

EÖTVÖS LORÁND TUDOMÁNYEGYETEM

Faculty of Humanities

Doctoral School of Linguistics

Doctoral Programme of English Linguistics

**The Intonation–Phonation Interface in Syrian
Arabic: An Acoustic Study of Voice Quality Across
Prosodic Position and Sentence Modality**

In-house Defence

Candidate: Firas Shbeeb

Supervisor: Dr. G. Kiss, Zoltán

Budapest, 2026

Contents

Chapter One: Introduction	5
1.1 Background and motivation.....	5
1.2 Syrian Arabic: Linguistic background	7
1.2.1 General overview.....	7
1.2.2 Phoneme inventory	9
1.2.3 Syllable structure	12
1.2.4 Stress patterns	13
1.2.5 Morphosyntax.....	18
1.2.6 Phonological processes.....	24
1.2.7 Prosodic phrasing	27
1.3 Research questions	30
1.4 Significance of the study	31
1.5 Overview of the dissertation.....	31
Chapter Two: Literature Review	33
2.1 Theories of intonation.....	33
2.1.1 Autosegmental–metrical (AM) theory.....	33
2.1.2 Alignment and starredness.....	36
2.1.3 Phrasing and edge marking.....	37
2.1.4 Scaling, downstep, and pitch range	38
2.1.5 ToBI conventions and cross-language adaptation	40
2.1.6 Alternative traditions: British “nuclear tone” and IPO/ToDI	41
2.1.7 Implications for Arabic dialects and the present dissertation	43
2.2 Voice quality: Acoustic correlates	47
2.2.1 Gender-related patterns in acoustic voice measures.....	49
2.3 Phonation types and continuum.....	50
2.4 Phonation, prosody, and cross-linguistic perspectives	51
2.5 Phonation, prosody, and cross-linguistic perspectives (Arabic focus)	54
2.6 The interaction of intonation and phonation in Arabic.....	55
2.7 English as a comparative reference for the intonation-phonation interface	58
2.8 Summary and research gap.....	61
Chapter Three: Methodology.....	64
3.1 Introduction	64

3.2 Research design	64
3.2.1 Pilot study	65
3.2.2 Main study	66
3.2.3 Stimulus structure	67
3.2.4 Rationale	68
3.3 Participants	69
3.4 Materials	69
3.4.1 Sample stimuli	71
3.5 Recording procedure.....	71
3.6 Data annotation.....	73
3.7 Acoustic analysis	74
3.7.1 Intonation analysis.....	75
3.7.2 Voice-quality analysis.....	79
3.8 Statistical analysis.....	80
3.9 Limitations of methodology	82
3.10 Summary.....	84
Chapter Four: Pilot Study	86
4.1 Introduction	86
4.2 Methodology.....	87
4.2.1 Participant.....	87
4.2.2 Materials	88
4.2.3 Recording procedure.....	89
4.2.4 Annotation	90
4.2.5 Acoustic analysis	90
4.3 Results	91
4.3.1 Intonation results	91
4.3.2 Voice-quality results	94
4.3.3 Regression results	97
4.4 Discussion.....	99
4.5 Summary.....	101
Chapter Five: Results of the Main Study (Modality and Position)	102
5.1 Introduction	102
5.2 Statements.....	102

5.2.1 Intonational labels in statements.....	103
5.2.2 Jitter (local %)	105
5.2.3 Shimmer (local %).....	106
5.2.4 Harmonics-to-noise ratio (HNR, dB)	107
5.2.5 Summary.....	108
5.3 <i>Wh</i> -questions	108
5.3.1 Intonational labels in <i>wh</i> -questions	108
5.3.2 Jitter (local %)	110
5.3.3 Shimmer (local %).....	111
5.3.4 Harmonics-to-noise ratio (HNR, dB)	112
5.3.5 Summary.....	113
5.4 Yes/no questions	113
5.4.1 Intonational labels in yes/no questions.....	113
5.4.2 Jitter (local %)	116
5.4.3 Shimmer (local %).....	117
5.4.4 Harmonics-to-noise ratio (HNR, dB)	118
5.4.5 Summary.....	119
Chapter Six: Gender Effects in the Main-Study Corpus	121
6.1 Introduction	121
6.2 Specific analytical procedure.....	121
6.3 Statements.....	123
6.3.1 Jitter (local %)	123
6.3.2 Shimmer (local %).....	123
6.3.3 Harmonics-to-noise ratio (HNR, dB)	124
6.4 <i>Wh</i> -questions	125
6.4.1 Jitter (local %)	125
6.4.2 Shimmer (local %).....	127
6.4.3 Harmonics-to-noise ratio (HNR, dB)	128
6.5 Yes/no questions	129
6.5.1 Jitter (local %)	129
6.5.2 Shimmer (local %).....	130
6.5.3 Harmonics-to-noise ratio (HNR, dB)	131
6.6 Summary.....	132

Chapter Seven: General Discussion	134
7.1 Chapter overview.....	134
7.2 Summary of key findings	134
7.3 Conceptual and measurement background for interpreting the outcomes	135
7.4 Discussion by research question.....	138
7.4.1 RQ1: How do prosodic features interact with voice quality measures in Syrian Arabic?.....	138
7.4.2 RQ2: Does prosodic position condition voice quality?	140
7.4.3 RQ3: How does sentence modality shape voice quality metrics?	142
7.5 Methodological considerations and limitations	144
7.6 Broader implications.....	146
7.7 Future research directions.....	148
7.8 Conclusion.....	149
References	152
Appendix A. Supplementary model outputs for §5.2	167
Appendix B. Supplementary model outputs for §5.3	168
Appendix C. Supplementary model outputs for §5.4	169
Appendix D. Supplementary model outputs for §6.3.3	170
Appendix E. Supplementary model outputs for §6.4.1	171
Appendix F. Supplementary model outputs for §6.4.2.....	172
Appendix G. Supplementary model outputs for §6.4.3	173
Appendix H. Supplementary model outputs for §6.5.3	174

Chapter One: Introduction

1.1 Background and motivation

Languages differ profoundly in how they use pitch. In some languages, pitch is an indispensable part of lexical identity and distinguishes word meanings; in others, pitch functions only at the post-lexical level to signal sentence type, information structure, and other discourse meanings. This fundamental contrast underlies the typological distinction between tone languages and intonation languages (Ladd, 2008; Yip, 2002).

Tone languages use pitch contrasts at the lexical level to signal differences in meaning between words. Mandarin Chinese, perhaps the best-known example, has four canonical tones: high level (T1, *mā* ‘mother’), rising (T2, *má* ‘hemp’), low-dipping (T3, *mǎ* ‘horse’), and falling (T4, *mà* ‘scold’). Each tone is an obligatory feature of the morpheme’s pronunciation and must be learned alongside its consonants and vowels. Thai similarly has five tones, typically labeled mid, low, high, rising, and falling (Abramson, 1978). In African languages like Yoruba, tone carries a substantial functional load, marking not only lexical distinctions but also grammatical categories such as tense and aspect (Pulleyblank, 1986). In essence, the defining characteristic of a tone language is that pitch differences serve as lexical contrasts.

Intonation languages, by contrast, do not use pitch for lexical contrast. In languages like English, German, or Arabic, pitch alone never produces minimal pairs in the lexicon. Instead, pitch is used at higher prosodic levels to mark pragmatic and discourse functions. For instance, English intonation distinguishes sentence types and information structure. A falling pitch contour in *She’s coming* signals a declarative statement, whereas a rising contour in *She’s coming?* indicates a question (Pierrehumbert & Hirschberg, 1990). Intonation in English also helps convey focus or contrastive emphasis. German exhibits similar intonational functions, although dialects vary in their preference for rising vs. falling contours (Grabe, 2004). Arabic likewise lacks lexical tone; distinctions in word meaning or grammar are realized through segmental contrasts (consonant/vowel patterns) and morphology, rather than through pitch. Nevertheless, pitch plays a central role in Arabic intonation, signaling sentence modality (e.g. question vs. statement), managing turn-taking, and adding paralinguistic information such as emotional tone (Hellmuth, 2006; El Zarka, 2017).

Although the tone–intonation contrast is convenient, it is not an absolute divide. Many tone languages also exhibit post-lexical intonational patterns. For example, even in Mandarin Chinese (a tone language), declarative sentences often show final pitch lowering, whereas yes–no questions typically end in a final rise (Xu, 2005). Conversely, some intonation languages develop lexicalized pitch contrasts. Stockholm Swedish, for instance, has a two-way pitch accent distinction that differentiates words, functioning somewhat like a tonal contrast (Bruce, 1977). These examples illustrate that “tone” and “intonation” are better viewed as end-points of a continuum rather than mutually exclusive categories.

Recent work in phonetics and laboratory phonology has highlighted another dimension of prosody: the interaction between intonation and phonation (voice quality). In many languages, laryngeal settings such as modal (normal voicing), creaky (laryngealized), and breathy phonation covary with F0 and contribute to prosodic contrasts, sometimes described in terms of “register” systems (Esposito & Khan, 2020). Perceptual evidence also indicates that voice-quality cues can influence pitch-height judgments, suggesting that phonation can function as an enhancement cue in the interpretation of pitch targets (Kuang & Liberman, 2018). For example, in English, glottalized voice quality, often realized as creaky phonation, is frequently observed at prosodic boundaries and in low-pitched phrase-final contexts (Redi & Shattuck-Hufnagel, 2001). Danish stød is commonly described as involving laryngealization (often realized as creaky phonation) and has a well-established role in Danish phonology and prosody (Basbøll, 2005). Importantly, however, the relationship between phonation and F0 is not uniform across languages or even across phonation contrasts: for Gujarati, Khan (2012) argues that F0 does not provide a systematic enhancement cue for the breathy–modal contrast.

Syrian Arabic (SyrA) has been analyzed for its intonation in the Autosegmental–Metrical (AM) framework (Al Hasan & Mahanta, 2022), but little is known about whether and how phonation interacts with its intonation. English, by contrast, was the original basis for the development of the AM framework and the ToBI transcription system (Tones and Break Indices); as a result, the intonation–phonation interface in English is comparatively well documented. Examining Syrian Arabic alongside English thus provides an opportunity to test cross-linguistic patterns in intonation–phonation interaction. For instance, does SyrA exhibit similar tendencies to English and other languages in showing systematic covariation between intonational structure and phonation-related acoustic cues? If so, that would suggest these intonation–phonation patterns reflect a universal tendency; if not, any

differences would point to language-specific adaptations. On this basis, the dissertation has two linked aims: (i) to provide a comprehensive description of Syrian Arabic intonation, and (ii) to examine whether phonation-related acoustic cues covary systematically with intonational structure in Syrian Arabic, with English serving as a comparative reference point. These aims are developed in the following sections through the linguistic background, the research questions, the methodological design, and the empirical analysis presented in the subsequent chapters.

1.2 Syrian Arabic: Linguistic background

1.2.1 General overview

Syrian Arabic (SyrA) is a cover term for the Arabic varieties spoken in Syria, which form part of the wider Levantine dialect continuum spanning (at least) Lebanon, Palestine, and Jordan. Within this continuum, Syrian varieties share many structural properties with neighbouring Levantine dialects, while also exhibiting regionally and socially conditioned differences in segmental phonology and prosody. A large proportion of descriptive work labelled “Syrian Arabic” has focused on the Damascus urban variety (often “Damascene Arabic”), including Cowell’s reference grammar, explicitly based on the dialect of Damascus (Cowell, 1964: 1–3).

Because of the status of Damascus as the political and cultural capital, Damascene Arabic is frequently treated, both by speakers and in the literature, as a supra-local reference point. This does not imply that it is linguistically representative of all Syrian dialects, nor that other varieties are peripheral; rather, it reflects sociolinguistic patterns of prestige and diffusion that are common in urban settings. Segmental variables such as the reflex of /q/ illustrate this point: urban Levantine varieties (including Damascene) often realise /q/ as [ʔ], whereas other Syrian varieties may maintain [q] or show alternative reflexes, and speakers’ choices are strongly conditioned by region and social meaning (Shetewi, 2023).

Syrian Arabic is used alongside Modern Standard Arabic (MSA) in a classic diglossic configuration in which MSA is associated with literacy, education, and formal public discourse, while colloquial varieties are used for everyday interaction (Ferguson, 1959). In practice, speakers command repertoires that include shifting styles and intermediate registers; this has implications for phonological description because the variety under analysis must be specified (e.g., read speech vs. spontaneous speech; “MSA-leaning” vs.

colloquial-leaning styles), and because contact with standard and media registers can introduce lexical and phonetic variability.

Within Syria, dialect differentiation is often described in terms of broad urban, rural (sedentary), and Bedouin groupings, though these labels compress substantial internal diversity and should not be treated as discrete types (Watson, 2011). For example, rural varieties may exhibit consonantal features not typical of major urban centres, while Bedouin-associated varieties may preserve distinct reflexes of particular segments (including /q/), and such features can carry social indexicality (e.g., rural features may be locally salient or stigmatized in urban contexts). Also, emphatic consonants such as /ʕ/ and /d̤/ might be realized with varying degrees of pharyngealization across social groups. Given these sociolinguistic facts, this dissertation adopts Damascene Arabic as its primary point of reference, consistent with the major descriptive baseline in the literature, while treating “Syrian Arabic” as a wider dialectal domain rather than a single uniform system.

From a structural perspective, Syrian Arabic shows the core properties typical of Levantine Arabic: a consonant inventory including emphatic and guttural segments, a vowel system with length contrasts, and flexible word order conditioned by information structure and discourse. The prosody of Damascene Arabic, in particular, has attracted attention. Kulk, Odé, and Woidich (2003) characterise its intonation as “almost singing¹,” a rather subjective perceptual description linked to its melodic range and timing patterns, including phrase-final lengthening. This perceptual profile makes Damascene Arabic a useful testing ground for research on the phonetics–phonology interface in intonation, especially where tonal events interact with durational and voice-quality cues.

Recent AM-based work on Syrian Arabic intonation supports the view that tonal events are systematically coordinated with prosodic structure. In particular, Al Hasan & Mahanta (2022) provide a preliminary AM analysis of Damascene Arabic and argue that boundary tones and related prosodic events pattern regularly with sentence-level categories, and that Syrian Arabic appears to employ nuclear pitch accents in close alignment with stressed syllables, unlike Egyptian Arabic, where pitch accents are distributed more freely (Hellmuth, 2006). Complementing this, Al Hasan & Mahanta (2023) highlight the role of phrase-final lengthening (post-tonic lengthening) alongside F0 in shaping sentence-level

¹ A pilot study of Damascene Arabic describes a perceptually salient “singing” quality, which Kulk, Odé, and Woidich (2003) relate mainly to larger pitch range and pre-pausal vowel lengthening, sometimes with a rising movement on the lengthened phrase-final vowel.

prosody in Syrian Arabic. For comparison, phrase-final rises and rise–fall patterns are well-documented in Lebanese Arabic intonation, including work that explicitly analyses phrase-final tonal adjustments in Lebanese Arabic within an intonational framework (Kelly, 2023). These dialectal comparisons motivate a careful, empirically grounded account of what is truly shared across Levantine varieties versus what is specific to Damascene/Syrian Arabic.

In this dissertation, I therefore take Damascene Arabic as the primary variety of analysis, for two principled reasons: first, it provides the most established descriptive baseline for “Syrian Arabic” in the reference literature; and second, it is the Syrian variety for which recent work has begun to develop an AM-based account of intonational structure and its phonetic correlates. At the same time, I treat Damascene not as a representative of all Syrian dialects but as a controlled point of comparison. On this basis, the dissertation pursues two linked objectives: (i) to build upon existing descriptive resources of Damascene Syrian Arabic intonation, and (ii) to test whether phonation-related cues covary with intonational structure in this variety, using English as a comparative reference.

1.2.2 Phoneme inventory

The phonological system of Damascene Syrian Arabic combines properties that are widely shared across Arabic with variety-specific patterns that must be stated explicitly. Like other Arabic varieties, it has a relatively rich consonant inventory, including emphatic (secondary-articulated) consonants and pharyngeal/uvular (“guttural”) segments, alongside a full set of oral stops and fricatives. At the same time, the phonemic inventory and its phonetic implementation show dialect-specific reflexes that distinguish Damascene from other Arabic varieties. The inventory adopted in this dissertation follows Cowell’s (1964, pp. 9–17) description of the Damascus dialect as the primary reference point, with broader Arabic comparisons drawn from Watson (2002) and Ryding (2005).

1.2.2.1 Consonants

Damascene Arabic has 28 consonant phonemes. The inventory includes several laryngeal and posterior places of articulation that are comparatively uncommon in many European languages, notably the pharyngeals /ħ, ʕ/ and the uvulars /q, x, ɣ/, alongside a full set of oral stops and fricatives. It also includes what are known as emphatic (secondary-articulated) coronals /ṣ, ḍ, ṭ, ẓ/. These emphatics are not simply plain coronals with a modified acoustic quality; they involve a robust secondary constriction, often described as pharyngealization and/or uvularization, and they systematically affect adjacent vowels.

Such consonant–vowel interactions are relevant for the present dissertation because they can shift vowel quality and timing in ways that must be controlled when interpreting prosodic patterns and acoustic measures derived from the speech signal (Cowell, 1964; Watson, 2011).

A well-known sociophonetic variable in Levantine Arabic is the reflex of /q/. In Damascene Arabic, /q/ is typically realised as a glottal stop [ʔ] in colloquial urban speech (e.g., /qalb/ ‘heart’ → [ʔalb]), whereas other Syrian varieties may show [q] or [g] depending on region and social alignment (Cowell, 1964; Habib, 2010; Watson, 2011). Importantly, the distribution of these variants is socially meaningful: in many Levantine contexts, [ʔ] functions as an urban/prestige-associated variant, while [q] and/or [g] can index rural or Bedouin-associated identities, though the precise sociolinguistic values and constraints vary by community and style (Habib, 2010).

Beyond /q/, salient Levantine variables include affricated realisations of /k/ (e.g., [tʃ]) in some rural dialects (as in /kalb/ ‘dog’ → [tʃalb]). This feature is widely reported as socially marked and often stereotyped as “peasant” in Levantine Arabic, and it is not part of the mainstream Damascene urban norm (Abdel-Jawad, 2013). Taken together, these patterns underscore that describing a “Damascene inventory” requires both (i) a phonemic baseline and (ii) an explicit statement of which socially conditioned variants are included or excluded in the variety analysed in this dissertation. The consonant inventory is summarised in Table 1 (adapted from Cowell, 1964, pp. 2–8).

Table 1 Consonant inventory of Damascene Arabic (adapted from Cowell, 1964: 2–8)

	labial	velarized labial	plain coronal	emphatic coronal	velarized coronal	palatal	velar	uvular	pharyngeal	laryngeal
Stops	b	<u>b</u>	t d	ṭ ḍ			k g	q		ʔ
Fricative	f		s z	ṣ ẓ		ʃ ʒ		x ʁ	ħ ʕ	h
Liquid			l, r		ḷ, ṛ					
Nasal	m	<u>m</u>	n		<u>n</u>					
Glide	w					j				

1.2.2.2 Vowels

The vowel system is less complex than the consonant system and, in Damascene Arabic, is commonly analysed in terms of five vowel qualities /i e a o u/. Phonologically, however, the contrast is asymmetric across length: Damascene Arabic contrasts five long vowel phonemes /iː eː aː oː uː/, whereas only /i a u/ have short phonemic counterparts, with short [e o] and [ə] typically analysed as predictable allophonic variants (Cowell, 1964; Almbark & Hellmuth, 2015). Cowell (1964) further notes systematic allophonic modification of vowels adjacent to emphatics, including lowering and centralisation, which contributes to the perceptual salience of emphasis through consonant–vowel coarticulation.

Vowel quality also shows dialect- and style-conditioned variation. In Damascene Arabic, short high vowels may be realised with laxer qualities (e.g., /i/ ~ [ɪ], /u/ ~ [ʊ]), particularly in unstressed positions, whereas long vowels are generally more stable in quality (Cowell, 1964; Almbark & Hellmuth, 2015). Figure 1 schematises the relevant vowel space as a descriptive reference (adapted from Almbark & Hellmuth, 2015).

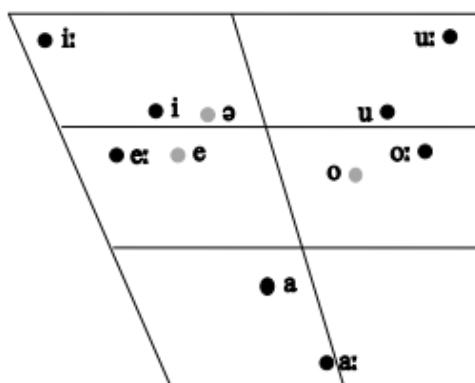


Figure 1 Vowel system of Damascene Arabic (adapted from Almbark & Hellmuth, 2015)

1.2.3 Syllable structure

The syllable structure of Damascene Arabic is described in detail by Cowell (1964, pp. 21–33). Relative to Classical/Standard Arabic, Damascene permits a wider range of surface consonant clustering (notably word-final clusters), while still maintaining tighter constraints than many North-African (Maghrebi) varieties, where extensive vowel deletion can yield highly complex clusters (Cowell, 1964; Dell & Elmedlaoui, 2002; Watson, 2002). At the level of individual syllables, the core types reported for Damascene include CV, CVC, CVV, CVVC, and CVCC, where CVVC and CVCC are commonly treated as superheavy (Cowell, 1964; Watson, 2002).

A widely reported property of Arabic syllable phonology is the avoidance of onsetless syllables in underlying representations. In Damascene Arabic, vowel-initial word forms are typically realised with an onset in careful speech, often through prothetic [ʔ] (frequently corresponding to orthographic hamza), as in /ʔism/ ‘name’. In more casual or connected speech, the presence and phonetic salience of [ʔ] may be reduced, and the satisfaction of avoiding onsetless syllables can interact with resyllabification across word boundaries (Cowell, 1964; Watson, 2002). This pattern contrasts with English, where vowel-initial lexical forms are permitted, although their phonetic realization may involve glottalization or other hiatus-management strategies in particular contexts.

Word-final superheavy syllables are permitted and are common in everyday vocabulary, especially in phrase-final position (e.g., /be:t/ ‘house’ as CVVC, /bint/ ‘girl’ as CVCC). Illustrative single-syllable shapes (Cowell 1964: 21–33) include:

- | | | |
|---------|--------|---------|
| (1) CV: | /la/ | ‘no’ |
| CVC: | /ʔism/ | ‘name’ |
| CVV: | /fi:/ | ‘in’ |
| CVVC: | /be:t/ | ‘house’ |
| CVCC: | /bint/ | ‘girl’ |

Syllable weight is closely tied to stress assignment in Levantine Arabic. In Cowell’s description of Damascene, final superheavy syllables attract stress; otherwise, stress typically falls on a heavy penult, and if no heavy syllable is available near the right edge, stress retracts leftward (Cowell, 1964). Thus, /mak'tu:b/ ‘written’ receives final stress (superheavy CVVC), while /'maktab/ ‘office’ receives penultimate stress (heavy CVC). These generalizations connect phonotactics (what syllables are permitted) with prosody

(where stress is assigned), and they also motivate careful control of segmental structure in prosodic analyses.

These structural facts matter for intonation for a straightforward phonetic reason: phrase-final syllables in Damascene are often heavy or superheavy and are frequently subject to pre-pausal lengthening, providing an extended acoustic interval in which boundary-related F0 movements can be realized (Kulk et al., 2003; Al Hasan & Mahanta, 2023).

In sum, Damascene Arabic allows light, heavy, and superheavy syllables, permits final consonant clusters, and avoids onsetless syllables in careful speech via onset-satisfaction strategies. The system is described in Cowell (1964, pp. 21–33) and can be related to moraic treatments of Arabic syllable weight (McCarthy, 1979; Kiparsky, 2003). These properties provide the necessary groundwork for the stress and intonation patterns analysed in later chapters.

1.2.4 Stress patterns

Stress in Damascene Arabic is widely described as predictable and weight-sensitive, and it plays a central role in prosodic organization at the word and phrase levels. In contrast to English—where stress may be lexically contrastive—stress placement in Damascene is typically analysed as deriving from consistent metrical generalizations based on syllable weight and position, with only limited scope for lexical specification (Cowell, 1964; McCarthy, 1979; Watson, 2002; Kiparsky, 2003). Because stress is systematically related to syllable structure (light/heavy/superheavy), it provides an essential foundation for analysing higher-level prosody: in AM-style accounts, pitch accents are expected to align with metrically prominent (stressed) syllables, and phrase-final lengthening can interact with stress-bearing syllables in ways that affect the realization and interpretation of intonational events. Accordingly, an explicit account of stress assignment is necessary for the present dissertation’s investigation of the intonation–phonation interface.

1.2.4.1 Descriptive rules

Cowell (1964) provides an early systematic description of stress in Damascene Arabic and treats stress placement as regular and non-contrastive. McCarthy (1980) refines this descriptive picture within a moraic–metrical framework and discusses cases where suffixation and prosodic constituency affect the surface location of stress, while preserving the core generalization that stress is weight-sensitive. In both accounts, syllable weight,

defined primarily in terms of vowel length and coda structure, plays the central role in determining stress placement.

For the purposes of this dissertation, the stress pattern can be summarised descriptively as follows (after Cowell, 1964; McCarthy, 1980):

1. **Stress a final superheavy syllable (CVVC, CVCC).**

- /maktu:b/ → [mak'tu:b] 'written'

2. **Otherwise, stress the penult if it is heavy or disyllabic (CVV, CVC).**

- /mada:ris/ → [ma'da:ris] 'schools'
- /katab/ → ['katab] 'wrote.3SG.M'

3. **Otherwise, assign stress to the antepenult.**

- /maktaba/ → ['maktabe] 'library'

McCarthy (1980) argues that these generalizations extend beyond simple monomorphemic forms and remain largely stable under affixation, insofar as morphological material changes syllable weight and prosodic structure in predictable ways. On this view, stress in Damascene Arabic is best treated as **computed** from prosodic form (syllable weight and footing), rather than stored lexically, while still requiring careful attention to morphologically complex and cliticized environments where syllabification and prosodic domains may alter the surface stress pattern.

1.2.4.2 McCarthy's (1980) contribution

McCarthy's "A note on the accentuation of Damascene Arabic" revisits the stress generalisations reported in earlier descriptive work (e.g., Cowell, 1964) and argues that the Damascene pattern is best captured in a metrical analysis in which foot structure is defined in moraic rather than purely syllabic terms. The paper explicitly compares multiple contemporary metrical proposals. It motivates the mora-counting approach by appealing to morphologically conditioned "anomalies," especially in verbal forms with pronominal suffixation, which require cyclic assignment of metrical structure. In particular, McCarthy discusses pairs such as *fātaḥet* 'she opened' vs. *fātaḥeto* 'she opened it', and *šāafet* 'she saw' vs. *šāafeto* 'she saw him', where stress remains on the earlier syllable. By contrast, forms such as *šāllamet* 'she taught' vs. *šāllamāto* 'she taught him', and *ḥākramet* 'she honoured' vs. *ḥākramāto* 'she honoured him' are more revealing, because the suffixed

forms display penultimate stress and therefore cannot be handled by a simple surface stress rule.

A central move in McCarthy's analysis is to treat the relevant foot in Damascene as bounded by moras: informally, the foot comprises the first mora of the stressed syllable plus up to two following moras, yielding a ternary window effect. This derives the familiar pattern whereby stress falls on a heavy penult when one is available, otherwise on the antepenult, while superheavy final syllables remain stress-attracting. On this account, the regular forms such as *fātaḥeto* and *šāafeto* follow the expected pattern, whereas the anomalous forms such as *šallamāto* require an analysis in which metrical structure is built cyclically rather than assigned only once to the final surface form. This way of stating the generalisation also helps explain why stress does not freely retract beyond the antepenult in longer forms, including items borrowed from Literary/Standard Arabic.

McCarthy further argues that the cyclic analysis is especially valuable because it extends beyond simple unsuffixed forms. For example, he shows that the derivation of forms such as *šāafeto* differs crucially from that of *šallamāto*: in the former, the prosodic structure established earlier remains compatible with the expected stress pattern, whereas in the latter, the inherited cyclic structure leads to penultimate stress in the suffixed form. Overall, McCarthy's contribution is therefore not a new descriptive rule-list as such, but a theoretically explicit account that makes the predictability of Damascene stress a consequence of moraic parsing, foot maximality, and cyclicity, thereby tightening the link between surface stress placement and independently motivated prosodic structure.

1.2.4.3 Theoretical interpretation

McCarthy's work on Arabic prosody situates Damascene stress within a broader view of prosodic structure in which syllable weight is represented moraically and stress follows from metrical parsing rather than lexical specification (McCarthy, 1979, 1980; Kager, 2009). In this approach, light syllables contain one mora, heavy syllables contain two, and superheavy syllables contain three. This provides a principled representation of the weight distinctions that underlie the Damascene stress pattern and allows stress to be analyzed as the outcome of prosodic structure rather than as a memorized lexical property.

