M. 2001. Pharyngealization in Arabic: Modelling, acoustic analysis, airflow and perception. *Revue de La Faculté des Lettres El Jadida*, 6, 51–70.
- Zawaydeh, B. 1999. *The phonetics and phonology of gutturals in Arabic* (Unpublished doctoral dissertation). Bloomington, IN: Indiana University.
- Zawaydeh, B., & de Jong, K. 2011. The phonetics of localising uvularization in Arabic. In: Heselwood, B., & Hassan, Z. (Eds.), *Instrumental studies in Arabic phonetics*, 257–276. John Benjamins. DOI: <https://doi.org/10.1075/cilt.319>
- Zeroual, C., & Clements, G. N. 2015. The feature [pharyngeal]. In: Rialland, A., Ridouane, R., & H. van der Hulst (Eds.), *Features in Phonology and Phonetics: posthumous Writings by George N. Clements and Coauthors*, 247–276. Mouton de Gruyter: Berlin.
- Zeroual, C., Esling, J. H., & Hoole, P. 2011. EMA, endoscopic, ultrasound and acoustic study of two secondary articulations in Moroccan Arabic. In: Heselwood, B., & Hassan, Z. (Eds.), *Instrumental studies in Arabic phonetics*, 277–297. John Benjamins. DOI: <https://doi.org/10.1075/cilt.319>

**How to cite this article:** Al-Tamimi, J. 2017 Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1): 28, pp.1–40, DOI: <https://doi.org/10.5334/labphon.19>

**Submitted:** 25 April 2016    **Accepted:** 20 June 2017    **Published:** 20 November 2017

**Copyright:** © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 