McCarthy's (1980) contribution is therefore important not only for the description of Damascene Arabic itself, but also for metrical theory more broadly. By deriving stress from mora-based foot parsing and bounded foot structure, the analysis explains why stress

remains confined to a limited right-edge window and why morphologically complex forms can preserve the effects of earlier prosodic structure. On this view, the familiar antepenultimate default is not an extra stipulation, but a consequence of quantity-sensitive footing under prosodic constraints.

More generally, this analysis connects Damascene Arabic to wider typologies of quantity-sensitive stress systems and helps explain why the language has often been treated as a theoretically informative case in discussions of moraic structure, bounded stress windows, and cyclicity in metrical phonology (McCarthy, 1980; Kager, 2009; Kiparsky, 2003).

1.2.4.4 Comparative perspective

Stress assignment in Damascene Arabic broadly patterns with other Levantine varieties in being quantity-sensitive and constrained to a limited right-edge window (i.e., stress falls within the last few syllables under most analyses), although the precise details can vary across communities and speech styles (Watson, 2002; Kager, 2009; Kelly, 2021). In this sense, Damascene aligns closely with Lebanese Arabic, for which recent work also reports a weight-based system prioritizing final superheavy syllables and otherwise targeting heavy penults within a right-edge window (Kelly, 2021; Kager, 2009).

Egyptian Arabic is likewise weight-sensitive, but comparative descriptions highlight differences in how dialects treat certain right-edge configurations and “window” effects. For example, Kager (2009) notes systematic differences between Cairo and Damascus in words of comparable shape: the word *madrása* ‘school’ is stressed on the penult in Cairo, whereas *mádrase* ‘school’ is stressed on the antepenult in Damascus, indicating that dialects may diverge in how they compute stress within the final three syllables and how they handle particular sequences of light syllables near the right edge. These contrasts matter because they show that “weight sensitivity” is not a single uniform algorithm across Arabic dialects, but a family of closely related systems with dialect-specific parameter settings.

North-African varieties such as Moroccan Arabic present an additional point of comparison because extensive vowel reduction and syncope can yield highly complex consonant clusters, including medial CCC sequences, which complicate syllabification and the phonetic diagnostics of word-level prominence (Dell & Elmedlaoui, 2002). Accordingly, the literature does not converge on a single, simple stress algorithm for Moroccan Arabic comparable to Levantine; rather, stress and prominence have been argued to interact with

segmental reduction and prosodic structure in ways that can obscure straightforward weight-based generalisations (Dell & Elmedlaoui, 2002; Boudlal, 2004; Bruggeman et al., 2021). McCarthy (1980) uses cross-dialectal comparison of this kind to motivate a formally explicit metrical account of Damascene stress and to highlight Damascene as a particularly transparent case for theory-building in quantity-sensitive stress.

1.2.4.5 Stress and intonation

In Damascene Arabic, stress functions as a core component of prosodic structure and is directly relevant to intonational analysis. Within Autosegmental–Metrical (AM) approaches, pitch accents are typically modelled as associating with metrically prominent (stressed) syllables, and the predictability of stress in Damascene therefore provides a stable structural basis for locating potential tonal association sites (Cowell, 1964; Watson, 2002; Ladd, 2008). In practical terms, this means that when analysing F0 contours in Damascene Arabic, tonal events are expected to be coordinated with stressed syllables and prosodic boundaries rather than distributed arbitrarily across the segmental string.

Evidence from closely related Levantine varieties supports this general expectation. In her AM analysis of Lebanese Arabic, Chahal (1999, 2001) argues that stressed syllables are crucial for intonational description because they constitute the principal landing sites for pitch accents, while edge tones are independently associated with prosodic phrase boundaries. Acoustic work further shows systematic differences between stress, pitch accent, and nuclear accent in Lebanese Arabic, consistent with the claim that intonational prominence is built on metrical structure rather than on unstressed material (Chahal, 2003).

For Syrian Arabic, recent AM-based analyses likewise treat pitch accents as associated with stressed syllables and report that tonal alignment varies as a function of stress position within the accented word (Al Hasan & Mahanta, 2022). In addition, phrase-final timing patterns—especially post-tonic/phrase-final lengthening—provide a phonetic substrate that can facilitate the realization of boundary-related F0 movement in phrase-final positions (Kulk et al., 2003; Al Hasan & Mahanta, 2023). Taken together, these findings motivate treating stress and segmental timing as conditioning factors when analysing tonal alignment and boundary tones in Damascene Arabic.

Comparable conclusions have been reached for Egyptian Arabic. Hellmuth’s AM analysis treats Egyptian Arabic as a stress-accent language with post-lexical pitch accents and shows that pitch-accent alignment and distribution are sensitive to stressed syllable

structure (Hellmuth, 2006). The cross-dialectal convergence is therefore not that all Arabic dialects share identical stress rules, but that intonational events are systematically constrained by metrical structure: stress provides a recurrent locus for pitch-accent association, while prosodic phrase edges host boundary-related tonal events (Chahal, 1999, 2001; Hellmuth, 2006; Al Hasan & Mahanta, 2022).

1.2.5 Morphosyntax

The morphosyntax described in this dissertation is that of Damascene Syrian Arabic, understood as the educated urban variety centered on Damascus. The classic descriptive foundation remains Cowell's grammar (1964), which explicitly identifies its object of description as the everyday Arabic of educated city-dwelling Syrians, particularly natives of Damascus, while also noting its close relationship to other urban Levantine varieties. More recent syntactic and prosodic work likewise treats Damascene Syrian Arabic, or closely related Syrian Arabic varieties, as the relevant empirical domain for clause structure, interrogative syntax, and prosodic organization (Sulaiman, 2016; Al Hasan & Mahanta, 2022). This matters for the present dissertation because morphosyntax is not treated here solely as background grammar. Rather, subject expression, word order, left-peripheral structuring, and sentence modality all affect the overt prosodic skeleton of the clause and therefore condition the distribution of prominence and phrasing that later chapters investigate (Aoun, Benmamoun & Choueiri, 2010; Al Hasan, 2022).

1.2.5.1 Agreement system and null subjects

Finite verbs in Damascene Syrian Arabic inflect for person, number, and gender. Descriptive accounts of the dialect show that the agreement system is morphologically rich, but also that it includes some syncretism. This is important for the present dissertation because agreement morphology often provides substantial information about the subject, but does not always identify the referent uniquely without discourse support. In recent syntactic work on Syrian Arabic, rich agreement is likewise treated as a central property of the clause, but not as sufficient in itself to explain the interpretation of unexpressed subjects. A more accurate characterization is therefore that subject reference in Damascene Syrian Arabic is commonly recoverable through the interaction of verbal agreement and discourse context, rather than through agreement morphology alone (Cowell, 1964; Sulaiman, 2016). Table 2 illustrates the perfective paradigm of *katab* 'write' (selected forms):

Table 2 Perfective agreement paradigm (Damascene Syrian Arabic): *katab* ‘write’

Person/Number	IPA	Gloss
1SG	[ˈkatabt]	I wrote
2SG.M	[ˈkatabt]	You (m.sg) wrote
2SG.F	[ˈkatabti]	You (f.sg) wrote
3SG.M	[ˈkatab]	He wrote
3SG.F	[ˈkatabat]	She wrote
1PL	[ˈkatabna]	We wrote
2PL	[ˈkatabtu]	You (pl) wrote
3PL	[ˈkatabu]	They wrote

Two points follow directly from this paradigm. First, the morphology distinguishes most person–number–gender values overtly. Second, it also shows syncretism²: 1SG and 2SG.M both surface as *ˈkatabt*. This means that the verb often narrows the interpretation of the subject considerably, but does not always determine it uniquely. In practice, interpretation depends on agreement together with the surrounding discourse context (Cowell, 1964).

This can be illustrated with the form *ˈkatabt*, which is compatible with either first singular or second singular masculine interpretation:

1. [ˈkatabt ədˈdars]
write.PFV.1SG/2SG.M DEF-lesson
‘I wrote the lesson’ / ‘you (m.sg.) wrote the lesson’ (subject pronoun omitted)

In isolation, the verbal form does not distinguish these two readings. The intended subject must therefore be recovered from the discourse situation. If an independent pronoun is added, the ambiguity disappears immediately:

2. [ˈʔana ˈkatabt ədˈdars]
1SG write.PFV.1SG DEF-lesson
‘I wrote the lesson.’
3. [ˈʔanta ˈkatabt ədˈdars]
2SG.M write.PFV.2SG.M DEF-lesson
‘You wrote the lesson.’

² Syncretism refers to a situation in which different grammatical categories are expressed by the same morphological form. In the present context, for example, the Damascene Syrian Arabic form *ˈkatabt* can correspond to both 1SG and 2SG.M, so the two categories are morphologically syncretic.

A further point is that not all agreement forms are equally ambiguous. For example, plural verbal morphology typically identifies the grammatical subject more narrowly:

4. [raħ 'jiktbu əd'dars]

FUT write.IPFV.3PL DEF-lesson

‘They will write the lesson.’

The significance of overt subject pronouns in Syrian Arabic is therefore primarily discourse-pragmatic rather than purely grammatical. As Cowell’s grammar makes clear, since the verbal predicate already encodes subject agreement, an independent pronoun is usually unnecessary unless the speaker wishes to convey contrast, emphasis, or greater explicitness (Cowell, 1964). This is consistent with the expected pattern in a null-subject language, where omission of the subject pronoun is structurally normal and overt expression is often pragmatically marked. More generally, theoretical work on null subjects has shown that the relation between rich agreement and subject omission is not mechanically one-to-one, and that the formal licensing of null subjects must be distinguished from the interpretation of their reference in discourse (Chomsky, 1981; Rizzi, 1982). Syrian Arabic is therefore best described as a null-subject, rich-agreement variety in which subject reference is commonly recoverable through the interaction of verbal morphology and discourse context. For the purposes of the present dissertation, this matters because the presence or absence of an overt subject changes the clause’s overt prosodic organization and therefore affects which constituents are available to bear prosodic prominence.

1.2.5.2 Word order, subject expression, and discourse structure

Damascene Syrian Arabic allows both VSO and SVO orders, but the alternation is not best treated as free variation. Recent work on Syrian Arabic syntax argues that, in the past tense, VSO is the unmarked order, while SVO is a common alternative whose distribution is constrained by the discourse and interpretive properties of the preverbal subject. In particular, preverbal subjects tend to be definite, specific, or otherwise discourse-prominent, which means that SVO order is often associated with topic-like or information-structurally marked interpretations rather than with a purely neutral subject-first order (Sulaiman, 2016). This is consistent with broader analyses of Arabic clause structure and subject placement, which likewise relate preverbal subjects to properties of definiteness,

specificity, and information structure rather than with a purely neutral subject-first order (Aoun et al., 2010). Consider the following examples:

5. ['katab l'walad əd'dars]
write.PFV.3SG.M DEF-boy DEF-lesson
'The boy wrote the lesson.' (VSO)
6. [l'walad 'katab əd'dars]
DEF-boy write.PFV.3SG.M DEF-lesson
'The boy wrote the lesson.' (SVO)
7. ['katab əd'dars]
write.PFV.3SG.M DEF-lesson
'(He) wrote the lesson.' (null subject)

Example (5) illustrates the verb-initial order often treated as structurally unmarked, while (6) shows the preverbal-subject pattern that is frequently associated with a more discourse-prominent or topic-like subject. Example (7) shows that an overt subject need not appear at all when it is recoverable. The key point for the present dissertation is that these alternatives change the number and order of overt prosodic words and thereby affect where prominence may be realized.

Cowell's descriptive grammar (1964) points in the same direction. He notes that subjects may precede the verb under particular structural and discourse conditions, including cases in which the subject phrase is relatively long or made prominent for emphasis. He also treats ordinary subject–predicate clauses as closely related to topic–comment organization. Taken together, these observations support a central claim for the present dissertation: subject realization and subject placement in Damascene Syrian Arabic are tightly linked to information packaging, and this in turn affects where intonational prominence is likely to be realized in the clause.

1.2.5.3 Topicalization, left-dislocation, and focus

A further important aspect of Damascene Syrian morphosyntax is the productivity of topic–comment and left-dislocation³ structures. Cowell (1964) describes constructions in which a constituent is placed in the left periphery and resumed by a pronoun inside the following clause, producing a structure comparable to left-dislocation. Such constructions

³ Left-dislocation is related to topicalization, but is more specific: the left-peripheral constituent is resumed by a pronoun or other coreferential element in the next clause.

are not limited to one clause type and are best understood as part of a broader discourse-organizing system in which a left-peripheral constituent is made especially salient. This makes them directly relevant for prosodic analysis, since a constituent placed in the left periphery is a natural candidate for prominence, especially when it carries topic- or focus-related interpretation. Here such an example:

8. [əd'dars 'katabt-u]
 DEF-lesson write.PFV.2SG.M-3SG.M.OBJ
 ‘The lesson, did you write it?’

Here the fronted object *əd'dars* ‘the lesson’ is placed at the left edge of the clause and resumed on the verb by the object suffix *-u*. A comparable illustrative subject-dislocation pattern is given in (9):

9. [l'walad huwwa 'katab əd'dars]
 DEF-boy 3SG.M write.PFV.3SG.M DEF-lesson
 ‘The boy, he wrote the lesson.’

In (9), the subject is first introduced in the left periphery and then resumed inside the clause by the independent pronoun *huwwa*. In both examples, the left-peripheral constituent is discourse-prominent and naturally available for prosodic highlighting. Because such structures alter the overt arrangement of constituents in the clause, they are also relevant to the distribution of pitch accents and nuclear prominence.

The relevance of these constructions becomes clearer when they are considered together with recent prosodic work on Syrian Arabic. Al Hasan and Mahanta (2022) report that VSO utterances with neutral focus assign pitch accents to all content words, whereas SVO utterances with subject focus place nuclear prominence on the subject and reduce the prominence of following material; likewise, VOS utterances with object focus place nuclear prominence on the object while the subject is deaccented. Their results therefore support the broader claim that morphosyntactic configuration and information structure jointly constrain prominence distribution in Syrian Arabic. In this sense, topicalization, left-dislocation, and focus are not marginal syntactic options, but part of the structural background that shapes later prosodic analysis.

1.2.5.4 Sentence modality and clause structure

Because sentence modality is a central factor in the present dissertation, it is useful to distinguish briefly among the clause-structural properties of declaratives, yes/no questions, and *wh*-questions in Damascene Syrian Arabic. Declaratives provide the baseline clause type. As noted above, they may occur in both VSO and SVO orders, and these alternatives interact with discourse structure and prominence distribution. For the purposes of the dissertation, declaratives therefore serve as the structural and prosodic reference point against which the interrogative sentence types are compared.

Yes/no questions are especially important because they are often not distinguished from declaratives by dedicated clause-typing morphology. Cowell (1964) describes the yes/no interrogative in Damascene Arabic as generally sharing the same clause type as the corresponding declarative, with interrogativity signaled primarily through intonation, including rising pitch or a medium-high level contour with final lengthening. More recent summaries of Damascene Syrian Arabic make essentially the same point, namely that yes/no questions are typically formed through intonational means rather than through obligatory interrogative morphology (Al Hasan and Mahanta, 2022). The following published examples from Al Hasan and Mahanta (2022) are useful here:

10. [ˈnaːmet ˈleːlaʔ]

slept.3SG.F Laila

‘Did Laila sleep?’

11. [radˈdeːt ʕal muˈbajlʔ]

answered.2SG on DEF-mobile

‘Did you answer the mobile?’

These examples are useful because they show that the yes/no interrogative can be segmentally very close to a declarative clause, while the interrogative reading is signaled prosodically.

Wh-questions, by contrast, are structurally more distinct. In descriptive and syntactic accounts alike, *wh*-words are normally clause-initial in Syrian Arabic. Cowell (1964) notes that question words usually appear first in simple substitution questions. More formal work by Sulaiman (2016) goes further, arguing that Syrian Arabic *wh*-questions generally require verb–subject order with most *wh*-phrases. This means that *wh*-questions are not simply intonational variants of declaratives: they alter the clause’s left edge and therefore change

the set of prosodically comparable positions available for analysis. This is directly relevant to the present dissertation, since later chapters compare sentence types partly in terms of prosodic position, and *wh*-initial structure restricts the possibility of an utterance-initial lexical target word in ways that do not apply equally to statements and many yes/no questions. The following published examples from Al Hasan and Mahanta (2022) illustrate this point:

12. [we:n 'sa:ken 'ʕa:del?]
 where staying Adel
 ‘Where is Adel staying?’
13. [ʔemta na:m 'ʕali?]
 when slept Ali
 ‘When did Ali sleep?’

These examples make clear that *wh*-questions in Damascene Syrian Arabic differ from yes/no questions not only in meaning but also in clause structure and in the position of the prosodically prominent material. Therefore, sentence modality is not merely a semantic label in this dissertation; it is a morphosyntactic and prosodic organizer of the dataset.

1.2.5.5 Prosodic implications

The morphosyntactic options described above have direct consequences for prosodic analysis. Null-subject realization changes the number of overt prosodic words in the clause; word-order alternations change which constituent appears clause-initially and is thus available for early prominence; left-peripheral topic structures create natural sites for discourse-driven prominence; and modality affects both clause structure and the distribution of boundary-related prosodic effects. For this reason, morphosyntax in Damascene Syrian Arabic should not be treated as a merely contextual component of the grammar. Rather, it places real constraints on the intonational and phonatory patterns examined later in the dissertation. In this sense, Damascene Syrian Arabic is a null-subject, rich-agreement variety in which subject realization, word order, left-peripheral discourse structure, and sentence modality jointly shape the overt prosodic skeleton of the clause.

1.2.6 Phonological processes

In addition to its segmental inventory and stress system, Syrian Arabic exhibits regular phonological processes that can reshape syllable structure and segmental timing in

connected speech. Because the dissertation later evaluates phonation and intonation in running speech, it is important to expand on these processes that may introduce additional consonantal duration, create or remove vocalic material, or alter segmental identity, all of which can affect the phonetic realization of prosodic categories.

1.2.6.1 Definite-article assimilation

As in other Arabic varieties, the definite article in Syrian Arabic triggers a well-known assimilation pattern: the article *ʔal* ‘the’ assimilates to a following sun consonant, producing a surface geminate. The terms sun and moon are the traditional Arabic grammatical labels for the two consonant classes involved in this process: with sun consonants, the /l/ of the article is not realized separately and assimilates to the following segment, whereas with moon consonants, the /l/ is retained (Alfozan, 1989). For example:

14. *al-fams* → [aʃʃams]

DEF-sun

‘the sun’ (assimilation of /l/ to /ʃ/)

15. *al-tuffaḥ* → [attʊfʰaḥ]

DEF-apple

‘the apple’ (assimilation of /l/ to /t/)

16. *al-qamar* → [alʰʔamar]

DEF-moon

‘the moon’ (no assimilation before a moon consonant)

This alternation matters because assimilation lengthens the following consonant and, therefore, affects local timing and syllable structure. Such timing differences are relevant for the present dissertation because they may influence the phonetic realization of prosodic events, including the alignment of pitch movement and the durational support available near phrase edges.

1.2.6.2 Vowel epenthesis in consonant sequences

Syrian Arabic generally avoids highly complex consonant sequences. One repair strategy is the insertion of a short vowel to break up a cluster that would otherwise leave a consonant unsyllabified or yield a disfavored consonant sequence, especially in connected speech. Watson (2002) discusses epenthesis in sequences of three consonants (including across word boundaries) and illustrates the appearance of short epenthetic vowels in such

environments (Watson, 2002; Kiparsky, 2003). The relevant point for later chapters is that epenthesis can create new syllabic nuclei and thus create additional potential docking sites for tonal events—though whether those nuclei are prosodically strong enough to host pitch accents is an empirical question.

17. xubz-na → ['xubizna]
 bread-1PL.POSS
 ‘our bread’ (epenthesis between /bz-n/)

In addition to epenthesis, gemination is another phonological process that bears directly on prosodic analysis, because increased consonant duration can influence syllable weight, temporal organization, and the phonetic realization of prosodic events.

1.2.6.3 Gemination

Gemination is contrastive in Arabic and may arise both lexically and morphologically. In Lebanese Arabic, Khattab (2007) describes gemination as a phonological property of the language and notes that all 27 consonants may occur as geminates. Her materials illustrate medial singleton–geminate contrasts with items such as *tʕabaʔ* ‘dish’ versus *tʕabbaʔ* ‘he persuaded’, alongside experimental pairs such as *katab* ‘he wrote’ ~ *kattab* ‘he caused to write’, and *hakam* ‘he judged’ ~ *hakkam* ‘he gave someone authority to judge’, which were used to investigate the acoustic realization of medial C versus CC sequences. These examples make the core point transparent: the relevant contrast is not an extra syllable, but the greater duration of the consonant itself.

Morphologically, gemination is especially transparent in Form II verbs⁴. Ryding (2005) gives regular Form II examples such as *nazzaf-a* ‘to clean’, *qaddar-a* ‘to appreciate’, *qarrar-a* ‘to decide, determine’, and *sawwar-a* ‘to photograph’. These forms are useful for the present discussion because they show that gemination in Arabic is not only a lexical contrast, but also a regular morphological resource in the verbal system.

Phonetically, Khattab’s study (2007) shows that medial geminate consonants in Lebanese Arabic form durational distributions distinct from singleton consonants, and that consonant duration is the primary cue to the contrast, even though vowel duration also interacts with it proportionally. For the purposes of the present dissertation, the relevance of gemination is that a longer consonantal interval can affect local timing and may also bear on syllable

⁴ Form II verbs are one of the traditional derived verb patterns in Arabic, typically characterized by gemination of the middle root consonant (faʕʕala), and are often associated with causative or intensive meanings.

weight, depending on the analysis adopted. Near prosodic boundaries in particular, increased consonant duration can influence the temporal organization of the syllable and the amount of durational support available for the realization and alignment of pitch movement. Gemination is therefore relevant not only as a segmental contrast, but also as a factor that can shape the phonetic implementation of prosodic structure (Hellmuth, 2006).

1.2.7 Prosodic phrasing

Prosodic phrasing refers to how speech is grouped into hierarchical prosodic units (e.g., prosodic words, phonological phrases, intonational phrases). These groupings matter because they constrain where prominence can occur (stress and pitch accents) and where edge-related tonal events occur (boundary tones), and they shape rhythmic and temporal patterns such as phrase-final lengthening. In Syrian Arabic, prosodic phrasing is particularly relevant because syntactic choices that are common in the dialect—such as variable word order, topicalization, and optional subject expression (pro-drop)—can affect how syntactic structure maps onto prosodic domains. This section outlines the theoretical background and summarizes what is currently known about phrasing in Syrian Arabic and closely related Arabic varieties, with a view to how phrasing conditions the intonational and phonatory patterns examined later in the dissertation.

1.2.7.1 Theoretical background

The study of prosodic phrasing is grounded in Prosodic Phonology, which models speech as organized into a hierarchy of constituents (Selkirk, 1984; Nespor & Vogel, 1986). A common version of the hierarchy is: syllable (σ) $\rightarrow$ foot (Ft) $\rightarrow$ prosodic word (ω) $\rightarrow$ phonological phrase (φ) $\rightarrow$ intonational phrase (ι) $\rightarrow$ utterance (U). Each level is associated with characteristic phonological and phonetic phenomena—for example, stress at the level of the foot/word and boundary-related tonal events at the level of the intonational phrase.

A central question concerns how syntax maps to prosodic phrasing. In end-based or alignment approaches, boundaries tend to be inserted at the edges of syntactic constituents (Truckenbrodt, 1999). In Match Theory, prosodic constituents are expected to correspond to syntactic constituents by default, but mismatches are possible when prosodic well-formedness or information structure overrides a direct match (Selkirk, 2011). This is directly relevant for Syrian Arabic because topicalization and focus can motivate prosodic boundaries that do not straightforwardly mirror syntax.

1.2.7.2 Phonetic diagnostics of phrasing

Across languages, prosodic boundaries are not identified by a single cue, but by a cluster of phonetic signals that often occur together. The most widely used diagnostics are: (i) boundary tones (characteristic F0 movements at the edge of a phrase), (ii) phrase-final lengthening (slowing down and lengthening of segments near the boundary), (iii) pitch reset (a return to a higher or “baseline” pitch at the start of the next phrase), and sometimes (iv) pauses (Beckman & Pierrehumbert, 1986; Jun, 2005). Importantly, these cues are gradient rather than categorical: weaker boundaries may show only slight lengthening with no pause, whereas stronger boundaries are more likely to show multiple cues at once. In addition, the strength of these cues typically varies with boundary type—for example, an intonational phrase (ι) boundary is often more clearly marked than a phonological phrase (φ) boundary.

Within the Autosegmental–Metrical (AM) approach, these diagnostics map onto a basic structural division: pitch accents are associated with metrically strong (stressed) syllables, whereas boundary tones are associated with phrase edges (Beckman & Pierrehumbert, 1986; Jun, 2005). For this reason, the empirical chapters treat prosodic phrasing as an explicit analytic variable: phrasing can determine (a) where tonal events are expected to occur and how they align with stressed syllables, and (b) how timing effects such as phrase-final lengthening shape the realization of F0 and voice-quality measures.

1.2.7.3 Prosodic phrasing in Syrian Arabic

Existing work suggests that Syrian Arabic shows clear prosodic grouping and that phrase edges are marked by systematic F0 and timing patterns. In a pilot study of Damascene Arabic, Kulk, Odé, and Woidich (2003) report distinct phrase-final melodic patterns and highlight the perceptual salience of phrase endings in the dialect. More recent AM-based work proposes an explicit tonal inventory and phrasing structure for Syrian Arabic and reports that sentence modality (e.g., statements vs. yes/no questions) correlates with different phrase-final F0 patterns and with phrase-final timing effects (Al Hasan & Mahanta, 2022; Al Hasan & Mahanta, 2023). Importantly, these studies motivate the view that phrasing in Syrian Arabic is not merely “where pauses happen,” but a structured part of the grammar that constrains where tonal events and temporal lengthening occur.

1.2.7.4 Prosodic boundaries and syntax

Prosodic phrasing often aligns with syntax, but it does not copy it. In many utterances, speakers place prosodic boundaries at points where syntax also has a natural break—such as between coordinated clauses or between a topic constituent and the rest of the clause. These are cases where the syntax–prosody mapping is relatively straightforward: a major syntactic boundary tends to coincide with a stronger prosodic boundary.

At the same time, prosodic phrasing is not determined by syntax alone. Speakers can introduce boundaries—or weaken them—in ways that reflect information structure (topic vs. focus vs. background) and prosodic well-formedness (e.g., avoiding overly long phrases). This means that the same syntactic structure can be phrased differently depending on what is discourse-prominent, and conversely, prosodic boundaries can appear in places where syntax does not require a break. In end-based accounts, this is captured by the idea that prosodic boundaries are preferred at syntactic edges but may be overridden by competing prosodic constraints (Truckenbrodt, 1999). In Match Theory, the default expectation is that syntactic and prosodic constituents correspond, but systematic mismatches are permitted when required by prosodic or discourse factors (Selkirk, 2011). Syrian Arabic is therefore a useful test case because it frequently employs constructions (e.g., topicalization and focus) that are known to trigger such mismatches.

1.2.7.5 Cross-dialectal perspectives

Comparing Syrian Arabic with other Arabic dialects helps separate what is plausibly shared across Arabic prosodic systems from what is dialect-specific. At a broad level, many dialects show (i) stress-based metrical prominence, and (ii) systematic marking of phrase edges through a combination of F0 patterns and timing cues. However, dialects can differ substantially in (a) the prosodic units they motivate (how many phrase levels are needed), (b) how frequently pitch accents occur, and (c) which phonetic cues are most reliable for diagnosing boundaries (Hellmuth, 2006; Jun, 2005).

For Egyptian Arabic, Hellmuth’s AM analysis proposes a structured prosodic system in which tonal events are distributed systematically but not necessarily with the same phrasing assumed for Levantine varieties. In particular, her analysis motivates specific phrase-level domains and constraints on pitch-accent distribution, and she shows that edge marking and accent placement depend on prosodic structure rather than being a simple reflection of syntax or word boundaries (Hellmuth, 2006). The takeaway for the present

dissertation is that even when two dialects are both “stress-accent” and both use boundary-related tonal events, they may still differ in how prosodic domains are defined and how tonal events map onto those domains.

For Moroccan Arabic, the comparison is informative for a different reason. Extensive vowel reduction and complex consonant clustering can make traditional prominence cues less straightforward to interpret, because segmental structure changes the availability and duration of vowel-bearing intervals where F0 is most clearly measurable (Dell & Elmedlaoui, 2002). Recent acoustic work, therefore, emphasizes that diagnosing prominence in Moroccan Arabic requires close attention to duration (and other non-F0 cues) alongside pitch, rather than assuming that F0 excursions alone provide a stable indicator of prominence level (Bruggeman, Roettger, & Grice, 2021). The implication is methodological: across Arabic dialects, the relative weight of F0 vs. timing cues in signaling prominence and boundaries can differ, which matters directly for any dissertation—like yours—that also measures phonation-related variables sensitive to segmental timing.

These cross-dialectal comparisons motivate a balanced stance for Syrian Arabic: Syrian Arabic should be analysed as part of a broader Arabic prosodic landscape, but it still requires a dialect-specific account of (i) which prosodic constituents are needed, and (ii) how those constituents are diagnosed phonetically.

1.3 Research questions

This dissertation examines the interaction between intonation and phonation in Syrian Arabic, with comparative reference to English. The following questions guide the research:

1. How do prosodic features, i.e., pitch accents and boundary tones, interact with voice quality measures (jitter, shimmer, HNR) in Syrian Arabic?
2. Do specific prosodic positions, i.e., initial, medial, and final, correlate with systematic differences in phonation types?
3. How do different sentence modalities (declaratives, yes/no questions, *wh*-questions) affect the distribution of voice quality metrics?

In addition to these main questions, this dissertation proceeds with the following general expectations, drawing on the prior literature discussed in detail in Chapter 2. First, voice-quality measures are expected to vary systematically with prosodic structure rather than

pattern randomly across tokens. Second, prosodic position, especially phrase-final position, is expected to condition measurable differences in jitter, shimmer, and HNR. Third, sentence modality is expected to shape these acoustic patterns because statements, *wh*-questions, and yes/no questions differ in their intonational organization. Finally, the dissertation also considers whether the modality- and position-related patterns observed in the main corpus remain stable across male and female speakers, and whether any apparent gender effects are modified once token-level acoustic covariates such as F0 and intensity are taken into account.

1.4 Significance of the study

This study is significant in several respects, both theoretically and empirically. At the theoretical level, it contributes to our understanding of the interface between syntax, prosody, and phonation, an area that has received relatively little attention, especially as far as Arabic is concerned. While intonation has often been described within the Autosegmental-Metrical framework and prosodic hierarchy theory, these models rarely incorporate systematic phonation data. By examining how Syrian Arabic employs voice quality measures such as jitter, shimmer, and HNR in relation to prosodic structures, the dissertation offers fresh evidence that broadens the scope of intonational phonology. It tests whether established models are sufficient to capture both tonal and phonatory phenomena, or whether an expanded theoretical apparatus is needed.

Empirically, the dissertation fills a clear gap in the documentation of Arabic prosody. Compared with Egyptian or Lebanese Arabic, Syrian Arabic has been under-represented in instrumental studies of intonation and phrasing. The material for this study provides a systematic account of its intonational patterns and their interaction with phonation, thereby enriching the typology of Arabic dialects. The detailed acoustic analysis also makes the data relevant to the wider field of phonetic sciences, where comparable cross-linguistic work is ongoing.

1.5 Overview of the dissertation

Following this introductory chapter, Chapter 2 reviews the relevant literature on intonation, phonation, and the intonation–phonation interface, with particular attention to Autosegmental–Metrical theory, Arabic prosody, and cross-linguistic work on voice-quality correlates. Chapter 3 presents the methodology, including the research design, participants,

materials, annotation procedure, acoustic measures, and statistical modeling strategy. Chapter 4 reports the pilot study and its role in refining the main study. Chapter 5 presents the main results of the dissertation, organized by sentence modality and prosodic position. Chapter 6 extends the main-study analysis by examining gender-related patterns and the role of token-level acoustic covariates such as F0 and intensity. Finally, Chapter 7 discusses the findings in relation to the research questions, methodological considerations, and the broader implications of the dissertation.

Chapter Two: Literature Review

2.1 Theories of intonation

Research on intonation asks how fundamental frequency (F0) patterns—and their coordination with other prosodic dimensions such as duration, intensity, and voice quality—signal linguistic structure and pragmatic meaning beyond segmental content. While F0 is the primary acoustic dimension of intonation, the interpretation of intonational patterns in running speech often depends on their interaction with metrical prominence, phrasing, and information structure (Ladd, 2008; Jun, 2005).

A common starting point is that speech is organized into a prosodic hierarchy (e.g., syllable, foot, prosodic word, phonological phrase, intonational phrase, utterance), though languages differ in which intermediate constituents are required and in how sharply these levels are phonetically marked (Selkirk, 1984; Jun, 2005). Within this general view, intonation is treated as a structured system whose units are constrained by prosodic constituency and by metrical prominence.

2.1.1 Autosegmental–metrical (AM) theory

The dominant framework in contemporary intonational phonology is Autosegmental–Metrical (AM) theory, developed initially for English intonation and subsequently extended cross-linguistically (Pierrehumbert, 1980; Ladd, 2008). AM represents intonation as sequences of discrete tonal events drawn from a small inventory, canonically High (H) and Low (L) tones, which are associated with specific prosodic positions.

In standard AM/ToBI-style analyses, tonal events are typically grouped into:

1. Pitch accents, associated with metrically strong (stressed) syllables (e.g., H*, L*, L+H*), and
2. Edge tones, associated with the edges of prosodic constituents (e.g., phrase accents such as H- / L- in systems that posit an intermediate phrase, and boundary tones such as H% / L% at intonational phrase edges) (Pierrehumbert, 1980; Ladd, 2008).

Crucially, Autosegmental–Metrical (AM) theory offers a way of analysing intonation by distinguishing between what belongs to the phonological representation and what belongs to the phonetic realisation. In Pierrehumbert’s foundational account, the main goal is to develop an abstract representation that can describe what tunes a language allows and how

the same tune can be realised across different stress patterns, while also formalising the rules that map phonological structure onto phonetic output (Pierrehumbert, 1980). This distinction is motivated by the fact that raw F0 contours can vary greatly depending on speech rate, segmental context, and pragmatic emphasis, but such variation does not necessarily mean that different intonational categories are involved.

On the phonetic side, AM assumes that tones are realised as tonal targets and that the continuous F0 contour heard in speech is produced through interpolation between those targets (Arvaniti, 2015). These targets are defined by two main properties. First, scaling refers to the height of a target within the speaker's pitch range. Second, alignment refers to the position of the target relative to TBUs⁵. For this reason, pitch accents are expected to occur with stressed syllables, while edge tones are expected to occur with phrase-final or phrase-initial units such as final vowels (Arvaniti, 2015). Although interpolation is often assumed to be approximately linear, AM analyses also recognize that the contour between tonal targets may be non-linear, including cases where the F0 trace dips or “sags” between two high targets. As a result, apparently flat or plateau-like stretches of the contour can be difficult to interpret: they may represent a genuinely specified tonal target, or they may simply be the phonetic transition between neighbouring targets rather than an independent tonal event (Arvaniti, 2015; Pierrehumbert, 1980;1981).

At this point, it is useful to distinguish alignment from the broader notion of timing. Alignment refers specifically to the relative position of a tonal target with respect to TBUs and other segmental landmarks. Timing, by contrast, includes wider temporal effects caused by speech rate, tonal crowding⁶, and the amount of segmental material available for tonal realisation. The last of these is especially important when voiceless stretches interrupt or obscure the F0 track. Empirical work on segmental anchoring shows that alignment is systematic, but also language- and variety-specific. Atterer and Ladd (2004), for example, show consistent alignment differences between Northern and Southern German and argue that such differences are better analysed as quantitative phonetic realisation rules than as different phonological associations. Ladd (2004, 2006) makes a similar point, arguing that

⁵ TBU stands for tone-bearing unit, that is, the prosodic unit with which a tonal target is coordinated. In autosegmental work on lexical tone, the tone-bearing unit is often the vowel or mora; in intonational analyses, however, it is commonly approximated as the syllable nucleus. Thus, when AM studies discuss alignment, they mean the position of a tonal target relative to the relevant unit in the speech stream, such as the nucleus of a stressed syllable or a phrase-final vowel (Arvaniti, 2015).

⁶ Tonal crowding occurs when tones are packed into a small amount of segmental material, so that their realisation is compressed and their phonetic timing or height may adjust accordingly (Arvaniti, 2015).

segmental anchoring should be analysed as a measurable phonetic pattern, such as earlier or later peak alignment, rather than as an extra symbolic association in phonological structure. In other words, if one language or dialect places an F0 peak slightly earlier or later than another, that does not necessarily mean the tone is phonologically associated in a different way. It may simply mean that the same tonal category is realised with different quantitative timing.

Concrete examples make clear why these distinctions matter for phonological analysis. In English, Pierrehumbert (1980) shows that the same fall–rise melody in *Legumes are a good source of vitamins* is redistributed when focus shifts: when main stress falls on *vitamins* rather than on *good*, the melody is aligned differently, but its key landmarks still remain tied to nuclear stress and the phrase edge. Importantly, in a short utterance, a rise may appear to be a single movement. In a longer utterance, however, the same pattern may be spread out and show several parts with a plateau between them. This demonstrates that the visible contour shape is not the same thing as the underlying phonological tune (Pierrehumbert, 1980).

A comparable stress-conditioned alignment effect has been reported for Syrian Arabic by Al Hasan and Mahanta (2022). In their yes/no-question data, *'na:met 'le:la?* ‘Did Laila sleep?’ illustrates a case in which first-syllable stress in the nuclear-accented word is associated with L^*+H , whereas *rad'de:t ʕal mu'baʃl?* ‘Did you answer the mobile?’ illustrates a case in which stress on the second syllable is associated with $L+H^*$. In Lebanese Arabic, boundary configurations provide clear examples of scaling differences: Chahal and Hellmuth (2014) show that $H-H\%$ forms the high-rising edge typical of yes/no-question tunes, as in *'ʕindak 'jee* ‘Do you have it?’, while $!H-H\%$ represents a downstepped version of the same rising edge, as in *'miin 'ħama 'muna min 'lima?* ‘Who protected Muna from Lima?’. They further argue that contrasts such as $H-H\%$ versus $!H-H\%$, and $H-L\%$ versus $!H-L\%$, are best understood in terms of edge-tone scaling and local pitch-range level.

Finally, the AM distinction between representation and implementation also has important methodological consequences, which help explain why this chapter cross-references related concepts elsewhere. Decisions about implementation, for example, how a target window is defined, how interpolation is handled across voiceless stretches, or whether contours are smoothed to reduce microprosodic effects, can change the measured values of alignment and scaling and may therefore affect the tonal analysis that is ultimately

proposed. Al Hasan and Mahanta (2022), for instance, explicitly report smoothing some contours in order to reduce consonant-related microprosodic effects, while Al Tamimi and Khattab (2018) show that segmental contrasts in Lebanese Arabic interact with measures such as f0 in surrounding vowels, confirming that segmental phonetics can influence the signal used in intonational analysis. For this reason, it is useful to treat association and starredness (2.1.2) separately from scaling and pitch-range phenomena (2.1.4), while keeping both linked to the broader distinction between implementation and representation introduced here.

2.1.2 Alignment and starredness

Once alignment is treated as distinct from broader timing effects, an important question arises: how far can alignment be used to diagnose phonological accent categories? In AM research, this question is closely linked to the notion of starredness⁷. Starredness has been influential because it allows analyses to distinguish accents that may appear broadly similar on the surface but differ in their tonal association. For example, L+H* and L*+H both involve rising movement, but they are not equivalent: in L+H* the H is starred and linked to the stressed syllable, whereas in L*+H the L is starred and the following H is a trailing tone. In this way, starredness provides a way of distinguishing pitch accents not only by contour shape, but by the relation between tonal events and prosodic prominence (Ladd, 2008).

Greek has been central to the debate on this issue because detailed phonetic studies of Greek prenuclear rises showed that the relation between stress and tonal alignment is not always straightforward. In bitonal accents, the L tone aligned around the onset of the stressed syllable, while the H peak was realised later, typically in the following unstressed syllable. As a result, neither tone appeared to coincide with the stressed syllable in a simple way, and this made it unclear which tone should count as starred. The Greek case was therefore important not because it undermined the notion of starredness, but because it showed that starredness cannot always be identified simply by looking for the most obvious turning point in the surface contour. Instead, the analyst must distinguish carefully between phonological association and phonetic alignment (Arvaniti, Ladd, & Mennen, 2000). The main point is that starredness is still useful, but it cannot be identified just by looking at the raw F0 contour. What matters is whether repeated alignment patterns point

⁷ The status of the tone marked with an asterisk as phonologically associated with the metrically prominent syllable

to a real difference between accent categories, or whether they simply reflect small phonetic differences in how the same accent is realised.

2.1.3 Phrasing and edge marking

AM analyses treat prosodic phrasing as a structured component of intonational organisation rather than as a by-product of speaking rate or pausing. In this view, utterances are grouped into prosodic domains of different sizes, and these domains matter because they determine where tonal events are licensed and how the utterance is interpreted as a sequence of grouped units. Prosodic phrasing is therefore meaningful in the sense that it can affect how material is packaged for interpretation: for example, whether a speaker presents part of an utterance as complete or continuing, and how listeners recover the grouping of information across a sentence. Work in prosodic phonology and intonational phonology consistently treats such phrasing as hierarchical, with higher-level domains built from lower-level ones rather than as a flat sequence of local pitch events (Nespor & Vogel, 1986, 2007; Beckman & Pierrehumbert, 1986; Ladd, 2008).

Within AM-style analyses, phrase edges are marked by edge tones, but the exact phrasing inventory differs across languages. In the English-type model developed from Pierrehumbert's work, an intermediate phrase is marked by a phrase accent such as H- or L-, while an intonational phrase is closed by a boundary tone such as H% or L%. Beckman and Pierrehumbert's treatment of English and Japanese helped establish this hierarchical view, and later typological work showed that languages may differ in which intermediate phrase levels they posit and how those levels are phonologically marked. For example, Jun's work on Korean distinguishes an accentual phrase and an intonation phrase, while Beckman and Pierrehumbert's English model distinguishes an intermediate phrase and an intonation phrase. The common point is not that all languages use the same phrase inventory, but that phrase structure is organized into levels and that these levels are reflected in systematic intonational marking (Beckman & Pierrehumbert, 1986; Jun, 1998; Jun, 2005; Ladd, 2008).

This hierarchical organization is also important because different phrase levels are associated with different kinds of phonetic evidence. Major phrase boundaries are typically signaled not by a single cue but by a cluster of cues, including boundary-related pitch movement, phrase-final or pre-boundary lengthening, pitch reset in the following phrase, and sometimes pause. Experimental work on prosodic boundary perception and production

shows that these cues help listeners recover grouping and boundary strength. In one German study, for example, sequences with an internal intonation-phrase boundary showed a much larger pitch rise, substantial final-vowel lengthening, and a pause, whereas corresponding sequences without such a boundary lacked those stronger cues; specifically, the final vowel was about 1.8 times longer before the boundary and the pause averaged about 506 ms in the boundary condition (Wellmann et al., 2012). These patterns were described as consistent with the main acoustic cues associated with German intonational phrase boundaries in spontaneous speech (Peters et al., 2005, as cited in Wellmann et al., 2012). This kind of result helps clarify what it means to say that phrasing is meaningful: phrase edges are not merely abstract labels but are supported by robust phonetic cues that guide interpretation (Beckman & Pierrehumbert, 1986).

Cross-linguistic and cross-dialectal work further shows that phrasing is both structured and variable. Chahal and Hellmuth (2014), for instance, show that Lebanese and Egyptian Arabic share broad AM assumptions about phrasing and edge-tone combinations, but differ in the size of their pitch-accent inventories, in the distribution of pitch accents, and in how focus is realized. This is directly relevant to the present dissertation, because it suggests that Arabic varieties can share general prosodic architecture while differing in the details of phrasing and edge marking. For Syrian Arabic, this means that phrase boundaries cannot be assumed to behave exactly like those of English or even of neighboring Arabic dialects; they must be established from the language-specific patterning of edge tones, final lengthening, and related phonetic cues. This is why later discussion in the dissertation can treat phrase-final lengthening, pitch reset, and edge-tone realization as informative cues to prosodic structure rather than as incidental phonetic variation.

2.1.4 Scaling, downstep, and pitch range

As discussed in the previous section, AM distinguishes the alignment of tonal targets from their scaling⁸. Once the location of a tonal target has been identified, a second question arises: how high or low is that target realised relative to other tonal events in the same utterance? AM analyses treat this question as an important part of phonetic implementation, because two utterances may have the same tonal inventory and similar alignment patterns, yet differ significantly in the relative height of their peaks and valleys.

⁸ The relative tonal height in the speaker's pitch range.

A widely discussed case is downstep, often notated as !H, in which a high tone is realized lower than a preceding high tone because of tonal interaction rather than because it belongs to a different pitch category altogether. This matters because successive high targets in an utterance are not always expected to reach the same level: their relative height may contribute to the interpretation of continuation, completion, or broader discourse organization. More generally, tonal height also interacts with global pitch range and with pragmatic factors such as focus and prominence, so that differences in intonation may involve not only where tones are aligned, but also how high or low they are realized relative to one another (Ladd, 2008).

Real-language examples from Lebanese Arabic make this especially clear. Chahal and Hellmuth (2014) analyse the yes/no question *'šindak jee* ‘Do you have it?’ as H-H%, where the utterance ends in a high-rising edge. They also discuss the narrow-focus *wh*-question *'miin 'ḥama 'muna min 'lima?* ‘Who protected Muna from Lima?’, analysed as !H-H%. In both cases, the phrase ends in a rising configuration, but the second example contains a downstepped high boundary, that is, a final high realized at a lower level than the corresponding non-downstepped high. Chahal and Hellmuth’s discussion is important because it shows that Arabic edge tones cannot be distinguished only by their direction of movement, such as “rising” versus “falling,” but also by their relative height relations within the utterance. They make the same point for contrasts such as H-L% versus !H-L%, arguing that these are best understood partly in terms of edge-tone scaling and local pitch-range level.

Pitch range is also relevant to the realization of focus. Chahal and Hellmuth (2014) show that in Egyptian Arabic, narrow focus is often accompanied by a substantially expanded pitch excursion on the focused constituent, while following material remains accented but is produced in a compressed pitch range. A comparable pattern is reported for Syrian Arabic by Al Hasan and Mahanta (2022): in broad focus, content words are accented across the utterance, whereas in narrow focus, the focused constituent carries a high-pitched nuclear accent with an enlarged F₀ range, and the post-focal material is realized in a compressed pitch range. They illustrate this both for subject focus in SVO structures and for object focus in VOS structures. Taken together, these studies show that focus may be expressed not only through the presence or absence of pitch accents, but also through relative expansion and compression of pitch range across the utterance. In such cases, post-

focal material is not necessarily deaccented categorically; rather, tonal events may still be present, but implemented in a reduced pitch span.

The main point is that phrase-final tonal realization in Syrian Arabic cannot be reduced solely to contour shape or target location. It is also useful to keep in view the possibility that apparent differences between final rises and falls may reflect not only different tonal categories or edge-tone combinations, but also differences in the relative height of otherwise comparable tonal events. In other words, two contours may be similar in alignment but differ in the level at which their phrase-final tones are realized, and such differences may be phonologically or pragmatically relevant.

2.1.5 ToBI conventions and cross-language adaptation

To support empirical analysis within the AM framework, ToBI⁹ provides a set of conventions for annotating pitch accents, edge tones, and prosodic junctures in spoken corpora. The system was originally developed for American English, where it offered a practical way of linking phonological analysis to consistent corpus annotation, and it was later adapted for a range of other languages (Silverman et al., 1992; Beckman & Ayers Elam, 1997). In this sense, ToBI is not a theory separate from AM, but a method for applying AM categories in a transparent and replicable way.

A major strength of ToBI is that it makes analytical decisions explicit. Rather than describing intonation only impressionistically, it requires the analyst to specify which pitch accents are distinguished, which edge tones are recognized, and which levels of phrasing are assumed in the language under investigation. This is particularly useful in comparative work because it encourages consistency in annotation and helps make different analyses more easily interpretable across studies. At the same time, ToBI does not eliminate analytical uncertainty. Its usefulness depends on how well the tonal categories and phrasing levels chosen for annotation match the prosodic system of the language being studied (Ladd, 2008).

For this reason, ToBI systems have been adapted in language-specific ways rather than transferred mechanically from English. A well-known example is GToBI for German, which retains the general AM logic of annotating pitch accents, phrase accents, boundary tones, and break indices, but adapts the inventory and diagnostics to the phonological and

⁹ Tones and Break Indices

phonetic properties of German intonation (Grice, Baumann, & Benzmüller, 2005). Such adaptations show that cross-language ToBI work is not based on the assumption that the same labels automatically have the same meaning in every language. Instead, the value of the system lies in providing a common representational framework within which language-specific contrasts can be described.

This point is especially important for the present dissertation. Since Syrian Arabic does not have a fully standardized ToBI system comparable to those established for languages such as English or German, any annotation of its intonation necessarily involves language-specific analytical choices. As a result, comparison with other languages must be made cautiously: labels such as H*, L+H*, or H-H% may be useful descriptively, but their interpretation depends on the tonal contrasts, alignment patterns, and phrasing structure motivated for Syrian Arabic itself. In this sense, ToBI conventions are best understood here as a structured descriptive tool rather than as a fixed cross-linguistic template. The main analytical priority is therefore not labelling equivalence in itself, but a clear account of what the labels are intended to represent in the Syrian Arabic data.

The analysis of Syrian Arabic intonation in the present study depends on identifying recurring pitch accents and boundary configurations in a way that is systematic enough to support comparison across sentence types, prosodic positions, and speakers. A ToBI-style approach is helpful here because it provides a principled format for organizing the annotation of tonal events and prosodic boundaries. More broadly, this also helps explain why annotation is not a purely technical stage of the analysis. For example, decisions about whether two surface contours should be assigned the same label depend on prior assumptions about the prosodic system being described. For an underdescribed variety such as Syrian Arabic, this explicitness is especially important, since it allows the proposed categories to remain open to refinement while still providing a coherent descriptive basis for the empirical chapters that follow.

2.1.6 Alternative traditions: British “nuclear tone” and IPO/ToDI

AM/ToBI approaches coexist with earlier traditions in intonational analysis, and these traditions remain important because they highlight different ways of describing the relation between pitch movement, meaning, and prosodic structure. Before the widespread adoption of AM theory, English intonation was often described in terms of larger contour types such as fall, rise, and fall–rise, treated as functionally meaningful units associated especially

with the nucleus of the intonational phrase. These approaches offered a rich descriptive vocabulary for linking contour type to discourse meaning, speaker attitude, and pragmatic interpretation, even though they did not always analyse such patterns in terms of discrete tonal events associated with specific prosodic positions (Crystal, 1969; Cruttenden, 1997).

The British tradition, in particular, places strong emphasis on the nuclear tone as the central carrier of intonational meaning. In this approach, distinctions such as fall, rise, fall–rise, and rise–fall are treated as meaningful contour categories that help describe completion, continuation, reservation, contrast, and related discourse functions. This tradition has been highly influential in practical description and pedagogy, especially for English, because it offers an accessible account of how overall contour shape relates to communicative meaning. At the same time, its categories are not always directly comparable to AM pitch accents and edge tones, since the British tradition typically starts from larger contour units and functional interpretations rather than from a decomposition into discrete tonal events linked to prominence and phrasing (Crystal, 1969; Cruttenden, 1997).

A different but equally important line of work is represented by the IPO tradition¹⁰, developed especially for Dutch. Rather than beginning from abstract tonal categories, the IPO model focuses on perceptually salient pitch movements and motivates contour description through stylization, perceptual testing, and manipulation of speech signals. In this tradition, the goal is not simply to label contours impressionistically, but to identify the pitch movements that listeners actually hear as relevant. This makes the IPO approach especially valuable as a bridge between phonetic detail and perceptual interpretation, since it asks which aspects of the contour are functionally salient enough to matter for listeners (’t Hart, Collier, & Cohen, 1990).

ToDI¹¹ occupies an especially interesting position in this domain. Historically, it is informed by IPO-style work on Dutch intonation, but it also adopts a more explicitly tonal representational format for transcription and analysis. In this sense, ToDI shows that older perceptually oriented traditions and AM-style representational approaches are not always sharply opposed; rather, later systems may combine insights from both. What distinguishes these approaches is therefore not simply whether they are “formal” or “informal,” but the

¹⁰ The IPO tradition was developed at the Institute for Perception Research (IPO) in Eindhoven.

¹¹ Transcription of Dutch Intonation

kind of analytical priority they give to contour shape, perceptual salience, functional interpretation, or discrete tonal representation (Gussenhoven, 2005).

These alternative traditions are important because they show that intonation can be approached from more than one descriptive angle. The British tradition is especially useful in reminding us that contour categories are closely tied to discourse interpretation, while the IPO tradition highlights the perceptual salience of pitch movement as a phonetic object of study. AM theory, however, is adopted here because it provides a particularly explicit way of relating intonation to prosodic structure, including the association of tones with stressed syllables and phrase edges, and because it offers a widely used framework for cross-linguistic comparison. This does not mean that other traditions are descriptively inadequate; rather, it means that AM is better suited to the aims of the present dissertation, which require a representation of intonation that can be linked systematically to prominence, phrasing, and cross-linguistic phonological analysis (Ladd, 2008; Gussenhoven, 2005).

2.1.7 Implications for Arabic dialects and the present dissertation

For Arabic dialects, the AM framework provides an explicit way of modelling the relation between stress-based prominence, intonational categories, and dialect-specific tonal inventories. This is particularly important in Arabic because intonational structure must be analysed in relation to stress, phrasing, and information structure, while also allowing for dialectal differences in how tonal categories are distributed and realised. Detailed AM accounts already exist for varieties such as Egyptian Arabic and Lebanese Arabic, where researchers have examined pitch-accent inventories, phrase-edge marking, focus structure, and the distribution of tonal events across different sentence types (Hellmuth, 2006; Chahal & Hellmuth, 2014). These studies show that Arabic dialects may share a broadly similar stress–intonation interface, while still differing in accent density, in the relative frequency of particular pitch-accent categories, and in the way phrase edges are marked.

This point is especially relevant because it warns against treating “Arabic intonation” as a single uniform system. In Egyptian Arabic, Chahal and Hellmuth (2014) show that contrastive focus is often expressed through expanded pitch range on the focused item and compressed pitch range on following material, but without categorical deaccenting; for example, in *'maama bitit 'ʕallim ju 'naani bi-l-'leel* ‘Mum is learning Greek in the evenings’, the post-focal object *ju 'na:ni* ‘Greek’ still bears a pitch movement, although it is

compressed. In Lebanese Arabic, by contrast, narrow focus may be followed either by deaccenting or by strong post-focal pitch-range compression, and the same chapter illustrates this with 'muna 'hamet 'lama min 'lima 'Muna protected Lama from Lima', where non-focal material may be produced in an extremely compressed range or analysed as deaccented in narrow-focus contexts. These differences show that Arabic dialects can share a general AM-style architecture while differing in accent distribution, boundary marking, and focus realization.

2.1.7.1 Al Hasan and Mahanta (2022): A preliminary AM analysis of Damascene Syrian Arabic

A particularly important point of reference for the present dissertation is Al Hasan and Mahanta's (2022) preliminary AM analysis of Damascene Syrian Arabic. The study is significant because it is presented as an initial attempt to describe the dialect's intonational patterns, tonal inventory, and prosodic phrasing within the autosegmental–metrical framework. The authors explicitly note that previous work on Syrian Arabic had been limited to preliminary descriptions and brief comparisons, and that no earlier study had offered a full investigation of the dialect's tonal inventory or prosodic structure. Their aim, therefore, was not only to describe recurring intonational patterns but also to establish a first AM-based account of Damascene Syrian Arabic that could serve as a basis for later research.

Methodologically, the study is based on two experiments using speech from six native speakers of Syrian Arabic. The first experiment included 22 utterances, covering declaratives, questions, continuations, alternatives, surprise, imperatives, and four types of focus. The second experiment consisted of 27 sentences designed around the three major Syrian Arabic word orders VSO, SVO, and VOS, with target items representing penultimate, antepenultimate, and ultimate stress positions. The corpus was segmentally labelled in Praat, transliterated using IPA symbols, and annotated using a ToBI-style transcription. The authors also report smoothing some pitch contours by 10–30% in order to reduce micro-prosodic consonantal effects.

One of the study's main contributions is its proposal for the prosodic phrasing of Syrian Arabic. Al Hasan and Mahanta assume three levels of phrasing in the dialect: the Accentual Phrase (AP), the Intermediate Phrase (ip), and the Intonational Phrase (IP). Boundary identification was based both on perception and on phonetic cues such as F0 reset, pre-

boundary lengthening, pause, and phrase-edge tones. In their analysis, the Syrian Arabic AP is typically realized as $[H^* La]$ in broad-focus contexts, while IP-final APs often show $[!H^* La]$ with a downstepped pitch accent. They also argue for an intermediate phrase marked by H-, illustrated in longer utterances where a higher peak, lengthening, and pause together support a boundary stronger than an AP edge but weaker than a full IP boundary.

In their discussion of declaratives, the authors describe the default Syrian Arabic pattern as an initial rise followed by a final fall. Declaratives typically show pitch accents on content words, with the pitch rising on the first accented syllable and then falling toward the phrase edge. In the controlled data, broad-focus declaratives commonly showed H* pitch accents in each AP together with a final L% boundary tone, while longer utterances displayed a gradual downtrend in successive H peaks. The study also notes that the pitch maximum often extends over the stressed syllable and sometimes into the following syllable, producing plateau-like or hat-like contours.

The treatment of questions is even more directly relevant to the present dissertation. For *wh*-questions, Al Hasan and Mahanta report a rise to the nuclear pitch accent followed by downdrift, with the nuclear accent almost always occurring on the *wh*-word, which occupies clause-initial position in Syrian Arabic. They analyse the *wh*-question AP as L+H* followed by La, and they identify two phrase-final patterns in the tail: either !H-!H% or !H-L%. Their examples include *we:n 'sa:ken 'ʕa:del?* ‘Where is Adel staying?’ and *'ʔemta na:m 'ʕali?* ‘When did Ali sleep?’. A central observation is that Syrian Arabic *wh*-questions often show an exaggerated phrase-final drawl, which produces vowel lengthening and can yield a final mid-rise before the boundary. The authors stress that this part of the model remains preliminary and may need refinement once the pragmatic functions of *wh*-questions are studied in more detail.

For yes/no questions, the authors report a different contour type. These utterances begin with an initial low tone on the nuclear-accented word, followed by a rise to a higher level. The nuclear pitch accent occurs on the first word, typically the finite verb, rather than on the final content word. Crucially, the shape of the nuclear pitch accent varies with stress position: if the first syllable is stressed, the accent is analysed as L*+H; if the second syllable is stressed, it is analysed as L+H*. Their examples include *'na:met 'le:la?* ‘Did Laila sleep?’ and *rad'de:t ʕal mu'bajl?* ‘Did you answer the mobile?’. As in *wh*-questions, the final rise is accompanied by a pre-pausal drawl, which they treat as a characteristic

feature of Damascene question intonation and link to earlier descriptions of the dialect's "singing intonation."

A further important contribution of the paper is its preliminary tonal inventory. On the basis of the controlled corpus, Al Hasan and Mahanta propose six nuclear pitch-accent types for Syrian Arabic: H*, L+H*, L*, L*+H, !H*, and H+H*, together with a more limited H+L* accent in certain declaratives. They also posit phrase accents La and Ha for accentual phrases, H- and !H- for intermediate phrases, and boundary tones L%, H%, and !H% for intonational phrases. Since the paper is explicitly preliminary, these categories should not be treated as fully settled, but they are nevertheless highly valuable because they provide the first explicit AM-based tonal inventory proposed for Damascene Syrian Arabic.

The second experiment extends the analysis to word order and focus structure. In VSO broad-focus utterances, pitch accents occur on every content word. In SVO utterances with subject focus, the subject bears the nuclear pitch accent, while the verb and object are described as deaccented, with La controlling the following stretch and an L% boundary at the right edge. In VOS utterances with object focus, the object carries the nuclear accent, the verb carries a pre-nuclear accent, and the subject is deaccented. The authors report that the nuclear pitch accent in VOS was realized as H+H* in 80% of the utterances, while the pre-nuclear accent was usually H*. They also note that narrow focus is associated with expanded F0 range on the focused constituent and compressed pitch range on post-focal material, which they interpret as suggesting deaccenting. On this basis, they characterize Syrian Arabic as a right-headed language and contrast it with Egyptian Arabic, where post-nuclear accents may continue after the narrow-focus item because every content word tends to receive a pitch accent.

For the present dissertation, the value of Al Hasan and Mahanta's study lies in two areas. First, it provides a language-specific AM account of Syrian Arabic, rather than forcing the dialect into an English-based template. Second, it offers an explicit mapping between tonal events and prosodic locations such as accented syllables, AP edges, intermediate phrase edges, and intonational phrase boundaries. At the same time, because the study is based on a controlled dataset and repeatedly presents its model as preliminary, it is best used here as a strong starting point rather than as a final authority on Syrian Arabic prosodic structure. That makes it especially useful for the present dissertation: it provides a principled framework to build on, while still leaving room for refinement and empirical testing.

2.1.7.2 Relevance for the present dissertation

Seen in this way, AM is useful in the present dissertation not merely as a labelling system, but as the representational framework that identifies where prosodic events are expected to occur in Syrian Arabic. If earlier AM-based work shows that tonal events are anchored to accented syllables and phrase edges, then these locations can serve as principled reference points for investigating whether phonation-related measures such as jitter, shimmer, and HNR vary systematically with prosodic structure. The central question is therefore not simply whether phonation differs across raw F0 shapes, but whether it covaries with specific prosodic categories and positions that have independent phonological motivation. For that reason, AM functions here as the analytical backbone of the dissertation, while the empirical chapters remain responsible for determining how far particular tonal categories and prosodic locations are supported by the Syrian Arabic data itself.

2.2 Voice quality: Acoustic correlates

Waveforms and spectrograms are useful starting points for identifying non-modal phonation, because they can reveal irregular periodicity, aperiodic noise, and changes in harmonic structure. For example, creaky phonation often involves irregular glottal cycles and reduced periodic energy, while breathy phonation typically involves increased turbulent airflow and greater noise energy (Gordon & Ladefoged, 2001). However, visual inspection alone is rarely sufficient, especially in connected speech, where segmental and prosodic variation makes the acoustic signal much less stable than in sustained vowels. The same phonation type may therefore show different surface patterns depending on surrounding consonants, stress, intonation, and speaking rate. For this reason, most instrumental work relies on quantifiable acoustic measures that can be extracted consistently and compared statistically.

A wide range of measures has been proposed for voice quality, but the present dissertation focuses on F0, jitter, shimmer, and harmonics-to-noise ratio (HNR) because they provide complementary information about periodicity, stability, and noise in the signal, and they are widely used in phonetic and clinical voice research (Boersma, 1993; Brockmann et al., 2011; Teixeira et al., 2013; Yumoto et al., 1982). Importantly, these measures are best treated as acoustic correlates of phonation-related properties, not as direct labels of categorical phonation types (e.g., “creaky” vs. “breathy”).

Fundamental frequency (F0) is the acoustic correlate of perceived pitch and reflects the rate of vocal-fold vibration. F0 is central to prosody because intonational events are implemented through changes in F0 over time. Although some phonation types are often associated with particular pitch patterns, for instance, creaky voice often occurs at very low pitch, F0 should not be treated as a direct measure of voice quality, since it is also shaped by vocal-fold tension, aerodynamic conditions, and speaking context (Ladd, 2008). Because F0 is also sensitive to speaker physiology and situational factors, it is typically treated as a prosodic variable that must be interpreted relative to context rather than as a standalone indicator of phonation.

Jitter measures cycle-to-cycle variability in the fundamental period (i.e., small irregularities in the timing of successive glottal cycles). It is commonly used as an index of periodicity stability: lower jitter indicates more stable periodic vibration, while higher jitter indicates greater irregularity (Teixeira et al., 2013). Since irregular periodicity is a common acoustic consequence of some non-modal phonations, jitter can help distinguish more stable modal phonation from segments showing increased irregularity. At the same time, jitter is sensitive to recording quality and to confounds such as F0 and intensity, which motivates careful measurement settings and consistent sampling procedures (Brockmann et al., 2011).

Shimmer measures cycle-to-cycle variability in amplitude (i.e., micro-perturbations in the size of successive cycles). Like jitter, shimmer is often interpreted as reflecting stability, but in the amplitude domain. Higher shimmer values are typically associated with greater amplitude instability and can correlate with increased noise or reduced periodic energy in the signal (Teixeira et al., 2013). As with jitter, shimmer is influenced by contextual factors—especially intensity—and therefore needs consistent recording conditions and careful segmentation (Brockmann et al., 2011).

Harmonics-to-noise ratio (HNR) quantifies the relative amount of periodic (harmonic) energy compared to aperiodic (noise) energy in the voice signal. In general, higher HNR indicates a more periodic, harmonically structured signal, while lower HNR indicates more noise and reduced harmonic structure (Boersma, 1993). HNR has been widely used in clinical and phonetic research because it correlates with perceived hoarseness and with the degree to which noise replaces harmonic structure (Yumoto et al., 1982). In this dissertation, HNR is therefore treated as a complementary measure to jitter and shimmer, capturing periodicity vs. noise at the signal level rather than focusing only on cycle-to-cycle perturbation.

In summary, waveforms and spectrograms are valuable for qualitative inspection, but systematic analysis of phonation–prosody relations in running speech requires acoustic measures that can be extracted consistently. F0 provides the prosodic backbone of intonation, while jitter, shimmer, and HNR provide complementary indices of periodicity stability and noise. These measures are not direct labels of phonation types, but they provide an operational basis for testing whether particular prosodic categories and positions are associated with systematic changes in voice-quality correlates.

2.2.1 Gender-related patterns in acoustic voice measures

Across studies of healthy speakers, gender effects in acoustic voice measures are often mixed rather than uniform. For perturbation measures such as jitter and shimmer, some studies report higher values for one gender under particular recording conditions, whereas others find only small or inconsistent differences; outcomes depend on the task, the language, and measurement settings (Sørensen & Horii, 1983; Teixeira & Fernandes, 2014). One important reason is that these measures are sensitive to local production factors, especially vocal intensity and fundamental frequency (F0), so apparent gender contrasts may partly reflect differences in pitch range or loudness rather than stable phonatory differences alone (Orlikoff & Kahane, 1991; Brockmann et al., 2008, 2011). This point is especially relevant in connected speech, where variation in F0 and intensity is part of the speech event itself rather than an incidental source of noise.

Measures tied more directly to periodicity and noise, such as harmonics-to-noise ratio (HNR), have sometimes shown clearer gender differentiation in normophonic samples. For example, Dehqan et al. (2010) reported higher HNR in female than in male speakers even where jitter and shimmer did not show strong differences, and Shabnam and Pushpavathi (2022) likewise found gender-linked differences in HNR and related voice-quality indices. At the same time, such findings should not be treated as universally stable biological patterns, since the literature also shows that acoustic voice measures are shaped by task, speaking style, and measurement context. In connected speech especially, where pitch, intensity, and segmental structure vary continuously, any apparent gender effect must be tested against local acoustic covariates rather than assumed to reflect voice quality alone.

This issue is particularly relevant for the present dissertation because the main study examines connected Syrian Arabic speech rather than sustained phonation. If the acoustic measures under investigation are themselves influenced by F0 and intensity, then a proper

account of prosody–phonation relations must also consider whether modality- and position-related patterns are homogeneous across male and female speakers, or whether they are partly shaped by gender-linked differences in local acoustic realization. That question becomes especially important for HNR, which has sometimes been treated as a relatively stable gender-sensitive measure in healthy voices, but may behave differently once token-level acoustic covariates are modeled directly in running speech.

2.3 Phonation types and continuum

This section clarifies what is meant by modal, breathy, and creaky phonation, and why these categories are often treated as points on a continuum rather than as sharply separated types. In broad articulatory terms, phonation types differ in how tightly the vocal folds are adducted, how they vibrate, and how much airflow/noise accompanies voicing. Classic descriptions treat modal voice as the default setting in which the vocal folds are sufficiently adducted to vibrate regularly, producing a stable periodic signal (Ladefoged, 1971; Laver, 1980; Catford, 1977).

In breathy voice (murmur), the vocal folds vibrate but are held more open than in modal voice, allowing increased airflow and typically more aperiodic noise in the acoustic signal (Laver, 1980; Gordon & Ladefoged, 2001). In creaky voice (laryngealization), vibration is often irregular and may involve greater constriction and reduced airflow, yielding a signal with less stable periodicity (Gordon & Ladefoged, 2001).

A useful way to conceptualize these differences is Ladefoged’s proposal that phonation types can be arranged on a continuum of glottal stricture, from more open to more closed: roughly voiceless → breathy → modal → creaky → glottal closure (Ladefoged, 1971; Gordon & Ladefoged, 2001). The key point for this dissertation is that phonation is often gradient: speakers can produce “more breathy” or “more creaky” realizations, and the same speaker may shift along the continuum depending on prosodic position, speech style, and communicative goals.

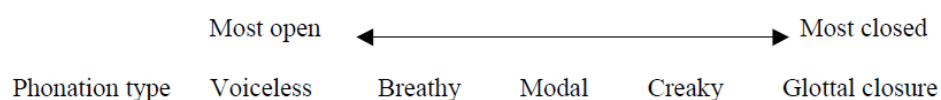


Figure 2 Continuum of phonation types (Ladefoged, 1971)

Many accounts also discuss intermediate settings near modal voice, commonly labelled lax/slack and tense/stiff phonation. These intermediate categories are useful because they

capture differences in laryngeal setting that are not as extreme as prototypical breathy or creaky voice. Importantly, tense and lax phonations are not always simply “halfway to creaky” or “halfway to breathy”: empirical work suggests that tense/lax contrasts can be relatively close to modal voice in their vibratory patterns and may require dedicated diagnostics (Kuang & Keating, 2014; Gordon & Ladefoged, 2001).

Because phonation is gradient and multi-dimensional, categories are not absolute. Ladefoged (1971) already emphasizes that speakers and languages can differ in what they treat as “modal” versus “breathy/creaky,” and that individual baselines vary. This means that it is risky to treat labels like “breathy” and “creaky” as fixed acoustic objects. Instead, they should be treated as phonetic/phonological categories whose realization can vary, and whose interpretation depends on the language system and the speaker.

Phonation differences can also function as social or physiological signals in addition to linguistic ones. For example, voice quality changes with age are widely discussed in the voice literature (Hollien, 1987). Separately, perception and synthesis work shows that listeners are sensitive to systematic voice-quality differences (including breathiness) across talkers and speaking styles, which can become socially meaningful in particular communities (Klatt & Klatt, 1990). For this dissertation, the takeaway is simply that voice quality can be influenced by factors other than linguistic prosody, so analyses must control for speaker and context when interpreting phonation–intonation relations.

Finally, many languages use different regions of the phonation space to express linguistic contrasts. Cross-linguistic overviews document phonation contrasts such as breathy vs. modal such as Hmong (Huffman, 1987), or creaky vs. modal such as Mundurukú (Picanço, 2005), and some languages show complex interactions among phonation, tone, and stress (Gordon & Ladefoged, 2001; Chávez-Peón, 2010). These typological facts motivate treating phonation as a legitimate linguistic dimension, not merely a paralinguistic voice quality effect, and they provide a clear precedent for examining whether Syrian Arabic uses phonation systematically in relation to intonation.

2.4 Phonation, prosody, and cross-linguistic perspectives

A recurring finding in the phonetics literature is that voice quality is not distributed randomly: it often covaries with prosodic structure, including prominence and phrase position. In a pilot study on American English, Epstein (2003) reports systematic differences in voice-quality measures across prosodic contexts, including contrasts between

prominent vs. non-prominent words and between phrase-initial vs. phrase-final prominent words. Importantly, she also reports increased rates of non-modal phonation in association with low boundary tones compared with high boundary tones, suggesting that phonation can function as an additional cue to phrase-edge tonal categories (Epstein, 2003).

The relationship between phonation and prosody becomes more complex in tone languages, where phonation may participate in lexical contrasts and may interact with tone in language-specific ways. For San Lucas Quiaviní Zapotec, earlier descriptive work has been interpreted as proposing a strong association between tone and phonation (Munro & Lopez, 1999). However, later instrumental work argues that tone and phonation influence each other, but tone cannot be fully predicted from phonation alone (Chávez-Peón, 2010). This shift in interpretation matters methodologically: it cautions against assuming that phonation can straightforwardly predict tone without demonstrating that predictability in the phonetic record.

Perceptual work also shows that listeners can weight phonation cues differently depending on the language and the contrast. In White Hmong, Garellek, Keating, and Esposito (2013) show that listeners can use spectral tilt cues—particularly measures such as H1–H2 and related slope manipulations—to perceive breathiness in tone identification tasks, with breathy phonation playing a particularly important role for certain tones (Garellek et al., 2013). A key implication is that phonation may contribute to the perceptual robustness of prosodic contrasts, but the relevant acoustic cues are not necessarily the same across phonation types (e.g., breathiness vs. creakiness) or across languages.

A further complication is that creaky voice itself is not a single acoustic object. Keating, Garellek, and Kreiman (2015) argue that there are several kinds of creak, and that no single acoustic measure identifies all of them equally well. Instead, different measures are more useful for different kinds of creak and in different contexts. As a result, measures such as F0 irregularity, spectral measures, and noise-related measures (including HNR) may pattern differently depending on the subtype of creak and the context (Keating et al., 2015). This point is directly relevant to this dissertation: if phonation is examined instrumentally in running speech, results should be interpreted as changes in acoustic correlates of phonation, not as categorical proof that a segment is “creaky” or “breathy.”

Clinical reference values for perturbation measures are sometimes used in the voice literature as broad benchmarks for identifying potentially pathological phonation. Teixeira,

Oliveira, and Lopes (2013), for example, summarize threshold values shown in Table 3 below for several commonly used measures, including Jitt (1.04%), Jitta (83.2 μ s), RAP (0.68%), Shim (3.81%), and ShdB (0.35 dB). However, these values were compiled for clinical voice assessment and are based on sustained-vowel phonation, not on connected speech. They should therefore not be treated as direct diagnostic cutoffs for the present study, which examines phonation in running speech rather than in time-invariant vocalic tasks. Instead, they are included here only as general reference points that help situate the present acoustic measures within the wider clinical literature on perturbation and voice quality.

Table 3 Clinical reference values for selected perturbation measures in sustained vowels, adapted from Teixeira et al. (2013).

Parameter	Threshold value
Jitt (%)	1.04
Jitta (μ s)	83.2
Rap (%)	0.68
ppq5 (%)	-
Shim (%)	3.81
ShdB (dB)	0.35
Apq3 (%)	-
Apq5 (%)	-

This limitation is important methodologically. In the present dissertation, jitter, shimmer, and HNR are not used diagnostically, and no attempt is made to classify tokens as “normal” or “pathological” on the basis of threshold values derived from sustained vowels. Rather, these measures are analysed as continuous acoustic correlates whose distributions may vary across prosodic position, sentence type, and speaker. This approach is more appropriate because previous work has shown that perturbation measures are affected by factors such as vowel quality, gender, vocal intensity, and F0, even in typical clinical tasks based on sustained vowels. As a result, transferring sustained-vowel thresholds directly to connected speech would risk overinterpreting context-sensitive acoustic variation as if it were stable evidence of voice pathology.

Finally, age-related voice changes have been reported for F0 and perturbation measures, with patterns differing across sexes and across measures (Nishio & Niimi, 2008). These findings motivate treating ‘speaker’ as a critical factor (and typically a random effect)

when testing whether prosodic position or modality is associated with systematic shifts in jitter, shimmer, and HNR.

Taken together, cross-linguistic research supports two conclusions that guide the present dissertation. First, phonation-related variation can align with prosodic structure (e.g., phrase edges and tonal categories), but the mapping is language- and context-dependent (Epstein, 2003; Garellek et al., 2013). Second, because both phonation and its acoustic correlates are multidimensional, the strongest empirical approach is to test whether specific prosodic categories and positions predict systematic shifts in acoustic correlates rather than assuming a one-to-one mapping between phonation type labels and individual measures.

2.5 Phonation, prosody, and cross-linguistic perspectives (Arabic focus)

This subsection reviews research on intonation in Arabic dialects, with an emphasis on Syrian Arabic, and identifies where evidence is currently strongest—and where it is still thin—regarding the prosody–phonation interface. Across dialects, Arabic is commonly analysed as having lexical stress with post-lexical intonation, meaning that pitch patterns are primarily used to express sentence modality, information structure, and prosodic phrasing rather than lexical contrasts (Hellmuth, 2006; El Zarka, 2017).

At the same time, the literature strongly supports the view that there is no single “Arabic intonation system.” Comparative overviews show both shared tendencies and substantial dialect-specific differences in the inventory and alignment of pitch accents and edge tones, as well as in phrasing structure and the density of tonal events (El Zarka, 2017; Hellmuth, 2014). This motivates treating Syrian Arabic not as a minor variant of a uniform Arabic pattern, but as a system requiring its own empirically grounded description.

Work on Lebanese Arabic has been influential in establishing Levantine intonation within the Autosegmental–Metrical (AM) framework. In a preliminary AM analysis, Chahal (1999) proposes a tonal inventory with pitch accent categories such as H*, L*, and downstepped high (!H*) as well as bitonal accents (e.g., L+H*), alongside edge tones (e.g., H%/L%). These categories support analyses of canonical falling declaratives and rising yes/no questions, while also highlighting how continuation patterns and phrase boundaries are signalled through tonal events at prosodic edges. Later comparative work on Lebanese

and Egyptian Arabic further underscores that dialects may share a stress–intonation interface while differing in phrasing and tonal distribution (Chahal & Hellmuth, 2014).

For Syrian Arabic, early instrumental work drew attention to the perceptual impression of “singing” intonation in Damascene Arabic. Kulk, Odé, and Woidich (2003) compared recordings from two Damascene and two Cairene male speakers, using both spontaneous speech and read questions. They report that in both dialects there are two basic statement contours and three basic question contours, but they argue that the perceptual difference between Damascene and Cairene Arabic is due above all to two prosodic properties: a larger pitch range in Damascene Arabic and pre-pausal vowel lengthening, which is absent in the Cairene data. In Damascene statements consisting of more than one utterance, they note that the vowel before the pause may be substantially lengthened and may also carry a rising movement; in one example, the final /a/ in *sakanna* ‘we lived’ is lengthened to 300 ms before a pause of 490 ms. They therefore identify pre-pausal lengthening and enhanced pitch range as central cues to the characteristic prosodic impression of Damascene Arabic. Although the study is explicitly a pilot study with a small number of speakers, it is still important because it identifies timing- and range-based properties that later work also treats as central to Syrian Arabic intonation. A fuller AM-based account of Damascene Syrian Arabic is discussed in Section 2.1.7.1.

Taken together, the Arabic-focused literature supports two conclusions that motivate the present dissertation. First, Syrian Arabic shares broad properties associated with stress-based intonation systems (accented prominence plus edge marking), while also showing dialect-specific characteristics such as strong phrase-final timing effects in certain utterance types (Kulk et al., 2003; Al Hasan & Mahanta, 2022, 2023). Second, despite progress on intonational structure, systematic evidence on how phonation covaries with Syrian Arabic prosody remains limited. This gap provides a concrete motivation for analysing phonation-related measures (e.g., jitter, shimmer, HNR) with respect to accented positions vs. boundary zones and to modality-driven configurations, rather than treating voice quality as an incidental by-product of speaking style.

2.6 The interaction of intonation and phonation in Arabic

Research on intonation has traditionally treated fundamental frequency (F0) as the primary acoustic correlate of prosodic structure. More recent work, however, shows that prosodic contrasts are often accompanied by systematic differences in voice quality—especially in

phrase-final domains and in association with low tonal targets—suggesting that phonation can function as an additional cue to prosodic structure (Gordon & Ladefoged, 2001; Redi & Shattuck-Hufnagel, 2001; Garellek, 2014). Importantly, the present dissertation examines voice quality instrumentally using jitter, shimmer, and HNR. These measures should be interpreted as acoustic correlates of phonation-related properties (periodicity stability and noise), not as direct categorical labels such as “creaky” or “breathy,” particularly in connected speech.

Within the Autosegmental–Metrical (AM) tradition, Arabic dialects have been analysed using dialect-specific tonal inventories and ToBI-style conventions, and the literature has established robust generalizations about accent distribution, alignment, and edge marking in several varieties (Hellmuth, 2006; Chahal & Hellmuth, 2014; El Zarka, 2017). What remains much less developed is the integration of phonation into these AM accounts: explicit, instrumental phonation analyses are still relatively rare for Arabic dialects, and claims about voice quality often remain impressionistic or peripheral to the main intonational analysis.

For Egyptian Arabic, Hellmuth (2006) provides a detailed AM analysis of pitch-accent distribution and phrasing, including evidence for strong edge marking and systematic alignment patterns. These results are important for the present study because they define the prosodic environments—accented syllables and phrase edges—where phonation-related variation would be most theoretically expected. For Lebanese Arabic, AM analyses propose relatively rich pitch-accent inventories and systematic use of edge tones to signal continuation versus finality, again providing clearly defined prosodic positions that could plausibly condition voice-quality correlates (Chahal, 1999; Chahal & Hellmuth, 2014). For Jordanian Arabic, work on word-level prosody shows strong boundary-related timing effects and F0 correlates of stress and juncture, underscoring the general relevance of prosodic structure for interpreting acoustic measures (de Jong & Zawaydeh, 1999). Across dialects, comparative syntheses emphasize that Arabic intonation is not uniform: dialects differ not only in inventories but also in alignment detail and in the phonetic implementation of prosodic structure (El Zarka, 2017; Hellmuth, 2014).

For Syrian Arabic, the available literature highlights both distinctive phrase-edge timing and emerging AM descriptions of tonal structure. A pilot study of Damascene Arabic reports a perceptually salient “singing” quality and strong pre-pausal vowel lengthening, especially in utterance-final contexts, where local melodic movement is also prominent

(Kulk, Odé, & Woidich, 2003). More recent work by Al Hasan and Mahanta (2023) develops this line of research by examining post-tonic lengthening and F0 alignment in Syrian Arabic statements and questions. On the basis of 20 utterances produced by six native speakers, they investigate the alignment of F0 peaks and valleys together with the duration of phrase-final segments, and argue that F0 cues prominence within the accentual phrase, while F0 and duration together cue larger prosodic domains. In particular, they report that phrase-final post-tonic short vowels are lengthened in Syrian Arabic questions, and that this lengthening contributes to the final rises observed in both *wh*-questions and yes/no questions. Although this is not a study of phonation as such, it is highly relevant here because it identifies boundary-related prosodic environments, especially phrase-final zones involving enhanced duration and F0 movement, in which phonation-related measures might also show systematic patterning.

On the basis of cross-linguistic findings and the Arabic intonation literature, three prosodic loci are especially relevant for testing intonation–phonation covariation in Syrian Arabic. First, phrase-final boundary zones are a strong candidate because phrase-final domains frequently involve low tonal targets, temporal expansion, and—in many languages—an increased incidence of non-modal phonation (Redi & Shattuck-Hufnagel, 2001; Garellek, 2014). Second, question-related final rises are relevant because Syrian Arabic is reported to show robust modality-driven phrase-final rises with accompanying temporal adjustments, which may alter the stability/noise properties captured by jitter, shimmer, and HNR (Al Hasan & Mahanta, 2022, 2023; Kulk et al., 2003). Third, prominence sites (especially nuclear accents) are theoretically relevant because prosodic strengthening can involve laryngeal and aerodynamic adjustments that may affect perturbation and noise measures, even when the segmental material is held constant.

Despite detailed AM descriptions of intonation for several Arabic dialects, the phonation dimension remains under-documented—particularly for Syrian Arabic. To date, there is no systematic study that models how jitter, shimmer, and HNR pattern with pitch-accent categories, boundary-related structure, prosodic position, and modality in Syrian Arabic connected speech. The present dissertation addresses this gap by jointly analysing intonational structure and phonation-related measures, testing whether voice quality provides redundant/enhancing cues to prosodic structure and whether such cues differ across sentence type and phrase position.

2.7 English as a comparative reference for the intonation-phonation interface

English is relevant here not as a second experimental language, but as a reference literature for a stress-accent language in which both intonational categories and some prosody-phonation covariation are comparatively well described. Earlier British-school descriptions treated English contours largely as holistic nuclear tones such as fall, rise, fall-rise and rise-fall, linking them to broad discourse meanings (e.g. Crystal 1969; O'Connor & Arnold 1973; Hirst 1998). A familiar textbook contrast is the Southern British pair *It's YOUR fault!* versus *MY fault?*, where a falling statement is opposed to a rising echo question. For cross-linguistic comparison, however, the decisive development is the autosegmental-metrical analysis of English inaugurated by Pierrehumbert (1980) and elaborated by Pierrehumbert and Hirschberg (1990), in which contours are decomposed into pitch accents associated with metrically strong syllables and edge tones associated with prosodic boundaries. In this framework, a canonical declarative may be represented as H* L-L%, while a typical yes/no question may be represented as L* H-H%. English therefore offers an explicit model in which prominence-related and boundary-related tonal events can be analysed separately rather than being folded into one holistic tune description.

English is also useful because it shows how much variation can exist even in a well-studied intonation system. Work from the IViE project¹² demonstrates substantial cross-dialect and within-dialect variation in both tune choice and phonetic implementation across British Isles varieties. For example, Grabe and Post (2002) show that in the same short declarative materials, speakers from Cambridge overwhelmingly preferred falling nuclei, whereas speakers from Belfast overwhelmingly preferred rise-plateaux. In the fuller IViE discussion, this contrast is illustrated with the same assertion, *It's Melanie!:* in Southern British English the utterance is produced with a falling contour, while in Belfast English the same utterance is produced with the characteristic rise-plateau contour associated with Belfast declaratives (Grabe, Kochanski, & Coleman, 2005). The same point extends beyond statements. In the IViE inversion-question materials, including utterances such as *May I lean on the railings?*, Cambridge speakers no longer strongly favored declarative-like falls, whereas Belfast speakers continued to prefer rise-plateaux, showing that the

¹² IViE stands for Intonational Variation in English. It was a large corpus-based research project on cross-varietal and stylistic variation in the intonation of nine urban dialects of English spoken in the British Isles, and it also developed a comparative prosodic labelling system designed for this purpose (Grabe & Post, 2002).

mapping between sentence type and tune is dialect-specific rather than one-to-one (Grabe & Post, 2002). These results therefore support two important conclusions: first, that the same English sentence can receive different intonational realizations in different dialects; and second, that sentence type does not determine contour choice in a categorical way.

Other dialects depart even more strongly from the Southern British pattern. In Belfast English, statement rises are not marginal but form part of the ordinary declarative system, and later work argues that these rises are better understood as a local dialect pattern than as simple “uptalk” borrowed from elsewhere (Sullivan, 2012). In Glasgow English, researchers developed a dedicated GlaToBI system precisely because the dialect’s characteristic rise-plus-plateau contours were not well captured by standard English ToBI assumptions (Mayo, Aylett, & Ladd, 1997). More generally, IViE-based work shows that even when dialects share the same broad phonological categories, they may differ in how those categories are realized phonetically: for example, Cambridge and Newcastle speakers tend to show compression of rises and falls, Leeds speakers show truncation in comparable contexts, and Belfast speakers show rise-plateaux where Cambridge, Leeds, and Newcastle often show simple rises or falls.

More recent work on American English likewise suggests that not all tune distinctions predicted by AM/ToBI are equally robust in production and perception, especially when discourse context is removed. In an imitation study¹³ of 12 predicted American English nuclear tunes, combining three pitch accents (H*, L+H*, L*+H) with four edge-tone sequences (H-H%, H-L%, L-H%, L-L%), Steffman, Shattuck-Hufnagel, and Cole (2022) found that the full theoretical inventory did not emerge cleanly in production: their analysis found fewer clearly distinct tune types than the full AM/ToBI system predicts, with the clearest difference being between generally rising contours and rise–fall contours. Their results therefore suggest that some distinctions posited in the AM/ToBI system are phonetically subtle and may not always be equally robust without supporting discourse context. This is important because it shows that even in English, the relation between phonological categories and surface realization is not always straightforward.

English also provides independent evidence that phonation-related phenomena are sensitive to prosodic structure. In American English, Redi and Shattuck-Hufnagel (2001)

¹³ An imitation study is a study in which participants hear spoken utterances and then repeat them as closely as possible. In intonation research, this helps show which prosodic patterns are clearly perceived and reliably reproduced.

show that glottalization is more frequent at utterance ends than at utterance-medial phrase boundaries, and more frequent at full intonational phrase boundaries than at intermediate phrase boundaries. A different but related pattern is word-initial glottalization, where vowel-initial words such as *apple* may be realized with an initial glottal stop, e.g. [ʔæpl]. Garellek (2014) argues that this pattern is best understood as a form of prosodic strengthening: prominent word-initial vowels show greater glottal constriction, whereas non-prominent phrase-initial vowels do not. Taken together, these studies support a view in which non-modal phonation in English is not random, but patterned by prosodic position, with utterance-final domains favoring phrase-final creak and prominent or strongly initial positions favoring glottalization.

At the same time, the literature makes clear that English “creaky voice” is not a single phenomenon. Garellek (2015) compares phrase-final creak with /t/-glottalization and shows that the two may be acoustically similar yet perceptually distinct. In his perception experiment, listeners were asked to identify near-minimal pairs such as *button* and *bun*, where only the first can undergo /t/-glottalization. Even when /t/-glottalization co-occurred with phrase-final creak, listeners generally still identified the words correctly, which suggests that /t/-glottalization and phrase-final creak remain perceptually distinguishable to English listeners.

Broader review work also cautions against reducing English creaky voice to popular stereotypes about a single demographic group. In their systematic review, Dallaston and Docherty (2020) identified only ten studies that met their inclusion criteria. Across those ten studies, the total speaker sample was only 297, of whom 64% were women; moreover, nine of the ten studies examined present-day American English, and many of these focused on young female college students. Dallaston and Docherty argue that the literature has concentrated heavily on one social group, not because that group has been shown to use more creaky voice than all others, but because it has been disproportionately studied.

Their review also shows that the underlying studies differed sharply in speech material. For example, Henton and Bladon (1988) examined eleven read sentences from speakers of Received Pronunciation and Modified Northern British English, whereas Yuasa (2010) recorded approximately ten-minute conversations about food but analysed only a 401-word segment from each conversation. Other studies used different tasks: the two studies by Abdelli-Beruh and colleagues (2012; 2014) analysed selected sentences from *The Rainbow Passage*, while Borrie and Delfino (2017) examined speech from the first five minutes of a

cooperative picture-matching task. This means that the reviewed literature was not the same: some studies measured creak in highly controlled read speech, while others examined short extracts from spontaneous interaction.

The methods used to identify and quantify creaky voice were also inconsistent. Dallaston and Docherty report three different detection methods across the ten studies: auditory coding, auditory-visual coding, and automated acoustic detection. In the auditory and auditory-visual studies, researchers variously defined creak by a low pitch, a rough quality, resolvable glottal pulses, or even a “popping corn” sound. Only one study used an automated acoustic method, and that method classified a syllable as creaky if most of its F0 measurements fell below 150 Hz. Dallaston and Docherty argue that this methodological diversity makes direct comparison extremely difficult: even two studies that appear to use the “same” measure may not in fact be measuring the same thing.

For that reason, their conclusion is deliberately cautious. They do not argue that young American women never use more creaky voice, but rather that the existing evidence is too limited and too inconsistent to support broad claims about a recent increase in English creaky voice or about its concentration in one demographic group. What the review does support is a much narrower point: the prevalence of creaky voice in English appears to vary across groups, individual speakers, and speaking contexts, but the exact pattern of that variation remains unclear because the evidence base is sparse and methodologically uneven.

2.8 Summary and research gap

The literature reviewed in this chapter shows that intonation is now routinely modelled as a structured system within the Autosegmental–Metrical (AM) framework, with English playing a foundational role in the development of AM representations and ToBI-style annotation practices (Pierrehumbert, 1980; Ladd, 2008). Within Arabic, AM analyses have expanded the empirical base beyond English by proposing dialect-specific inventories and phrasing accounts for varieties such as Egyptian and Lebanese Arabic, and broader syntheses emphasize substantial dialectal diversity rather than a single uniform “Arabic intonation system” (Hellmuth, 2006; Chahal & Hellmuth, 2014; El Zarka, 2017). For Syrian Arabic, the descriptive record remains smaller but is growing: early instrumental work highlights strong phrase-final timing effects in Damascene speech (Kulk et al., 2003),

and recent work proposes an initial AM analysis of tonal categories and alignment patterns (Al Hasan & Mahanta, 2022).

Parallel research on phonation demonstrates that voice quality is often prosodically patterned: non-modal phonation and related acoustic correlates can cluster at prosodic boundaries and may covary with tonal categories (Gordon & Ladefoged, 2001; Redi & Shattuck-Hufnagel, 2001; Garellek, 2014). However, for Arabic dialects—and especially for Syrian Arabic—this interface is still under-documented in instrumental terms. Existing Arabic intonation studies provide strong motivation about where prosodic structure is (accented syllables vs. boundary zones), but they rarely test whether phonation-related measures change systematically at those locations.

This review, therefore, motivates three concrete research gaps. First, despite emerging AM descriptions of Syrian Arabic intonation, there is not yet a systematic account of whether phonation-related acoustic measures covary with Syrian Arabic prosodic structure in connected speech—specifically with pitch-accented positions, phrase-final boundary zones, and modality-driven configurations. Second, cross-linguistic results suggest that phonation can function as an additional cue that enhances prosodic contrasts, but this hypothesis has not been tested quantitatively for Syrian Arabic using replicable acoustic measures (Redi & Shattuck-Hufnagel, 2001; Garellek, 2014).

A further gap concerns the role of speaker-related variation within these prosodic patterns, especially variation associated with gender. Studies of healthy voices have reported mixed gender effects for jitter and shimmer, and somewhat clearer but still task-dependent differences for HNR, suggesting that apparent male–female contrasts may partly reflect differences in F0, intensity, and speaking condition rather than stable phonatory properties alone (Orlikoff & Kahane, 1991; Brockmann et al., 2008, 2011; Dehqan et al., 2010; Shabnam & Pushpavathi, 2022). This issue has not been systematically examined for Syrian Arabic connected speech. Accordingly, the present dissertation not only asks whether voice-quality measures covary with prosodic structure and sentence modality, but also whether these patterns remain stable across male and female speakers once local acoustic covariates are taken into account.

This dissertation addresses these gaps by analysing how jitter, shimmer, and HNR—treated as acoustic correlates of periodicity stability and noise rather than direct categorical labels—vary as a function of prosodic position, intonational structure, and sentence

modality in Syrian Arabic, with English serving as a comparative baseline where prosody–phonation covariation is independently documented.

It also examines whether the modality- and position-related patterns observed in the main corpus remain stable across male and female speakers, and whether any apparent gender effects are modified once local acoustic covariates such as F0 and intensity are taken into account. In this way, the dissertation moves beyond a purely structural account of intonation–phonation relations and tests how prosodic organization, acoustic realization, and speaker-related variation interact in connected Syrian Arabic speech.

Chapter Three: Methodology

3.1 Introduction

This dissertation investigates the intonation–phonation interface in Damascene Syrian Arabic by combining (i) an Autosegmental-Metrical (AM) analysis of intonational structure, implemented in an AM/ToBI-style labelling system, with (ii) acoustic measures commonly used as correlates of phonatory behaviour. In this study, phonation is not treated as a categorical label (e.g., “breathy,” “creaky”) inferred directly from acoustics; rather, it is operationalized through jitter, shimmer, and harmonics-to-noise ratio (HNR), interpreted as acoustic correlates that may covary with prosodic structure under controlled conditions (Baken & Orlikoff, 2000; de Krom, 1993; Ferrand, 2002).

The methodology adopted in this dissertation is explicitly instrumental and annotation-based. This follows directly from the aims of the study, which are concerned not only with describing intonational categories, but also with examining the distribution of fine-grained acoustic properties in voiced intervals and their systematic relation to prosodic structure. Because perturbation- and noise-related measures such as jitter, shimmer, and HNR are sensitive to segment selection, recording quality, and local phonetic context, the analysis requires clearly defined measurement protocols and controlled elicitation conditions (Brockmann et al., 2011; de Krom, 1993). Similarly, the analysis of intonational events requires explicit prosodic annotation and temporally grounded acoustic inspection in order to relate phonetic measurements to accented syllables, boundary zones, and other prosodically relevant positions (Beckman & Ayers Elam, 1997; Ladd, 2008; Jun, 2014).

This chapter presents the methodological framework of the dissertation. It first describes the overall research design, then details the participants, materials, and recording procedure. It subsequently specifies the annotation workflow and the acoustic extraction procedures (including the operational definitions used for jitter, shimmer, and HNR), and finally outlines the statistical modelling strategy used to test the hypotheses of the study.

3.2 Research design

The dissertation employs a two-stage design: a pilot study and a main study. The pilot study is used to test and refine materials, elicitation procedures, and annotation/extraction settings, with the goal of minimizing avoidable sources of variance (e.g., unintended item

effects or unstable measurement settings) before the main data collection. The main study provides the primary dataset used for hypothesis testing.

Across both stages, the design is controlled and factorial in the sense that it targets systematic variation in prosodic structure, focusing on (i) sentence modality (declaratives, yes/no questions, *wh*-questions) and (ii) positional variation of the target word (initial, medial, final). The dependent outcomes are acoustic measures extracted from defined voiced intervals—jitter, shimmer, and HNR—treated as correlates potentially associated with prosodic phrasing, pitch accents, and boundary tones as represented in an AM/ToBI-style analysis (Beckman & Ayers Elam, 1997; Ladd, 2008). English is included as a comparative reference to contextualize patterns observed in Syrian Arabic, without assuming that English intonational structure is typologically default.

3.2.1 Pilot study

The pilot study was conducted with a single native speaker of Syrian Arabic (Damascene). Its purpose was formative: to test the feasibility of the experimental design and the end-to-end workflow (elicitation → recording → prosodic annotation → acoustic extraction) prior to expansion to multiple participants in the main study.

Five target words were selected for this stage. Their distribution across sentence modalities and positions was as follows. In declaratives, each target word occurred in three positions (initial, medial, final). In *wh*-questions, target words were elicited in two positions (medial and final) because the *wh*-phrase must occupy clause-initial position in the intended structures, preventing an additional lexical item from appearing initially without altering the construction. In yes/no questions, target words were also elicited in two positions (medial and final); placing a non-*wh* lexical item in initial position would typically require topicalization (a structure reserved for the main study), which would introduce an additional syntactic–information-structural manipulation not intended for the pilot.

Unlike the shorter and more tightly controlled utterances used in the main study, the pilot materials included longer sentences with richer contextual content, and some items were produced as two intonational phrases (i.e., two IP-sized units in an AM framework). Longer materials can increase perceived naturalness but also increase opportunities for variation in prosodic phrasing and tonal realization, which complicates consistent ToBI-style annotation and temporally anchored acoustic measurement (Jun, 2014; Ladd, 2008).

This pilot served two goals. First, it assessed whether target words could be embedded in the intended sentence positions without compromising fluency or acceptability. Second, it evaluated whether the intended measurement workflow (intonational annotation plus extraction of acoustic correlates of phonatory behavior) remained stable in more complex utterances. Pilot observations suggested that although longer sentences often sounded more natural, they yielded greater prosodic variability (e.g., in phrasing choices), which reduced measurement comparability across tokens. For this reason, the main study employed shorter, more controlled sentences while preserving the core manipulations of sentence modality and target-word position. For a detailed description of the pilot study design, procedures, and sample materials, see Chapter 4 (Pilot Study).

3.2.2 Main study

The main study employed a controlled read-speech task. This elicitation mode was selected to support (i) high and consistent recording quality and (ii) reproducible segmental and prosodic environments for detailed acoustic analysis of intonation and voice-related measures. Although spontaneous speech can provide a wider stylistic range, it also introduces substantial and difficult-to-control variation in speech planning, phrasing, and discourse structure. A controlled design therefore better matches the present goal of testing how sentence modality and target-word position pattern with ToBI-style intonational structure and acoustic correlates of phonatory behavior (Xu, 2010).

The study used a within-subjects (repeated-measures) design, such that each speaker produced all conditions. This design choice reduces the impact of stable individual differences (e.g., baseline voice quality and habitual pitch range) on comparisons across conditions, because each speaker serves as their own control (SAGE Research Methods, n.d.).

Three sentence modalities were elicited: declaratives, yes/no questions, and *wh*-questions. Within each modality, target words were placed in systematically controlled positions as far as the morphosyntax of Damascene Syrian Arabic allowed in order to evaluate positional effects. As outlined in the morphosyntactic overview in Chapter 1, each target word appeared in initial, medial, and final position in declaratives. In yes/no questions, target words were likewise elicited in three positions, but the initial-position condition required fronting of the lexical target. This follows from the clause structure of Damascene Syrian Arabic: yes/no questions are typically morphosyntactically parallel to declaratives

and are distinguished primarily by intonation, while verb-initial order is an attested basic order in the dialect. In *wh*-questions, target words occurred only in medial and final position, since the clause-initial slot is reserved for the *wh*-phrase in the relevant structures (Cowell, 1964; Sulaiman, 2016; Al Hasan & Mahanta, 2022). The distribution of target-word positions in the materials therefore reflects not only experimental control, but also the clause-structural properties of Damascene Syrian Arabic that determine which positions are available without changing sentence type or discourse structure.

This design was intended to support evaluation of main effects of sentence modality and target-word position, as well as their interaction, while keeping the positional contrasts as structurally comparable as possible within the morphosyntactic constraints of Damascene Syrian Arabic.

3.2.3 Stimulus structure

The main study included ten native speakers of Syrian Arabic. Each speaker produced three target words, with three repetitions of each stimulus in each applicable condition. Three repetitions were elicited in order to reduce the effect of ordinary production variability in controlled sentence reading and to ensure that at least two usable tokens would be available for analysis in cases of hesitation, disfluency, segmental error, or technical irregularity. For the final dataset, two repetitions per item were included. This selection was made after manual inspection of all repetitions in the waveform, spectrogram, and F0 trace. Tokens showing performance errors, hesitations, clipping, or contours that were insufficiently interpretable for prosodic analysis were excluded, and the remaining two repetitions judged to be the most accurate and acoustically reliable were retained. The stimulus set was structured around sentence modality and target-word position, summarized in Table 4.

Table 4 sentence design and distribution

Sentence Type	Positions Tested	Target Words (per speaker)	Repetitions (per condition)
Declaratives	Initial, Medial, Final	3	3
Yes/No Questions	Initial (Topicalization), Medial, Final	3	3
Wh-Questions	Medial, Final	3	3

This yields 72 utterances per speaker, calculated as follows:

- Declaratives: $3 \text{ positions} \times 3 \text{ words} \times 3 \text{ repetitions} = 27$
- Yes/No questions: $3 \text{ positions} \times 3 \text{ words} \times 3 \text{ repetitions} = 27$
- *Wh*-questions: $2 \text{ positions} \times 3 \text{ words} \times 3 \text{ repetitions} = 18$
- Total per speaker = $27 + 27 + 18 = 72$.

With 10 speakers, the total dataset comprises 720 recorded utterances (10×72). However, because only two repetitions per item were retained for the final analysis after manual inspection, the analyzed dataset was smaller:

- Declaratives: $3 \text{ positions} \times 3 \text{ words} \times 2 \text{ repetitions} = 18$
- Yes/no questions: $3 \text{ positions} \times 3 \text{ words} \times 2 \text{ repetitions} = 18$
- *Wh*-questions: $2 \text{ positions} \times 3 \text{ words} \times 2 \text{ repetitions} = 12$
- Total analyzed per speaker = $18 + 18 + 12 = 48$.

With 10 speakers, the final analyzed dataset therefore consisted of 480 utterances (10×48). This balanced structure supports estimation of the effects of modality and position within a repeated-measures framework.

3.2.4 Rationale

The stimulus design reflects both theoretical and practical considerations. Theoretically, the dissertation asks whether variation in intonational structure (as represented in an AM/ToBI-style analysis of pitch accents, boundary tones, and phrasing) covaries with acoustic correlates associated with phonatory behavior (jitter, shimmer, HNR). Practically, this requires materials that allow (i) consistent prosodic labeling and (ii) stable acoustic extraction across comparable voiced intervals, which is more feasible under controlled elicitation than under unconstrained spontaneous speech.

Manipulating target-word position is intended to test whether positional environments that typically differ in prosodic organization (e.g., phrase edges vs. phrase-medial contexts) show systematic differences in the acoustic correlates under study. Including multiple sentence modalities (declaratives, yes/no questions, *wh*-questions) enables evaluation of whether modality-linked intonational patterns are associated with measurable differences in these correlates. Finally, a within-subjects design enhances comparability across conditions by reducing the impact of stable between-speaker differences in baseline voice measures.

3.3 Participants

The participant group of the main study consisted of ten native speakers of Syrian Arabic, balanced by gender (five males, five females) and aged 20–30 years. Restricting the age range was intended to reduce variability associated with age-related changes in voice measures; nevertheless, sex-, intensity-, and F0-related effects remain relevant for perturbation measures and must be controlled or modeled statistically (Brockmann et al., 2011).

All participants reported that their everyday spoken variety was the urban Damascus (Damascene) variety. Although some speakers originated from other regions of Syria, they reported long-term residence in Damascus and regular daily use of the Damascene variety. Because recruitment occurred in Hungary, participants were drawn from the Syrian student community in Budapest (including ELTE and other Hungarian universities). This sampling strategy reflects practical constraints of recruiting a diasporic population and should be treated as a potential source of inter-speaker variation to be addressed analytically.

Inclusion criteria required self-reported normal speech and hearing and no history of speech or language impairment. All participants were literate in Arabic. Participants were recruited on a voluntary basis. Informed consent was obtained prior to recording. Participants were informed of the study procedures and data handling, and they were explicitly informed of their right to refuse participation or withdraw at any time without penalty.

Participation was based on verbal informed consent obtained prior to recording. Participants were informed of the purpose of the study, the recording procedure, and the intended research use of the recordings. They were also informed that the data would be handled in anonymized form: speakers were identified only through coded labels in file names and analysis tables, and no personally identifying information was included in the dissertation. Recordings were used only for the present research and were not linked to publicly reported personal data. Participant anonymity was therefore protected through de-identification in storage, analysis, and reporting.

3.4 Materials

The materials were designed to elicit three sentence modalities—declaratives, yes/no questions, and *wh*-questions—while manipulating the sentence position of a target word.

For declaratives and yes/no questions, target words were tested in initial, medial, and final positions. For *wh*-questions, target words were tested in medial and final position only, because the clause-initial slot is occupied by the *wh*-phrase in the targeted *wh*-constructions.

For the yes/no questions, the initial-position condition required fronting/topicalization in those items whose baseline clause structure in this study was verb-initial; this ensures that the target word appears clause-initially without altering the intended interrogative construction. Importantly, this implementation detail creates a structural difference between the yes/no “initial” condition and the medial/final conditions in verb-initial clauses, which is treated as a design constraint rather than evidence that “position” is purely prosodic.

Three target words were selected: *ka:se* ‘glass’, *walad* ‘boy’, and *maktab* ‘office’. All three carry stress on the first syllable. Selection was motivated by lexical familiarity, comparable stress structure, and the availability of a segmentable stressed vowel from which a usable voiced interval could be extracted. The latter consideration was especially important for two reasons. First, perturbations measures such as jitter, shimmer, and related noise measures are highly sensitive to segment selection and are most interpretable when calculated from relatively stable, nearly periodic vowel portions rather than from onset–offset transitions or consonant-adjacent regions (Choi et al., 2012). Second, vowel-onset F0 is systematically influenced by preceding obstruents, and voiceless obstruents in particular tend to raise F0 at vowel onset (Hanson, 2009). Therefore, these words were not assumed to be acoustically ideal in every respect; rather, they provided a practical compromise between lexical naturalness and the possibility of identifying a sufficiently stable mid-vowel measurement window.

All stimuli were constructed in colloquial (Damascene) Syrian Arabic, with corresponding IPA and English translations. Compared with the pilot materials, the main-study sentences were intentionally shorter and more tightly controlled in their lexical, syntactic, and prosodic design. This redesign followed directly from the pilot study, which showed that longer, more contextually elaborated sentences introduced greater variability in phrasing and tonal realization, and in some cases yielded more than one intonational phrase. In the main study, sentence frames were therefore simplified, extraneous contextual material was reduced, and target-word placement was manipulated systematically across sentence modality and prosodic position. The aim was to keep the utterances natural while

increasing comparability across tokens in the prosodic and phonetic environments used for annotation and acoustic measurement.

3.4.1 Sample stimuli

Table 5 Sample declarative sentences with all three positions

Arabic	IPA	Gloss	Translation
كاسة المي نظيفة	'ka:se-t-l maj nd'i:fe	glass-of-the water clean	The glass of water is clean.
شفت ولد عم بيكي	ʃift 'walad ʕam jebki	see.PST.1SG boy PROG cry	I saw a boy crying.
خلصت شغل بالمكتب	xallʕaʃt ʃəɣəl bil 'maktab	finish.PST.1SG work in.the-office	I finished work at the office.

Table 6 Sample Yes/No Questions with all three positions

Arabic	IPA	Gloss	Translation
كاسة المي نظيفة؟	'ka:se-t-l maj nd'i:fe?	glass-of-the water clean?	Is the glass of water clean?
شفت ولد عم بيكي؟	ʃift 'walad ʕam jebki?	see.PST.2SG boy PROG cry?	Did you see the boy crying?
خلصت شغل بالمكتب؟	xallʕaʃt ʃəɣəl bil maktab?	finish.PST.2SG work in.the-office?	Did you finish work at the office?

Table 7 Sample Wh-Questions with medial and final positions

Arabic	IPA	Gloss	Translation
وين كاسة المي؟	we:n 'ka:se-t-l maj?	where glass-of-the water?	Where is the glass of water?
إيمتى خلصت شغل بالمكتب؟	'ʔe:mta xallaʃt ʃəɣəl bil maktab?	When finish.PST.2SG work in the-office?	When did you finish work at the office?

3.5 Recording procedure

Recordings were conducted in the sound booth of the Hungarian Research Centre for Linguistics (Budapest) to minimize background noise and reverberation, supporting reliable acoustic analysis. However, according to Awan et al. (2024; 2025), it is important to recognize that acoustic voice measures can vary across recording chains and

instrumentation. This motivates the standardization of equipment and procedure, which is described below.

Participants were seated comfortably. A head-mounted condenser microphone was used to maintain a consistent mouth-to-microphone geometry across speakers. Consistent microphone placement is a standard recommendation in instrumental voice assessment because variation in distance and angle can introduce randomness in acoustic measures (Patel et al., 2018). Audio was recorded at 44.1 kHz sampling rate and 16-bit resolution. This recording format is widely used in phonetic research and is compatible with standard acoustic-analysis workflows; however,

Stimuli were presented using SpeechRecorder, which supports scripted presentation and structured storage of recordings for speech data collection (Draxler & Jänsch, 2004; BAS, 2025). Sentences were displayed on a computer screen, and each sentence was read three times. The software was configured to save each token as a separate audio file to facilitate subsequent annotation and acoustic extraction.

The recording session began with a short warm-up phase of five practice sentences, followed by an instruction screen asking participants to read the materials in their natural voice. The reason for this was to try to avoid exaggerated pitch movements or deliberate articulation. The main experimental block was then presented in randomized order. Each trial included a 1 second pre-recording delay and a fixed 4 seconds recording window (4.5 s in the warm-up items), followed by a minimal post-recording delay. In addition to the core target stimuli, the script also contained a small set of filler/distractor items interspersed in the randomized sequence. The script itself did not include programmed break screens; accordingly, any pauses taken during the session to reduce fatigue were handled manually rather than built into the stimulus script.

Recordings were monitored in real time to ensure a consistent signal level and to avoid clipping. In this sense, standardization was achieved not only through the recording format itself, but also through constancy of environment, microphone setup, stimulus delivery, and monitoring procedure to produce recordings suitable for ToBI-style annotation and extraction of jitter, shimmer, and HNR as acoustic correlates.

3.6 Data annotation

All recordings were annotated in Praat using multi-tier TextGrids aligned to the waveform and spectrogram (Boersma & Weenink, 2024). A multi-tier annotation structure was adopted because the dissertation relates intonational structure (pitch accents, boundary tones) to acoustic correlates associated with phonatory behavior (jitter, shimmer, HNR).

Each TextGrid contained five tiers, designed to keep prosodic and segmental information aligned while maintaining analytic separation:

1. **Tone tier:** Intonational events were labeled using an AM/ToBI-style inventory (Beckman & Ayers Elam, 1997; Ladd, 2008). Pitch accents (e.g., H*, L*, L+H* as used in this dissertation) were assigned with reference to lexical stress and perceived prosodic prominence, and phrase-final tonal events were labeled using boundary tones (e.g., L%, H%) to mark intonational-phrase (IP) edges. This approach follows standard AM assumptions about the association of pitch accents with metrically prominent syllables and boundary tones with prosodic phrase edges, while allowing language-specific inventories and implementation choices (Jun, 2014; Ladd, 2008). The categories and criteria used here are informed by earlier AM analyses of Arabic varieties and recent work specifically on Damascene Syrian Arabic (Al Hasan & Mahanta, 2022).
2. **Voice Quality (VQ) tier:** This tier marked the intervals used for subsequent extraction of jitter, shimmer, and HNR. The annotated intervals correspond to the target vowels, because vowel nuclei typically provide the most stable voiced portions for perturbation and noise-related measures when compared with consonantal regions that may contain aperiodicity or reduced voicing. Importantly, at this stage, the tier contains interval definitions only; acoustic measurements were not performed during annotation. The measurement protocol and extraction settings are specified in the subsequent section on acoustic analysis.
3. **Syllable tier:** Utterances were segmented into syllabic units to support alignment between tonal events (e.g., pitch accents) and metrical structure. The syllable tier serves as an explicit bridge between (i) tonal annotation and (ii) the segmental windows from which acoustic correlates are later extracted.

4. **Word tier:** Lexical items were transcribed (phonemically) and target words were explicitly identified in their designed positions (initial/medial/final as defined in the materials). This tier supports straightforward indexing of tokens and provides a cross-check for the syllable and tone tiers.
5. **Gloss tier:** An English gloss was included to document stimulus meaning and intended modality (declarative vs. interrogative). Although not used directly in acoustic extraction, this tier supports transparency and reproducibility by enabling later verification of item type and function.

Annotation was performed manually by the researcher, and each file was reviewed in multiple passes to improve internal consistency. Manual annotation was preferred because fully automated pipelines remain sensitive to local segmental and voicing-related perturbations in the speech signal (Mertens, 2004), and because the automatic detection of fine prosodic events becomes less reliable when applied to spontaneous or new datasets, especially for non-silent boundaries and intermediate phrase distinctions¹⁴ (Rosenberg, 2010).

This multi-tier annotation system provides the structural basis for the analyses that follow: intonational events and prosodic boundaries are represented directly in the Tone tier, while voice-quality correlates are operationalized via the vowel intervals marked on the VQ tier and calculated in the subsequent acoustic analysis section.

3.7 Acoustic analysis

The acoustic analysis addressed two domains: intonation and voice-quality correlates. Intonation was analyzed using an AM/ToBI-style framework, combining auditory judgment with inspection of the waveform, spectrogram, and F0 track in Praat, following standard practice in intonation transcription (Ladd, 2008). Voice-quality correlates were quantified using a custom Praat script that extracted jitter (local %), shimmer (local %), and harmonics-to-noise ratio (HNR, dB) from the vowel intervals marked in the VQ tier (Boersma & Weenink, 2024).

¹⁴ Automatic detection is generally more reliable for strong prosodic events than for subtler ones. In particular, phrase boundaries that are not marked by silence and distinctions involving intermediate phrasing are harder to identify automatically, because they depend on weaker combinations of pitch and temporal cues rather than on a single robust signal (Rosenberg, 2010).

Because perturbation and noise-related measures are highly sensitive to pitch-tracking settings, segment selection, and signal quality, the script incorporated speaker-adaptive pitch settings that used the two-pass method for estimation of the f0 range adapted to each speaker (Al-Tamimi, 2022). The specific procedures and parameter settings used for intonational inspection and for the extraction of jitter, shimmer, and HNR are described in the following subsections.

3.7.1 Intonation analysis

Intonational structure was analysed in Praat through repeated inspection of the F0 track in relation to the waveform, wide-band spectrogram, and the aligned annotation tiers, including the tiers for tone, syllable, word, and gloss. The aim of this stage was not simply to identify visible rises and falls in the raw contour, but to interpret each token within an Autosegmental–Metrical (AM) framework in which tonal events are related to stressed syllables and prosodic boundaries rather than to overall contour shape alone (Beckman & Ayers Elam, 1997; Ladd, 2008). For this reason, each utterance was evaluated both segmentally and prosodically: segmental alignment was used to establish the location of stressed syllables and phrase-final material, while the F0 contour was interpreted with reference to prominence, phrasing, and tonal movement across the utterance.

For purposes of visual inspection and labeling, the displayed F0 contour was pitch-smoothed in Praat using the Praat Vocal Toolkit (Corretge, 2012-2025) to reduce the effect of local microprosodic perturbations in the track. Smoothing was used only as an aid to visual interpretation of intonational events and not as a replacement for auditory judgment or for the underlying segmental annotation. In other words, labels were not assigned mechanically from the smoothed contour itself; rather, smoothing was used to make broader melodic structure easier to inspect when local consonantal effects or brief tracking irregularities obscured the displayed F0 movement. This follows the same general Praat-based AM workflow used in prior work on Damascene Syrian Arabic intonation (Al Hasan & Mahanta, 2022).

Pitch accents were identified as tonal events associated with metrically prominent syllables, while boundary tones were used to mark the edges of prosodic phrases, in line with standard AM practice (Beckman & Ayers Elam, 1997; Ladd, 2008). In practical terms, tonal labels were assigned on the basis of several converging criteria rather than on the basis of raw “peaks versus valleys” alone. These criteria included: (i) the alignment of

tonal turning points relative to the stressed syllable, (ii) the local scaling of F0, and (iii) the presence of phrase-final patterns consistent with prosodic boundary marking. Thus, a rise aligned closely with a stressed syllable was not treated in the same way as a rise that emerged only at the phrase edge, and a final tonal movement was interpreted with reference to its position in the prosodic structure rather than as an isolated contour fragment.

Particular attention was paid to the distinction between accentual events and edge-related events. In clause-medial positions, the analysis focused on whether a tonal movement was anchored to the stressed syllable of the target word and how far that movement extended into the following material. In phrase-final positions, the analysis focused primarily on final lowering or raising and on the alignment of these movements relative to the final stressed syllable and the right edge of the utterance. Although phrase-final timing effects are discussed in the literature on Damascene Syrian Arabic, they were not independently quantified in the present analysis; accordingly, phrase-final contours were interpreted here on the basis of tonal movement and prosodic position rather than on direct durational measurement.

Because the materials were produced in connected speech, the analysis also took account of contexts in which local segmental perturbations could obscure the underlying intonational pattern. Voiceless obstruents, weakly voiced segments, and rapid transitions between syllables can produce irregularities in the displayed F0 track that do not themselves constitute phonologically relevant tonal events. For that reason, each token was inspected auditorily as well as visually, and tonal labels were assigned only after considering the contour in relation to the segmental string and the broader utterance pattern. In ambiguous cases, the surrounding tokens produced by the same speaker in comparable conditions were also consulted in order to maintain internal consistency of analysis.

Annotation was performed manually by the researcher, and each file was reviewed in multiple passes. A first pass established the segmental and prosodic alignment of tonal events. A second pass checked the consistency of pitch-accent and boundary-tone assignment across tokens belonging to the same sentence type and positional condition. Further inspection was used where necessary to resolve uncertain cases, especially near phrase boundaries or in tokens with reduced voicing. Where relevant, the same utterance was inspected in both raw and smoothed display formats in order to ensure that smoothing

aided interpretation without obscuring segmentally grounded tonal events. This multi-pass procedure was adopted in order to improve internal consistency and to avoid over-reliance on any single visual cue in the displayed contour.

Figures 3–5 provide worked examples of the annotation procedure from the pilot study. They are included not as representative summaries of the final tonal inventory established in the main study, but as illustrations of how pitch-accent and boundary-tone decisions were made in practice from the waveform, F0 track, and aligned annotation tiers.

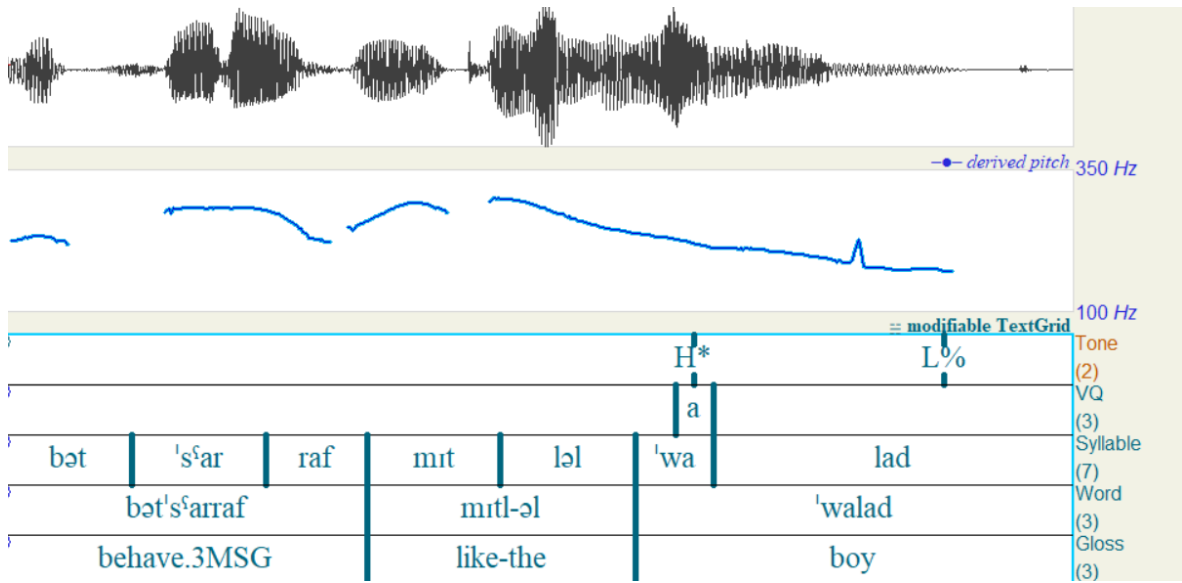


Figure 3 Pilot declarative token illustrating the annotation of H* and L%.

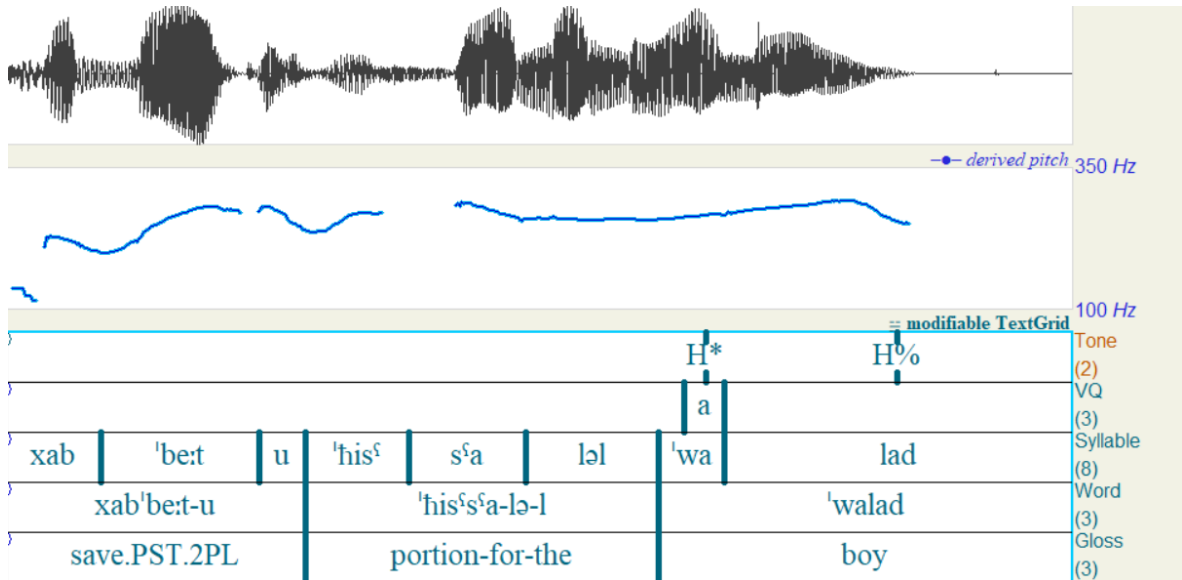


Figure 4 Pilot yes/no-question token illustrating pitch-accent and boundary-tone annotation.

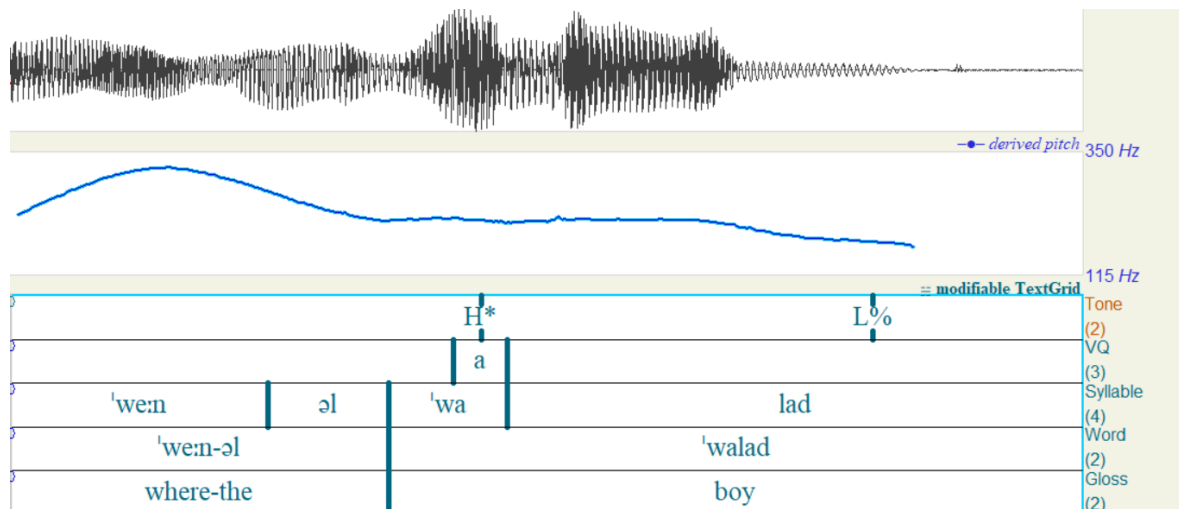


Figure 5 Pilot *wh*-question token illustrating annotation of a less expected contour.

The declarative and yes/no-question examples show relatively straightforward cases in which the local tonal event could be related clearly to the stressed syllable and the phrase edge. The *wh*-question example is methodologically useful for a different reason: it illustrates a less expected pilot contour, showing that early recordings could display greater variability and that tonal labels therefore had to be assigned through careful consideration of stressed-syllable alignment, phrase-edge position, and the broader utterance pattern, rather than by relying on sentence type alone.

This style of analysis is comparable to that adopted by Al Hasan and Mahanta (2022), who likewise interpret Syrian Arabic intonation through aligned Praat displays and relate tonal events to stressed syllables and prosodic boundaries. The present pilot figures follow the same general logic of interpretation, but are used here specifically to illustrate the practical annotation decisions that informed the later refinement of the main-study analysis.

The resulting labels were intended as phonological analyses of recurring intonational patterns, not as purely phonetic transcriptions of every local fluctuation in F0. In this respect, the procedure followed the general assumptions of AM analysis: contours were not labelled differently just because they looked slightly different on the surface, and they were not labelled the same just because they looked similar. Labels were assigned according to their relation to prominence and phrasing (Ladd, 2008). The goal of the intonational annotation was therefore to provide a principled representation of pitch accents and boundary tones that could later be related to prosodic position and to the acoustic correlates of phonation examined elsewhere in the dissertation.

3.7.2 Voice-quality analysis

Voice-related measures were extracted automatically from the vowel intervals marked in the VQ tier using a custom Praat script. For each token, the script opened the corresponding .wav file and its aligned TextGrid, retrieved the manually annotated VQ interval, and wrote the resulting measurements to a tab-delimited output file. In addition to the acoustic measures, the output included token metadata and annotation labels, namely, file name, speaker code, repetition round, sentence number, target word, gloss, VQ label, interval start time, interval end time, duration, jitterLocal%, shimmerLocal%, hnrMeanFulldB, f0MeanHz, intensityMean_dB, pitchAccent, and boundaryTone. In this way, each extracted value remained directly linked to the corresponding annotated interval and to the prosodic analysis. The pitch analysis method and settings follow Praat's documented options for periodicity analysis (Boersma & Weenink, 2024; Al-Tamimi, 2022). The custom extraction script can be made available as supplementary material or upon request.

Because perturbation- and noise-related measures are sensitive to pitch settings and segment definition, the script implemented a speaker-adaptive two-pass pitch procedure. In the first pass, Praat's cross-correlation pitch tracker was run with a broad range (50–600 Hz) and standard settings (time step 0.005 s; 15 candidates; silence threshold 0.03; voicing threshold 0.45; octave cost 0.01; octave-jump cost 0.35; voiced/unvoiced cost 0.14). The first and third quartiles of the resulting F0 distribution were then used to derive a speaker-specific pitch floor and ceiling ($\text{minPitch} = Q1 \times 0.75$; $\text{maxPitch} = Q3 \times 1.5$), and pitch was recalculated in a second pass using this adapted range. This two-pass procedure was intended to reduce octave errors and improve stability for periodicity-based measurements in speakers with different F0 ranges.

Harmonicity was then computed with Praat's To Harmonicity (cc) command using time step 0.005 s, minimum pitch = minPitch, silence threshold 0.1, and periods per window = 1. Intensity was computed with To Intensity using minPitch, default time step (0), and mean pressure subtraction enabled. Cycle boundaries were obtained through Praat's Sound + Pitch → To PointProcess (cc) workflow, which created a PointProcess from the sound and its associated pitch object. All subsequent perturbation measures were calculated over the exact start and end points of the annotated VQ interval rather than over the whole token.

Three primary voice-related measures were extracted. Jitter (local %) was calculated with a minimum period 0.0001 s, a maximum period 0.02 s, and a maximum period factor 1.3. Shimmer (local %) was calculated with the same minimum and maximum period values and the same maximum period factor 1.3, together with a maximum amplitude factor of 1.6. HNR (dB) was taken as the mean harmonicity value over the annotated interval. In addition, the script also exported mean F0 (Hz) and intensity for the same interval. The extraction thus combined manual interval definition with automated measurement using fixed script settings, allowing the final dataset to relate voice-related values directly to prosodic position and tonal labels. The resulting measurements were exported for statistical analysis.

3.8 Statistical analysis

All statistical analyses were conducted in R (R Core Team, 2025). Acoustic measures extracted in Praat were exported to Excel and imported into R with `readxl` (Wickham & Bryan, 2025). Data preparation was then carried out with functions from `dplyr`, `tidyr`, and `tibble` (Wickham et al., 2026a, 2026b; Müller & Wickham, 2026). In particular, information encoded in the file names was automatically extracted in R, including speaker code, repetition number, and sentence identifier, so that each observation could be linked to the corresponding audio file and annotation record. Mixed-effects models were fit with `lme4` (Bates et al., 2015), with inferential tests obtained via `lmerTest` (Kuznetsova et al., 2017). Estimated marginal means and post hoc contrasts were computed with `emmeans` (Lenth & Piaskowski, 2026). Visualizations were produced with `ggplot2` (Wickham, 2016), and tables in the HTML outputs were formatted with `knitr` (Xie, 2025).

Because the factorial structure of the dataset was not fully symmetrical across sentence types, the main acoustic analyses were conducted separately for statements, yes/no questions, and *wh*-questions, and separately for each dependent variable. The main reason for fitting sentence-type-specific models was that *wh*-questions did not include an utterance-initial lexical target position, since the initial slot was occupied by the *wh*-word itself. A single omnibus model combining sentence type and position would therefore have contained structurally non-equivalent cells and would not have provided a straightforward basis for interpreting positional effects across modalities. For this reason, the Chapter 5 analyses treated position within sentence type as the primary fixed effect of interest, rather than fitting a global Sentence Type $\times$ Position interaction. The three acoustic outcomes,

jitter (local %), shimmer (local %), and HNR (dB), were also modelled separately, since they represent distinct acoustic properties and were not assumed to vary identically across prosodic conditions.

Within each sentence type, linear mixed-effects models were fitted with position as a fixed effect and random intercepts for speaker and target word. The final reported models were fitted in lme4 by maximum likelihood (REML = FALSE), with inferential tests obtained through lmerTest. Their general form was:

Outcome ~ position + (1 | speaker) + (1 | targetWord)

Random intercepts were included because observations were repeated within speakers and the same three lexical items recurred across tokens, so both grouping factors could contribute baseline variation unrelated to the positional effect of interest. In this specification, the speaker intercept captured overall differences across individuals, while the target-word intercept captured baseline lexical differences that could arise from the segmental and acoustic properties of the recurring items. In other words, the model did not assume that lexical items would necessarily respond differently to position; rather, it allowed them to differ in their overall mean level while estimating the average positional effect across items.

During model development, simpler alternative specifications were also examined, including models with reduced or absent random-effects structure, in order to assess whether the inclusion of speaker- and item-level variation materially improved the fit and interpretation of the analysis. The final mixed-effects specification was retained because it offered the most appropriate balance between the repeated-measures design of the dataset, the need to account for baseline variation across speakers and items, and the requirement for a comparable model structure across sentence types and outcomes. More complex random-effects structures, such as random slopes, were not pursued in the main Chapter 5 analyses because the grouping structure was relatively limited (10 speakers and 3 recurring target words) and the principal goal at this stage was to estimate the average effect of position within sentence type rather than speaker-specific or item-specific position effects. Under these conditions, a more elaborate random-effects structure would have increased the risk of overparameterization without directly serving the central inferential aim of the chapter.

The main-study models converged successfully. One exception concerned the statement-jitter model, which produced a singular fit, with the speaker-level random-intercept variance estimated at or near zero. This was interpreted not as a failure of convergence, but as evidence that the data for that particular outcome provided little support for between-speaker baseline variation beyond the residual structure. The model was therefore retained, but its random-effects estimate was interpreted cautiously. When a richer fixed-effects structure became necessary in the later gender reanalysis, this was modeled explicitly through Gender, Position, and their interaction (Gender $\times$ Position), together with token-level F0 and intensity covariates, while retaining the same basic random-intercept structure for speaker and target word.

When the position factor had three levels (statements and yes/no questions), omnibus tests were followed by pairwise comparisons of estimated marginal means in order to determine which positional contrasts were driving any overall effect. These contrasts were computed with the emmeans package by first estimating marginal means for position and then comparing all available pairs with Tukey adjustment. Thus, for statements and yes/no questions, the reported contrasts were beginning–middle, beginning–end, and middle–end. In wh-questions, where only middle and end positions were available, the same procedure yielded a single middle–end comparison. In that case, the pairwise comparison was functionally equivalent to the model’s fixed-effect contrast, but it was retained for consistency of reporting across sentence types and outcomes. Thus, no custom or theory-driven subset of contrasts was specified at this stage; the post-hoc procedure always compared all available position levels within the relevant sentence type.

3.9 Limitations of methodology

Although the methodological choices in this dissertation were designed to support controlled elicitation and reproducible acoustic analysis, several limitations must be acknowledged.

First, the participant sample comprised ten speakers aged 20–30, all Syrians residing in Hungary and primarily students. This sampling strategy was appropriate for an instrumentally controlled study, but it necessarily limited the sociolinguistic breadth of the dataset. Damascene Syrian Arabic is embedded in regional, social, and generational variation, and a small diaspora-based sample cannot be taken as fully representative of that wider sociolinguistic space. The results should therefore be interpreted as evidence about

the intonation–phonation interface in this particular speaker group under these recording conditions, rather than as an exhaustive characterization of Syrian Arabic more generally.

Second, the study relied on a read-speech task recorded in a sound booth. This design improved signal quality, reduced unwanted environmental noise, and increased comparability across tokens, all of which were important for the extraction of perturbation- and periodicity-related acoustic measures. At the same time, read speech differs from spontaneous interaction in planning, discourse structure, phrasing, and pragmatic meaning. The findings should therefore be understood as reflecting laboratory-style productions rather than the full range of intonational and phonatory variation that may occur in everyday conversational speech.

Third, although three repetitions of each stimulus were elicited, only two repetitions per item were retained for the final analysis after manual inspection of all recorded tokens. This procedure improved the quality and interpretability of the analyzed dataset by excluding tokens with hesitations, performance errors, or reduced acoustic usability. However, it also means that the final dataset represents a curated subset of the elicited material rather than the full set of recorded repetitions. While this was methodologically justified, it remains a limitation insofar as the analyzed data do not capture the full range of token-to-token variability that may occur even within the same speaker and condition.

Fourth, the voice-related acoustic measures were restricted to jitter (local %), shimmer (local %), and HNR (dB). These measures capture aspects of cycle-to-cycle variability and periodicity and are widely used in instrumental voice analysis, but they do not exhaust the acoustic dimensions relevant to phonation. In particular, the study did not include spectral-tilt measures such as H1–H2 or related harmonic-amplitude metrics, nor did it include cepstral measures such as CPP or CPPS. Future work could extend the acoustic operationalization to include such measures and thereby provide a broader characterization of the phonatory correlates that may pattern with prosodic structure.

Fifth, the lexical basis of the study was intentionally narrow. The main experiment was built around three recurring target words, selected to maximize comparability across conditions and to allow controlled extraction of vowel-based acoustic measures. This was appropriate for the aims of the dissertation, but it also means that the results should not be interpreted as demonstrating complete lexical generality. Some portion of the observed variation may still reflect segmental or lexical properties of the particular items used, even

though the statistical models included target word as a random intercept in order to account for baseline item-level differences.

Finally, the statistical design was necessarily constrained by the structure of the dataset. Because *wh*-questions lacked an utterance-initial lexical target position, the positional factor was not fully symmetrical across sentence types. For that reason, the main analyses were conducted separately by sentence type, rather than through a single omnibus model combining sentence type and position. This was the most defensible choice for the available data, but it also means that cross-modality comparisons were handled indirectly rather than through one fully unified factorial model.

Taken together, these limitations mean that the results should be interpreted as evidence for how intonational structure and a restricted set of acoustic correlates pattern in controlled productions of Damascene Syrian Arabic by this speaker group, rather than as a comprehensive account of all prosodic and phonatory behavior in Syrian Arabic. At the same time, they point clearly to directions for future work, including broader sociolinguistic sampling, spontaneous-speech data, and expanded acoustic operationalization.

3.10 Summary

This chapter has outlined the methodology employed in the dissertation. The research design combined a pilot study with a larger main study, allowing the procedures and stimulus design to be refined before full-scale analysis. The chapter described the participants, materials, and recording procedures, then presented the annotation framework, the acoustic-analysis workflow, and the statistical modeling strategy used to evaluate the research questions.

Particular attention was given to the logic of the main-study design. The materials were constructed to examine the interaction of sentence modality and prosodic position under controlled recording conditions, while also respecting the clause-structural constraints of Damascene Syrian Arabic. Intonational structure was analysed manually within an AM/ToBI-style framework, using repeated auditory and visual inspection in Praat. Voice-related measures were extracted automatically from manually defined vowel intervals using a custom Praat script, with a speaker-adaptive pitch-setting procedure designed to improve the stability of periodicity-based measurements. Statistical analysis was conducted

in R through mixed-effects modeling, with separate models fitted by sentence type and dependent variable in order to reflect the asymmetrical positional structure of the dataset.

Taken together, the methodology was designed to maximize transparency, replicability, and analytical control while remaining closely tied to the dissertation's central questions about the intonation–phonation interface. The combination of controlled elicitation, manual prosodic annotation, and quantitative acoustic analysis provided a principled basis for examining how prosodic structure relates to selected acoustic correlates of phonation in Damascene Syrian Arabic. Although the study remains subject to the limitations outlined above, it provides a robust methodological foundation for the analyses that follow.

The next chapter presents the results of the pilot study. That initial stage served to evaluate the feasibility of the recording, annotation, and extraction procedures, to identify sources of unwanted variability, and to motivate the refinements that shaped the main experiment.

Chapter Four: Pilot Study

4.1 Introduction

This chapter reports a pilot study conducted as a preparatory step toward the full-scale investigation of the intonation–phonation interface in Damascene Syrian Arabic. As outlined in Chapter 3, the broader study examines how prosodic position and sentence modality relate to (i) intonational structure (analyzed in an AM/ToBI-style framework) and (ii) acoustic correlates commonly used in instrumental voice analysis—jitter, shimmer, and harmonics-to-noise ratio (HNR). The pilot study was undertaken to evaluate the feasibility of the elicitation materials and to verify that the planned workflow (recording → annotation → acoustic extraction → statistical analysis) could be implemented consistently before scaling to a multi-speaker dataset.

Pilot studies are widely used to identify practical and analytic problems that may not be evident at the design stage (e.g., unintentionally ambiguous stimuli, unanticipated phrasing patterns, or instability in analysis settings). They can increase the likelihood that the main study yields interpretable results, especially when based on a single speaker (van Teijlingen & Hundley, 2001). In this study, the pilot therefore functions primarily as a methodological testbed rather than as a standalone test of population-level hypotheses.

The pilot was conducted with one male native speaker of Syrian Arabic and used five target words embedded in sentences across three modalities: declaratives, yes/no questions, and *wh*-questions. Prosodic position was manipulated by placing target words in initial, medial, and final positions in declaratives and in medial and final positions in interrogatives, reflecting the structural constraints discussed in Chapter 3.

The pilot addressed two feasibility questions. First, could pitch accents and boundary tones be identified and labeled consistently in the pilot materials using the AM/ToBI-style criteria, supported by auditory judgment and Praat-based inspection of F0 contours? Second, did the extraction of jitter, shimmer, and HNR from the annotated vowel intervals yield a set of measurements that was sufficiently usable, interpretable, and internally consistent across tokens to support the planned statistical modeling in the main study? The pilot outcomes were used to identify sources of unwanted variability and to motivate refinements to stimulus structure, annotation practice, and analysis settings in the subsequent main experiment.

This chapter is organized as follows. Section 4.2 presents the pilot-specific aspects of the methodology, focusing on those features of the design that differed from the main study or later informed its refinement. Section 4.3 reports the pilot results for intonation and voice-quality correlates. Section 4.4 discusses the implications of the pilot for the dissertation's broader aims, Section 4.5 outlines the pilot's limitations, and Section 4.6 summarizes how the pilot informed the design of the main study reported in Chapter 5.

4.2 Methodology

The pilot study followed the general methodological framework outlined in Chapter 3. To avoid repetition, the present section focuses only on those aspects of the design that were specific to the pilot, or that later informed revisions to the main study. The pilot was intended as a feasibility-oriented stage in which the elicitation materials, recording setup, and annotation and acoustic-extraction workflow could be tested on a smaller scale before expansion to the multi-speaker main study.

Given its single-participant design, the pilot is treated primarily as a methodological test rather than as a basis for population-level inference. Any patterns observed in the pilot are therefore reported as preliminary observations whose main function is to motivate later refinements to stimulus structure, annotation practice, and extraction settings, rather than to serve as confirmatory evidence for the dissertation's core hypotheses.

Accordingly, the subsections below describe only the participant, materials, recording procedure, annotation conventions, and analytic steps that were specific to the pilot or that proved important for shaping the design of the main study.

4.2.1 Participant

The pilot study involved a single male native speaker of Syrian Arabic, aged 28, who was residing in Hungary at the time of recording. The participant reported extended residence in Damascus and regular use of the Damascene variety. For this reason, the pilot is treated as a preliminary test of the workflow on speech approximating the Damascene variety, while recognizing that a single speaker cannot represent the broader sociolinguistic or intra-dialectal range of Damascene Syrian Arabic.

The participant also reported no history of speech or hearing impairment. The use of one speaker is a clear limitation of the pilot design; however, it was adequate for the pilot's intended purpose, namely, to evaluate the feasibility and internal consistency of the

elicitation, annotation, and acoustic-extraction procedures before scaling the study up to the multi-speaker main dataset (van Teijlingen & Hundley, 2001). In this sense, the pilot participant served not as a basis for population-level inference but as a controlled test case for assessing whether the methodological workflow could be implemented successfully.

4.2.2 Materials

The pilot materials were designed as an exploratory version of the stimulus set later used in the main study. As outlined in Chapter 3, the pilot employed five target words embedded in sentence frames across three modalities: declaratives, yes/no questions, and *wh*-questions. The purpose of this larger lexical set was not to maximize lexical coverage, but to test whether target words could be embedded naturally in different sentence types and prosodic positions while still allowing interpretable intonational annotation and acoustic extraction.

Target-word position was manipulated within each modality. In declaratives, each target word occurred in initial, medial, and final positions. In the interrogative materials, however, target words were restricted to medial and final position. For *wh*-questions, this restriction followed directly from clause structure, since the *wh*-expression occupies the clause-initial position in the relevant Syrian Arabic constructions, leaving no comparable initial slot for an additional lexical target without changing the structure of the utterance (Al Hasan & Mahanta, 2022). In yes/no questions, the pilot materials likewise used a verb-initial interrogative frame, so an initial lexical target would have required additional fronting or topicalization. Because the pilot was intended as a feasibility test rather than a fully expanded factorial design, that extra structural manipulation was not introduced at this stage.

Unlike the later main-study materials, the pilot sentences were intentionally longer and more contextually elaborated, and some extended to two clauses. This was an exploratory choice: richer sentence contexts could sound more natural and allowed the pilot to test whether longer utterances would create complications for prosodic phrasing, tonal labeling, and acoustic extraction. At the same time, the pilot also showed the cost of this design. The longer materials introduced greater variability in phrasing and reduced comparability across tokens, which later motivated the shorter and more tightly controlled sentence frames adopted in the main study. In this sense, the pilot materials served not only as elicitation items, but also as a test of how much contextual complexity the annotation and measurement workflow could tolerate before the data became less comparable across

conditions. Table 8 gives representative examples of the pilot materials across sentence type and target-word position.

Table 8 Sample stimuli illustrating target word distribution by sentence type and position

Target Word	Sentence Type	Position	Example Sentence (Arabic)	IPA (simplified)	Translation
Kitab 'book'	Declarative	Initial	كتاب الطبخ القديم فيو وصفات أطيب	/kita:b it ^t təbx ilʔa'di:m fi:o waʃfa:t ʔat ^t jab/	The old cookbook has tastier recipes.
walad 'boy'	Yes/No question	Medial	تأخذ الولد يلعب بالحديقة؟	/ta:xod ʔilwalad jilʕab bilha'di:ʔa/	Are you taking the boy to play in the garden?
shazara 'tree'	Wh-question	Final	كيف لازم أهتم بها الشجرة؟ ما عم تنمو بسرعة.	/ke:f la:zem ʔih'tem bəhʃaʒʒara? ma ʕam 'tenmo bɪsɪrʕa/	How should I take care of this tree? It's not growing fast.

4.2.3 Recording procedure

The pilot recordings were made under the same general recording conditions described in Chapter 3: the session took place in the sound booth of the Hungarian Research Centre for Linguistics (Budapest), using a head-mounted condenser microphone and a recording format of 44.1 kHz and 16-bit. Stimuli were presented with SpeechRecorder in randomized order, and the participant was instructed to read the sentences naturally in a clear conversational style.

What was specific to the pilot was the size and composition of the recording set. Each sentence was produced three times, yielding 105 recorded tokens in total: 45 declaratives (5 target words $\times$ 3 positions $\times$ 3 repetitions), 30 yes/no questions (5 target words $\times$ 2 positions $\times$ 3 repetitions), and 30 wh-questions (5 target words $\times$ 2 positions $\times$ 3 repetitions). The pilot thus provided a sufficiently varied but still manageable dataset for testing the recording, annotation, and extraction workflow before the procedures were expanded and streamlined in the main study.

4.2.4 Annotation

The pilot annotation followed the general procedures described in Chapter 3 and is summarized here only in its pilot-specific form. For each recording, a Praat TextGrid was created and aligned with the waveform and spectrogram. As in the later main study, the annotation was designed to support both intonational analysis and the later extraction of voice-related acoustic measures.

The TextGrids included tiers for the utterance, words, vowel intervals selected for acoustic measurement, and intonational events. Intonation was annotated in an AM/ToBI-style framework through combined auditory judgment and visual inspection of the F0 contour, waveform, and spectrogram in Praat. Voice-related measures were not calculated during the annotation stage. Instead, the annotation defined the vowel intervals from which the later script extracted jitter (local %), shimmer (local %), and HNR (dB). In this sense, the pilot annotation served two related purposes: first, to test whether tonal labels could be assigned consistently in the more complex pilot materials, and second, to determine whether the defined measurement intervals were suitable for subsequent acoustic extraction. The pilot annotation therefore functioned not only as a labeling step, but also as a test of whether the intended prosodic and acoustic workflow could be implemented consistently in the longer and more variable pilot materials.

4.2.5 Acoustic analysis

The pilot analysis followed the general workflow described in Chapter 3: acoustic measures were extracted in Praat from the annotated vowel intervals using a custom script, exported to Excel, and then imported into R for cleaning and analysis. The resulting dataset was organized by sentence type and target-word position.

What differed from the later main study was the inferential status of the analysis. Because the pilot consisted of a single speaker, the statistical treatment was explicitly exploratory and within-speaker. Its purpose was not to support population-level inference, but to assess whether the extraction workflow yielded usable, interpretable measurements and whether the planned predictors could be operationalized meaningfully in the pilot materials. Accordingly, the pilot used multiple linear regression models in R (via `lm()`) to examine how jitter, shimmer, and HNR patterned as a function of target-word position and pitch-accent category in this speaker's productions. Any statistical output, including *p*-values, is therefore interpreted only as a description of model fit within the pilot design, not as

evidence that the same effects generalize to the broader Damascene Syrian Arabic population.

In addition, the pilot analysis served a diagnostic role for the acoustic-extraction procedure itself. For the perturbation measures, periodicity estimates were derived through Praat’s point-process workflow using the cross-correlation option. This made it possible to identify problematic tokens, including cases of unstable periodicity or segmentation difficulty, and thereby to refine the extraction settings later adopted in the main study. In this sense, the pilot analysis was not only an initial statistical exploration, but also a methodological check on the reliability and usability of the end-to-end workflow.

4.3 Results

The pilot results are reported with two aims: (i) to document whether the materials and annotation workflow yield consistent intonational labels for this speaker and stimulus set, and (ii) to identify any design or analytic issues that require refinement before the main study. The section first presents the intonational results, then the voice-quality results, and finally the exploratory statistical patterns that helped motivate later methodological adjustments.

4.3.1 Intonation results

Inspection of the F0 contours, supported by auditory judgment, revealed systematic differences in sentence type in the right-edge tonal patterns within the pilot dataset. Boundary tones were coded in a ToBI-style manner as L%, H%, and, for a minority of *wh*-question tokens, HL%. Their distribution is summarized in Table 9.

Table 9 Boundary tones across all utterances (pilot speaker)

Sentence type	N	Boundary tone	Count
Declarative	45	L%	45
<i>Wh</i> -question	30	L%	20
<i>Wh</i> -question		HL%	10
Yes-no question	30	H%	30

For this speaker and stimulus set, all declaratives were realized with a final L% boundary tone, and all yes/no questions were realized with a final H% boundary tone. The predominance of falling declaratives and rising yes/no questions is broadly consistent with

previous descriptions of Damascene/Syrian Arabic intonation, including the central role of phrase-final low endings in statements and phrase-final rises in yes/no questions (Kulk, Ode, & Woidich, 2003; Al Hasan & Mahanta, 2022).

Wh-questions in the pilot were most often coded with L% at the right edge (20/30), with a minority coded as HL% (10/30). The HL% label was assigned when the utterance-final region showed a two-step edge contour: a local rise in F0 in the final post-nuclear region followed immediately by a final fall to a low endpoint at the phrase boundary. By contrast, L% was assigned when the right edge showed a simple low ending, that is, a monotonic fall or low plateau without a clear pre-final upturn. In descriptive terms, the HL% cases therefore reflect a more complex phrase-final movement than the simpler L% cases (often analyzed as H-L%, i.e., a high phrase accent followed by a low boundary tone), but in this dissertation, they are notated as HL% for descriptive convenience (Aguilar et al., 2024). This is compatible with Al Hasan and Mahanta’s preliminary account of Syrian Arabic *wh*-questions, which notes that phrase-final “tails” may show either a mid-rise or a final fall, and relates such behavior to the exaggerated phrase-final drawl often observed in Syrian Arabic.

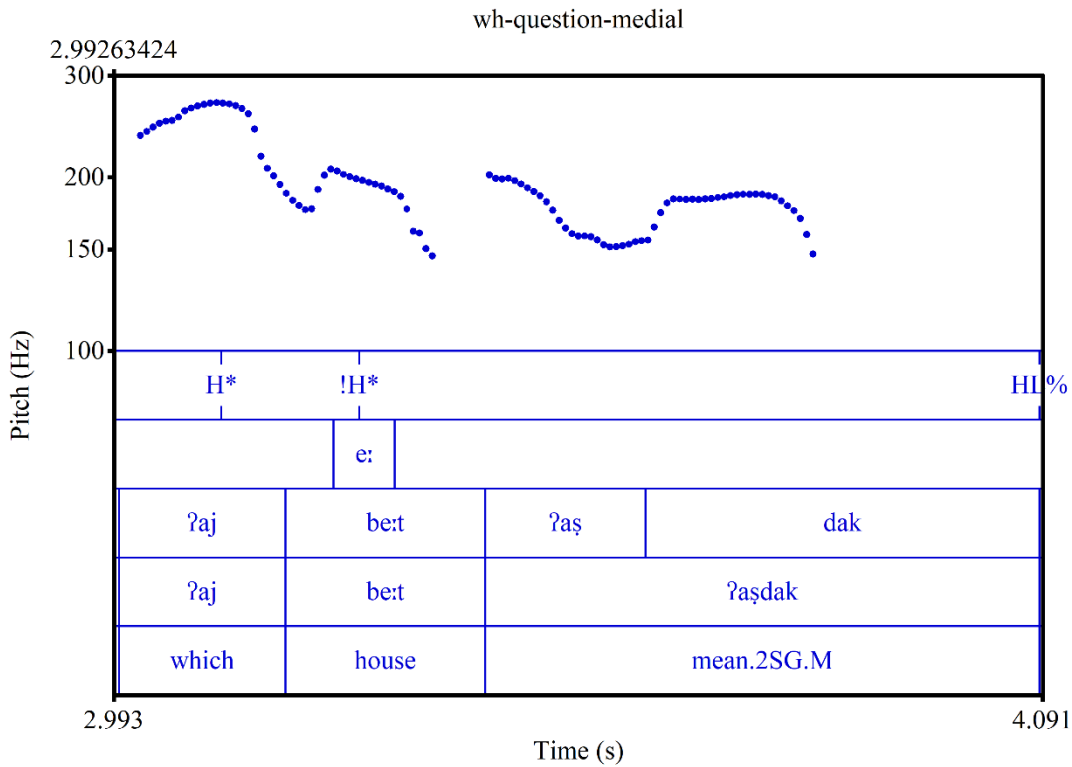


Figure 6 Pilot *wh*-question token illustrating the HL% boundary pattern.

Figure 6 provides a representative pilot token illustrating this less expected HL% pattern. The figure is included not as a summary of the final tonal system established later in the main study, but as a pilot-specific example of the contour variability encountered during early analysis. In this token, the right edge does not end in a simple low target; rather, the contour shows a brief pre-final rise followed by a final fall, which motivated the descriptive HL% label. The figure shows why pilot annotation could not rely on sentence type alone but required token-by-token evaluation of the final contour in relation to the phrase edge. This is also one respect in which the pilot resembles the preliminary observations of Al Hasan and Mahanta (2022), who report that Syrian Arabic *wh*-question tails may show different phrase-final realizations, including contours with a mid-rise before the final boundary movement.

Pitch accents were labeled in an AM/ToBI-style framework as tonal events associated with metrically prominent syllables. For this speaker, pitch accents were consistently anchored to stressed syllables, and the pilot materials enabled identification of recurrent accentual patterns across sentence types and target-word positions. In declaratives and yes/no questions, the accented syllable was typically associated with a high tonal event, while *wh*-questions showed greater contour variability in the pilot. Worked examples of the general annotation procedure, including how accentual events were distinguished from phrase-edge events, were already presented in Section 3.7.1 and are not repeated here.

Finally, it is important to distinguish the pilot design from the later main study. In the pilot, the yes/no questions were constructed with verb-initial interrogative frames, a common way of expressing yes/no questions in Syrian Arabic, and they were therefore not designed as string-identical counterparts of the declaratives. As a result, a clause-initial lexical target was not included in the pilot yes/no materials, since such a condition would have required additional fronting or topicalization beyond the exploratory design adopted at that stage. At the same time, this choice remains compatible with descriptive accounts of Damascene Syrian Arabic in which yes/no questions do not require obligatory interrogative particles and are often distinguished from declaratives primarily through intonation rather than morphology (Kulk et al., 2003). Only in the main study were more directly comparable statements and yes/no questions, with topicalization used to create an initial target-word condition.

4.3.2 Voice-quality results

This subsection summarizes the descriptive distribution of the three extracted acoustic measures, jitter (local %), shimmer (local %), and HNR (dB), in the pilot dataset and briefly comments on within-speaker tendencies across prosodic position and pitch-accent category. Because the pilot involved a single speaker, the results are treated as exploratory and are used primarily to evaluate whether the measurement workflow produced usable and interpretable acoustic outputs under the pilot design.

Across all pilot utterances, the overall mean values were 0.698% for jitter, 5.056% for shimmer, and 21.664 dB for HNR. The corresponding standard deviations were 0.298, 2.475, and 2.843, and the observed ranges were 0.128–1.601% for jitter, 1.523–16.340% for shimmer, and 16.534–29.226 dB for HNR (Table 10). These descriptive statistics provided a useful first estimate of the scale of variability that the later main-study design would need to control more tightly.

Table 10 Mean, standard deviation, and range of the acoustic measures across all pilot utterances

Measure	Mean	SD	Lowest	Highest
Jitter (%)	0.698	0.298	0.128	1.601
Shimmer (%)	5.056	2.475	1.523	16.340
HNR (dB)	21.664	2.843	16.534	29.226

The overall mean jitter remains within the broad range often reported in sustained-phonation studies of healthy young speakers, although such reference values should be interpreted with caution because they depend strongly on task type, recording conditions, and algorithmic settings (Teixeira et al., 2013; Brockmann et al., 2011). It suggests relatively stable periodic voicing, rather than supporting the identification of any particular phonation category, since jitter on its own cannot determine whether a token is creaky, breathy, or modal.

Shimmer values were more widely dispersed, which is not unexpected, since shimmer is sensitive to vocal intensity and measurement procedure (Brockmann et al., 2011). The mean HNR value, at 21.664 dB, suggests that the analyzed intervals were overall sufficiently periodic for Praat-based extraction to be feasible, since higher HNR values

indicate a signal with greater harmonic structure and relatively less noise (Boersma & Weenink, 2024).

Table 11 presents the same three measures by prosodic position. The positional means show that jitter was highest in the initial position (0.796%), lower in the medial position (0.717%), and lowest in the final position (0.641%). Shimmer was highest in the medial position (5.966%), compared with 4.658% initially and 4.367% finally. HNR showed a different pattern: it was 22.407 dB in the initial position, dropped to 20.690 dB in the medial position, and rose again to 22.275 dB in the final position.

Table 11 Mean acoustic values by position in the pilot dataset

position	N	Jitter Mean	Shimmer Mean	HNR Mean
beginning	18	0.796	4.658	22.407
middle	42	0.717	5.966	20.690
end	45	0.641	4.367	22.275

Figure 7 provides a compact visual summary of the raw pilot distributions by prosodic position, complementing the positional means reported in Table 11.

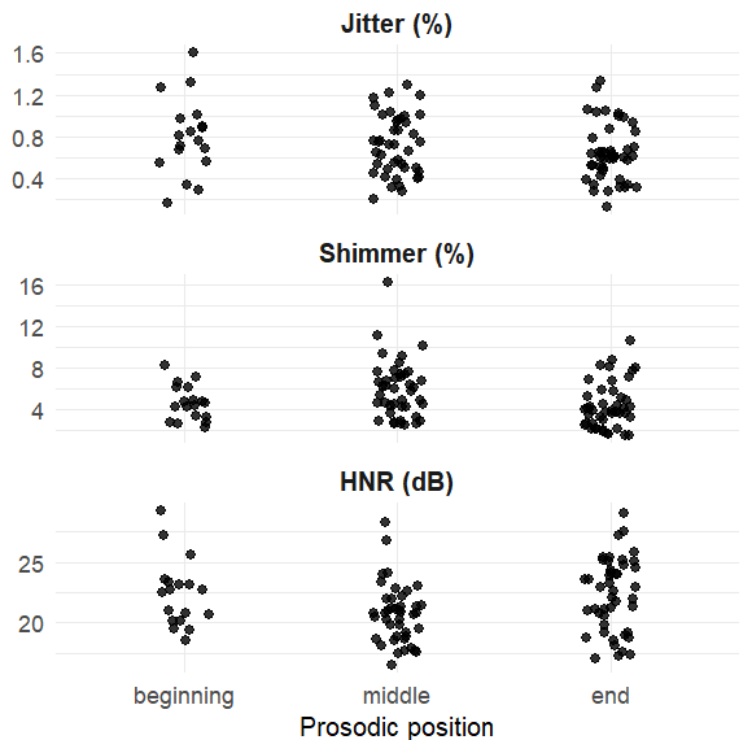


Figure 7 Raw pilot distributions of jitter, shimmer, and HNR by prosodic position.

These positional means make the pilot tendencies more explicit. Jitter was somewhat elevated at utterance onset relative to later positions, while shimmer was clearly highest in the medial condition. HNR, by contrast, did not show a simple monotonic decrease toward the utterance edge. In some English-focused work, phrase-final creak or reduced periodicity has been associated with lower HNR values near the right edge (Garellek, 2010; Keating et al., 2015), but that pattern was not apparent in this pilot speaker. The absence of a final reduction in HNR is therefore itself informative: it suggests that the selected phrase-final vowel intervals in this pilot did not systematically display noisier or less periodic phonation than earlier intervals.

Table 12 reports mean values by pitch-accent category. For jitter, the means were 0.677% for H*, 0.683% for !H*, and 0.970% for L+H*. For shimmer, the corresponding means were 5.000%, 5.087%, and 5.446%. For HNR, the means were 21.821 dB for H*, 21.069 dB for !H*, and 22.896 dB for L+H*. It is important to note, however, that the L+H* category in the pilot was based on only seven tokens, so these values should be interpreted with particular caution.

Table 12 Mean acoustic values by pitch-accent category in the pilot dataset

pitchAccent	N	Jitter Mean	Shimmer Mean	HNR Mean
H*	66	0.677	5.000	21.821
!H*	32	0.683	5.087	21.069
L+H*	7	0.970	5.446	22.896

Even so, Table 12 suggests that the L+H* tokens tended to show the highest mean values for all three acoustic measures in the pilot speaker. Because jitter and shimmer are both sensitive to local phonetic context, interval selection, and F0 tracking conditions, these differences are best treated as preliminary within-speaker tendencies rather than as evidence for a stable accent-specific phonation pattern. Their main value in the pilot was diagnostic: they helped show that acoustic values could vary with prosodic grouping factors and, therefore, justified tighter control of segmentation, extraction parameters, and stimulus structure in the main study.

Taken together, Tables 10–12 show that the pilot voice-quality measures were neither random nor uninterpretable. Rather, they yielded coherent descriptive patterns at the overall, positional, and accent-category levels, while also revealing enough variability to

motivate refinement of the later main-study design. In this sense, the pilot demonstrated that jitter, shimmer, and HNR could be extracted successfully from the annotated vowel intervals, while also showing why later analyses would require more controlled materials and more systematic statistical modeling.

4.3.3 Regression results

To explore whether prosodic factors patterned with the extracted acoustic measures in the pilot dataset, separate multiple linear regression models were fitted for jitter (local %), shimmer (local %), and HNR (dB), with phrase position and pitch-accent type as predictors. Because the pilot comprised a single speaker, these models are interpreted as exploratory within-speaker analyses rather than as a basis for population-level inference. Phrase-initial position and H* were used as the reference levels.

Table 13 Pilot regression results for jitter

Term	Estimate	SE	t	p
Intercept	0.731	0.073	9.965	0.000
Position: middle	-0.025	0.088	-0.287	0.775
Position: end	-0.125	0.087	-1.437	0.154
Pitch accent: !H*	0.036	0.065	0.555	0.580
Pitch accent: L+H*	0.293	0.120	2.437	0.017

Table 14 Model fit for jitter

Measure	Value
Observations	105
R-squared	0.091
Adjusted R-squared	0.054
Residual SE	0.290
F statistic	2.495

Table 13 presents the regression results for jitter. The model shows one significant effect: tokens labeled with L+H* had higher jitter values than the H* reference category ($\beta = 0.293$, $SE = 0.120$, $t = 2.437$, $p = .017$). Neither phrase-medial nor phrase-final position reached significance, and the !H* category also did not differ significantly from H*. The model fit was modest ($R^2 = .091$, adjusted $R^2 = .054$), indicating that the predictors accounted for only a limited proportion of variance in jitter. Within the pilot dataset, therefore, jitter variation appeared to be associated more clearly with the L+H* category than with phrase position. Given the small scale of the pilot and the very limited number of

L+H* tokens, this pattern should be treated cautiously as a preliminary within-speaker tendency rather than as evidence for a stable accent-specific phonatory effect.

Table 15 Pilot regression results for shimmer

Term	Estimate	SE	t	p
Intercept	4.409	0.604	7.296	0.000
Position: middle	1.499	0.721	2.078	0.040
Position: end	-0.195	0.715	-0.273	0.785
Pitch accent: !H*	0.184	0.538	0.343	0.733
Pitch accent: L+H*	1.120	0.989	1.132	0.260

Table 16 Model fit for shimmer

Measure	Value
Observations	105
R-squared	0.104
Adjusted R-squared	0.069
Residual SE	2.389
F statistic	2.916

Table 15 presents the regression results for shimmer. Here, phrase-medial position was associated with significantly higher shimmer than phrase-initial position ($\beta = 1.499$, $SE = 0.721$, $t = 2.078$, $p = .040$). No significant effects were found for phrase-final position, !H*, or L+H*. The model again showed only modest fit ($R^2 = .104$, adjusted $R^2 = .069$). These results suggest that, in the pilot speaker, shimmer varied more with prosodic position than with pitch-accent category, although this effect was limited to the medial position. As with the jitter findings, the result is best interpreted as exploratory and diagnostic rather than as a generalizable phonetic pattern.

Table 17 Pilot regression results for HNR

Term	Estimate	SE	t	p
Intercept	22.305	0.696	32.046	0.000
Position: middle	-1.345	0.831	-1.619	0.109
Position: end	0.308	0.823	0.374	0.709
Pitch accent: !H*	-0.872	0.620	-1.406	0.163
Pitch accent: L+H*	0.460	1.139	0.404	0.687

Table 18 Model fit for HNR

Measure	Value
Observations	105
R-squared	0.100
Adjusted R-squared	0.064
Residual SE	2.751
F statistic	2.775

Table 17 presents the regression results for HNR. In contrast to jitter and shimmer, none of the tested predictors reached significance. Phrase position and pitch-accent type, therefore, did not explain systematic variation in HNR in this pilot dataset. Model fit remained modest ($R^2 = .100$, adjusted $R^2 = .064$), suggesting that the included predictors captured little of the observed variance. Under the present extraction settings and measurement windows, periodicity as indexed by HNR thus appeared comparatively stable across the examined positional and accentual conditions in this speaker’s productions. At the same time, this null result should be interpreted cautiously, since variance in HNR may also reflect unmodeled segmental differences across the analyzed vowel intervals, including vowel-quality effects.

Taken together, the regression analyses support the pilot study’s methodological aim rather than providing strong inferential evidence. The workflow yielded analyzable outputs and highlighted potentially relevant tendencies in the relation between prosodic structure and acoustic voice measures, which informed the design of the multi-speaker main study.

4.4 Discussion

The pilot study was designed to test whether the experimental and analytic framework presented in Chapter 3 can be implemented successfully on Damascene Syrian Arabic data, specifically, whether the materials elicit the intended modal contrasts, whether AM/ToBI-style labeling can be applied consistently, and whether vowel-based extraction of jitter, shimmer, and HNR yields interpretable measurements under the chosen settings.

Within this pilot dataset, boundary-tone patterns emerged as expected across modalities: declaratives were consistently coded with L%, yes/no questions with H%, and *wh*-questions most often with L% (with a minority of HL% tokens). This aligns with prior

descriptions of Damascene Arabic intonation, which report robust falling declaratives and rising polar questions, while also documenting variability in interrogative realizations depending on phrasing and discourse context (Kulk et al., 2003; Al Hasan & Mahanta, 2022). Importantly, the pilot supports a methodological conclusion: the adopted AM/ToBI-style procedure (auditory judgment supported by Praat inspection) can be applied coherently to these materials for this speaker, and the stimuli generate contours that are sufficiently structured to annotate and compare across positions and modalities.

The pilot's descriptive and regression results suggest modest associations between prosodic configuration and perturbation measures in this speaker. Descriptively, jitter was highest in phrase-initial position, but in the regression model its only significant predictor was the L+H* category. Shimmer was significantly higher in phrase-medial position, whereas HNR did not show systematic variation with the tested predictors.

These tendencies are compatible with the general idea that prosodic structure can co-occur with changes in laryngeal and articulatory settings, particularly near prosodic boundaries, though the pilot cannot establish a mechanism. Work on prosodic strengthening shows that domain edges can condition systematic phonetic differences, including increased articulatory “strengthening” at prosodic boundaries (Fougeron & Keating, 1997). At the same time, jitter and shimmer are well known to be sensitive to F0, intensity, vowel quality, and measurement settings; these correlates show within-speaker differences across prosodic conditions under the chosen extraction procedure, and this motivates tighter parameter control in the main study.

The consistently high HNR values indicate relatively high periodic energy in the analyzed vowel windows for this speaker. However, the absence of an HNR effect does not imply that “creaky” or “breathy” phonation is absent, nor that phonation never varies with prosody. Creaky voice is not a single, uniform phenomenon, and different acoustic measures, such as cycle-to-cycle irregularity, spectral measures, and noise-related measures including HNR, can capture different aspects of it; consequently, a null result in one correlate does not preclude variation in others (Keating et al., 2015). In addition, phrase-final creak is often discussed in English and other languages as a prosodic phenomenon, but it is variable across speakers and contexts and may fall outside the specific vowel windows selected for measurement in a given protocol (Davidson, 2021; Keating et al., 2015).

Methodologically, the pilot validated the end-to-end pipeline of elicitation, annotation, acoustic extraction, and statistical modeling, while also highlighting concrete refinements that were implemented in the main experiment. The longer and sometimes bi-clausal pilot stimuli introduced additional phrasing variability, so the main study adopted shorter, more tightly controlled sentence frames. The pilot’s broader lexical set also added avoidable variability, which led to a reduction from five target words in the pilot to three in the main study. In addition, whereas yes/no questions in the pilot were limited to medial and final target-word positions, the main study used topicalisation to make an initial position available for that modality. More generally, because some positional cells remain structurally unavailable in interrogatives, especially in *wh*-questions, the main analyses were organized by modality rather than forced into a single full-factorial model. The pilot therefore did not simply demonstrate feasibility; it directly shaped the design choices that made the multi-speaker experiment more controlled and analytically tractable.

Although pitch accents were retained in the annotation and reported descriptively, they were not modeled inferentially in the main study. This is because the dissertation’s inferential focus is not on predicting tonal labels themselves, but on testing how annotated prosodic structure relates to continuous acoustic voice measures. Moreover, within each sentence type, the pitch-accent categories were relatively few and closely linked to the modal and positional patterns under description, making descriptive presentation more appropriate than separate outcome modeling.

4.5 Summary

This chapter has shown that the pilot study successfully tested the feasibility of the dissertation’s elicitation, annotation, and acoustic-extraction workflow on Damascene Syrian Arabic data. It also identified concrete design issues, especially stimulus complexity, positional comparability across modalities, and acoustic measurement control, that were addressed in the main study. The next chapter presents the results of the multi-speaker main experiment.

Chapter Five: Results of the Main Study (Modality and Position)

5.1 Introduction

This chapter reports the results of the main study investigating the intonation–phonation interface in Damascene Syrian Arabic. The main study used a controlled read-speech design to enable systematic comparison of sentence modality and target-word position with (i) AM/ToBI-style intonational structure, i.e., pitch accents and boundary tones, and (ii) three acoustic correlates derived from vowel intervals: jitter (local %), shimmer (local %), and HNR (dB). Full methodological details, such as participants, materials, recording, annotation, and extraction settings, are provided in Chapter 3.

The dataset comprises ten speakers: five females and five males, each producing three target words with three repetitions per applicable condition, yielding 72 utterances per speaker (27 declaratives, 27 yes/no questions, 18 *wh*-questions) and 720 utterances in total. Sentence type and position were manipulated as follows: declaratives and yes/no questions include initial, medial, and final target-word positions, while *wh*-questions include medial and final positions due to *wh*-initial structure.

Because *wh*-questions lack an initial-position level, statistical models are presented by sentence type, and each acoustic outcome is analyzed separately using mixed-effects modeling (see Chapter 3 for the analysis specification). Results are therefore organized to mirror this structure: for each sentence type, the chapter first reports the intonational patterns, then reports the model-based results for the three acoustic correlates, followed by a brief cross-type synthesis.

5.2 Statements

This subsection reports the main study results for declarative statements in Damascene Syrian Arabic. Analyses are presented (i) descriptively for the AM/ToBI-style tonal labels used in this dissertation and (ii) inferentially for three acoustic correlates of voice production: jitter, shimmer, and harmonics-to-noise ratio (HNR), modeled as a function of target-word position: initial, medial, and final.

5.2.1 Intonational labels in statements

Across the analyzed statement tokens (N = 180), the boundary-tone tier contained a single label: L% (n = 180). The pitch-accent tier contained two labels: H* (n = 120) and downstepped !H* (n = 60). Within the present dataset, statements therefore pattern with a consistent phrase-final low boundary tone (L%) and a high pitch accent that is realized either as a non-downstepped H* or as a downstepped !H* (Table 19; Figure 8).

Table 19 Counts of pitch accents and boundary tones in statements

type	category	label	n
statement	boundaryTone	L%	180
statement	pitchAccent	H*	120
statement	pitchAccent	!H*	60

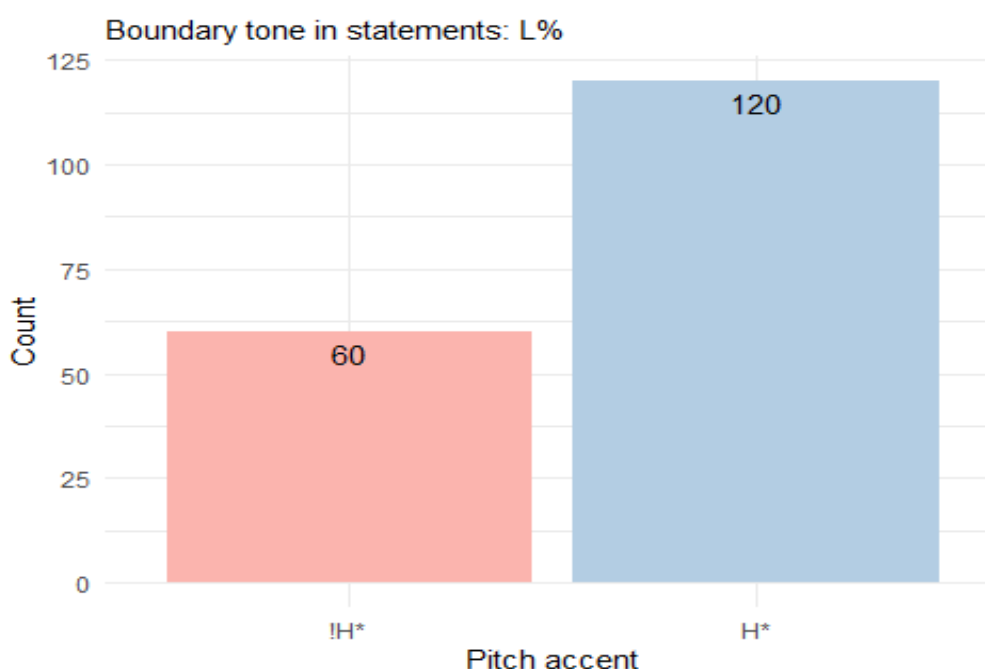


Figure 8 Visualising the distribution of pitch accents and boundary tones in statements

Figures 9-10 provide representative Praat illustrations for statement tokens in which the target word occurs in initial, medial, and final positions. Each figure displays the F0 track and the aligned TextGrid tiers used in this study: Tone, VQ, Syllable, Word, and Gloss. The Tone tier provides the AM/ToBI-style intonational labels: pitch accents and boundary tones, which are employed throughout the dissertation, such as H* and L%. These figures are included to make the segmentation and labeling decisions transparent and to provide representative token-level evidence for the tonal patterns summarized in the tables and figures above.

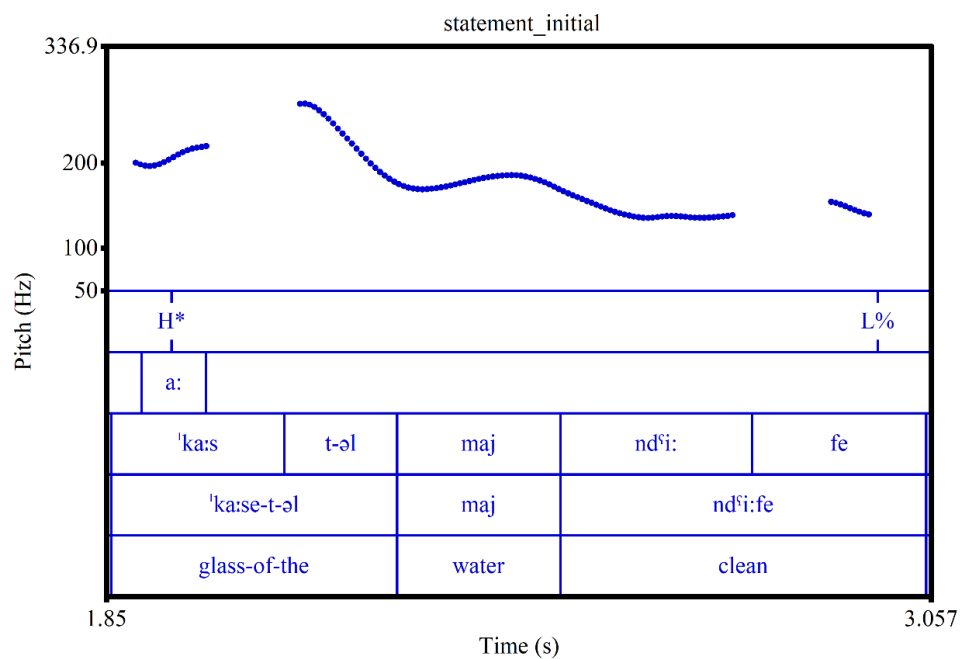


Figure 9 Statement (initial): F0 + tiers (H*, L%)

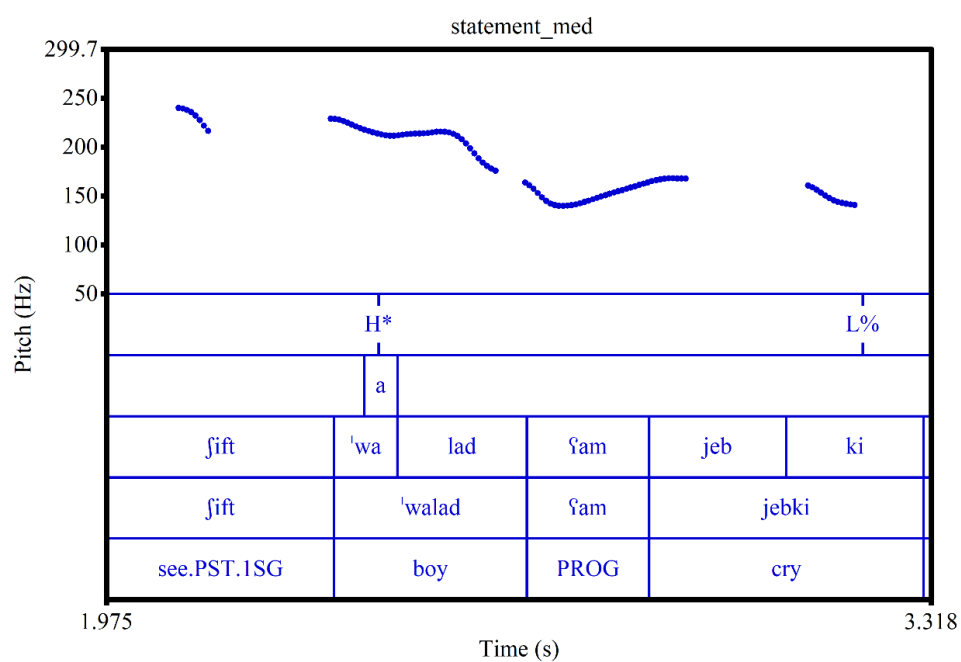


Figure 10 Statement (medial): F0 + tiers (H*, L%)

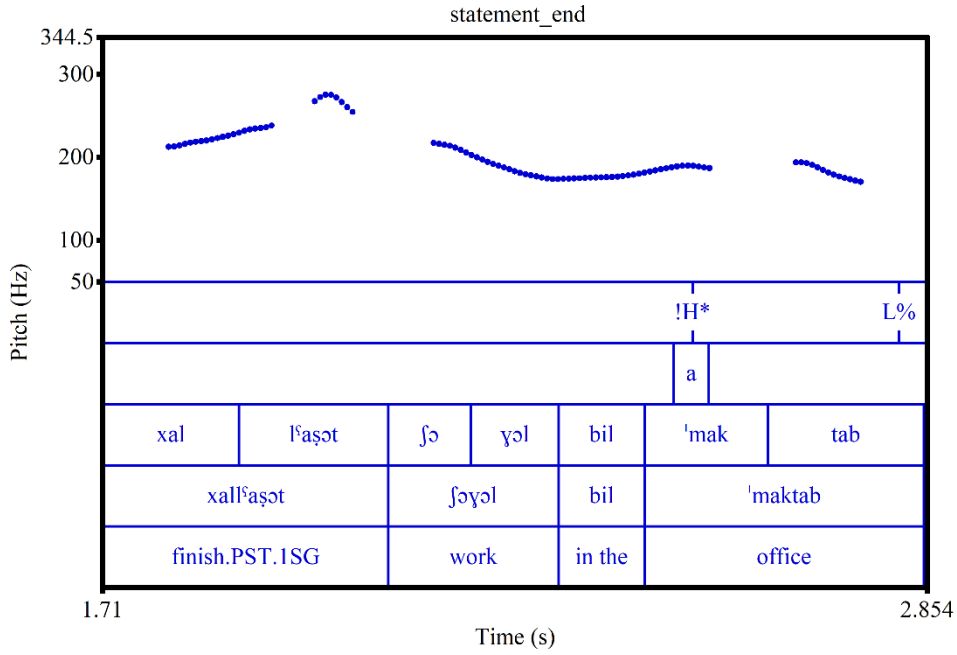


Figure 11 Statement (final): F0 + tiers (!H*, L%)

5.2.2 Jitter (local %)

Jitter was modeled using a linear mixed-effects model with fixed effect of position and random intercepts for speaker and target word (Table 20). The omnibus test provided weak evidence for an effect of position ($p = .071$; Appendix Table A1). Model-estimated means indicated the lowest jitter in medial position and the highest jitter in final position (Figure 12; Appendix A, Table A2). Tukey-adjusted pairwise contrasts were non-significant, although the medial-final contrast approached conventional thresholds ($p = .058$; Appendix A, Table A3). The model exhibited a boundary (singular) fit because the speaker-level random-intercept variance was estimated at (approximately) zero.

Table 20 Fixed effects for Jitter% in statements

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	0.885	0.084	10.484	19.733	0.000
fixed	positionmiddle	-0.125	0.115	-1.087	177.000	0.279
fixed	positionend	0.142	0.115	1.232	177.000	0.220

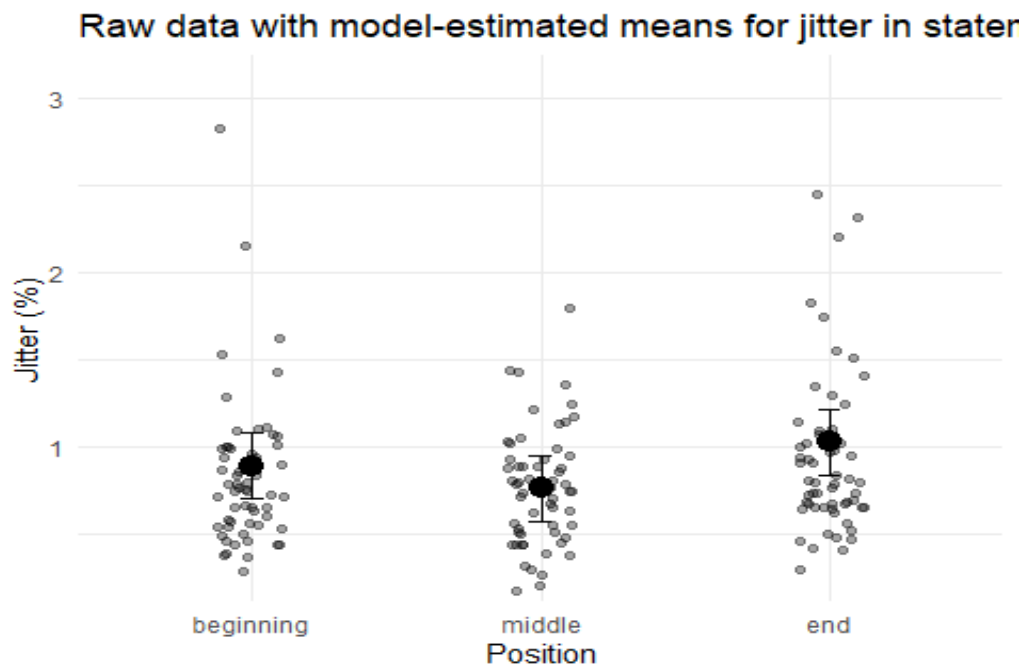


Figure 12 Raw data with model-estimated means for jitter in statements (1–99% zoom)

5.2.3 Shimmer (local %)

Shimmer was analyzed with the same modeling structure (Table 21). The omnibus test indicated no evidence for a position effect ($p = .372$; Appendix A, Table A1). Estimated means were similar across positions, and all Tukey-adjusted pairwise contrasts were non-significant (Figure 13; Appendix A, Tables A2–A3).

Table 21 Fixed effects for Shimmer% in statements

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	8.864	0.93	9.529	11.967	0.000
fixed	positionmiddle	-1.119	0.96	-1.165	167.358	0.246
fixed	positionend	0.101	0.96	0.105	167.358	0.917



Figure 13 Raw data with model-estimated means for shimmer in statements (1–99% zoom)

5.2.4 Harmonics-to-noise ratio (HNR, dB)

HNR was modeled as a function of position with random intercepts for speaker and target word (Table 22). The omnibus test indicated no evidence for a position effect ($p = .394$; Appendix A, Tables A1). Estimated means were highly similar across positions, and all Tukey-adjusted pairwise contrasts were non-significant (Figure 14; Appendix A, Tables A2–A3).

Table 22 Fixed effects for HnR in statements

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	14.255	1.285	11.093	7.870	0.000
fixed	positionmiddle	0.349	0.469	0.743	167.953	0.459
fixed	positionend	0.642	0.469	1.368	167.953	0.173

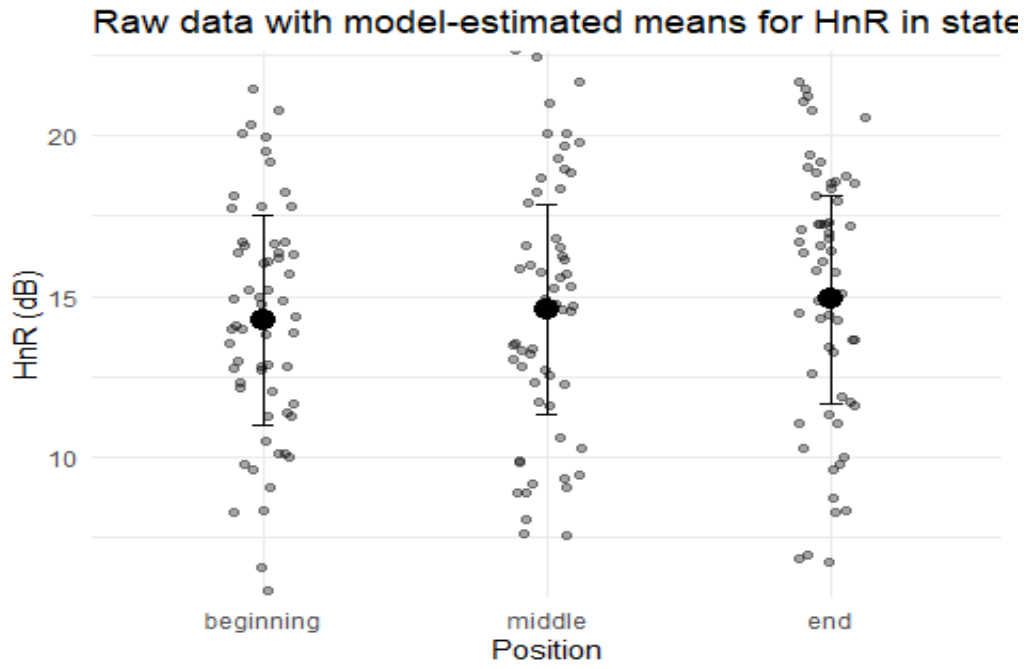


Figure 14 Raw data with model-estimated means for HnR in statements (1–99% zoom)

5.2.5 Summary

For statements, tonal annotation in this dataset yielded a uniform low boundary tone (L%) and high pitch accents (H* and !H*). Across the three voice-quality correlates analyzed, no robust position-related differences were observed. Jitter showed a weak trend toward higher values in final position relative to medial position, but this effect did not reach conventional significance after Tukey adjustment and should be treated as tentative pending the cross-modality comparison in later sections.

5.3 *Wh*-questions

This subsection reports the main study results for *wh*-questions in Damascene Syrian Arabic. Because the *wh*-element occupies the clause-initial slot in this stimulus set, target words were realized only in medial and final positions; accordingly, Position has two levels (middle, end). Intonational labels are summarized first, followed by mixed-effects analyses of acoustic voice-quality correlates: jitter, shimmer, and harmonics-to-noise ratio (HNR).

5.3.1 Intonational labels in *wh*-questions

Across the analyzed *wh*-question tokens (N = 120), the boundary-tone tier contained a single label (L%; n = 120). The pitch-accent tier contained two low-accent labels (L and L*), each occurring in 60 tokens (Table 23; Figure 15). These counts summarize the distribution of labels under the current annotation conventions.

Table 23 Counts of boundary tones and pitch accents in *wh*-questions

type	category	label	n
<i>wh</i> -question	boundaryTone	L%	120
<i>wh</i> -question	pitchAccent	L	60
<i>wh</i> -question	pitchAccent	L*	60

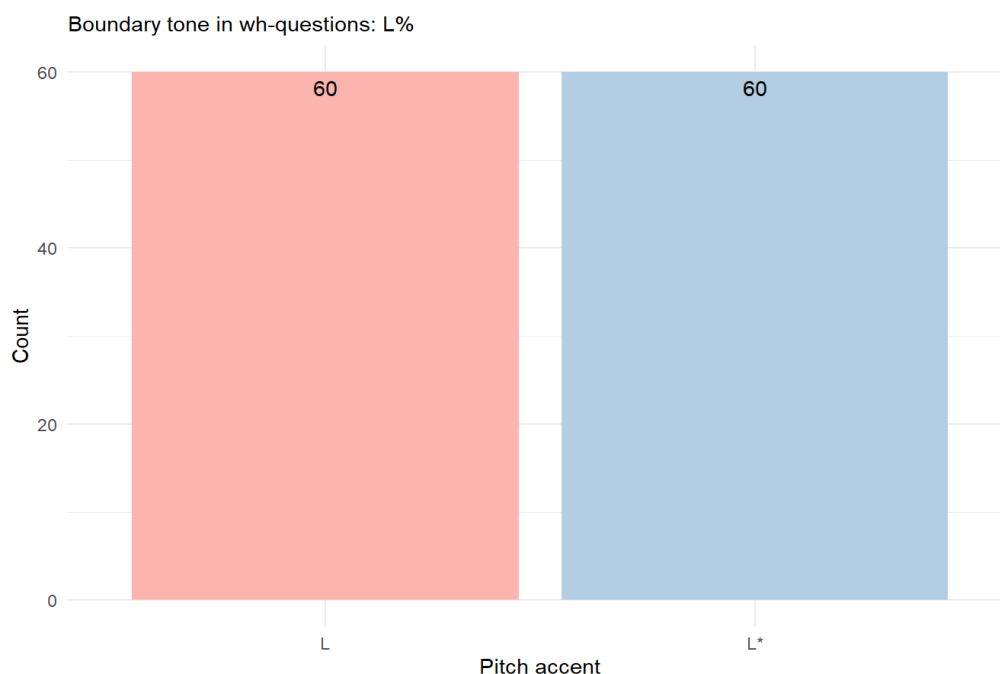


Figure 15 Distribution of pitch accents and boundary tones in *wh*-questions

Figures 16–17 present Praat picture-window figures for *wh*-question tokens in which the target word occurs in medial and final position. Each figure shows the F0 track and the aligned TextGrid tiers used in this study: Tone, VQ, Syllable, Word, and Gloss. The Tone tier contains AM/ToBI-style intonational labels, such as H* and L%. These figures are included to document the segmentation and labeling procedure for *wh*-questions and to provide illustrative token-level evidence corresponding to the tonal patterns summarized in the preceding tables and figures.

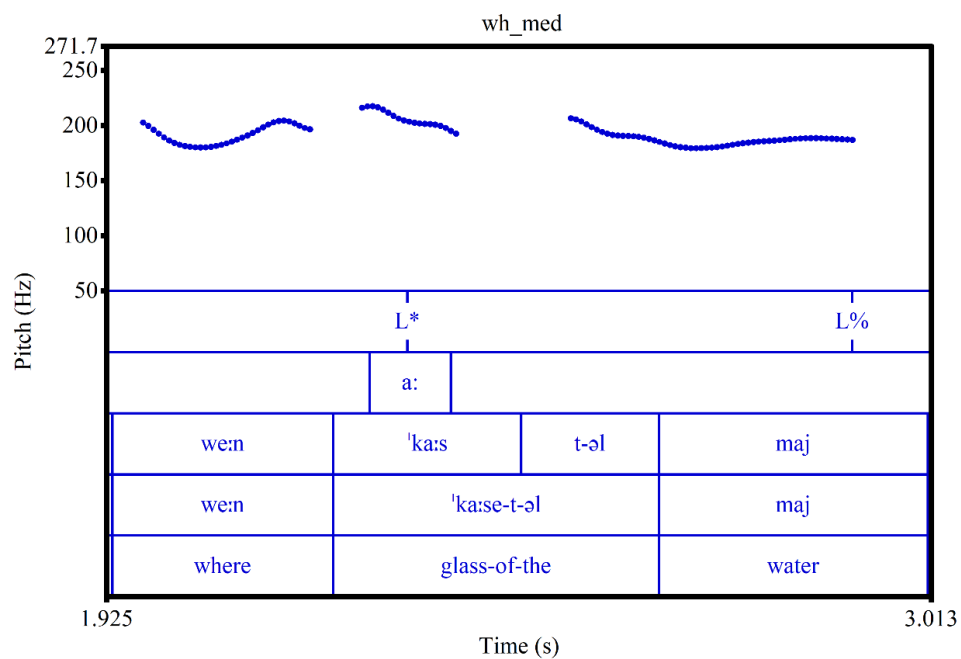


Figure 16 Wh-questions (medial): F0 + tiers (L*, L%)

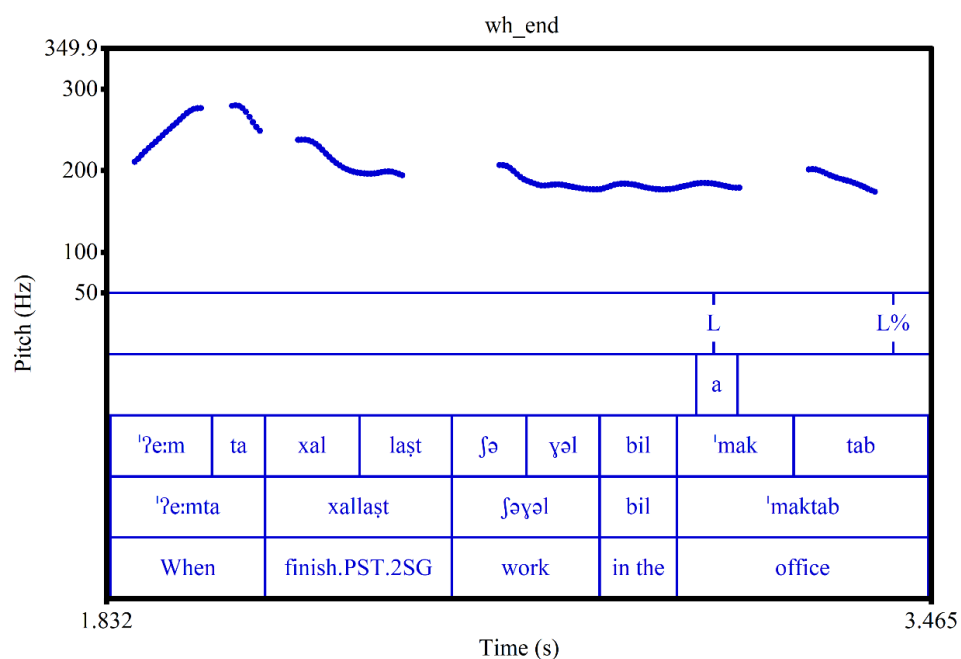


Figure 17 Wh-question (final): F0 + tiers (L, L%)

5.3.2 Jitter (local %)

Jitter (local %) was modeled using a linear mixed-effects model with Position as a fixed effect and random intercepts for speaker (code) and target word (N = 120; 10 speakers; 3 target words). The omnibus test indicated a reliable effect of Position ($p < .001$; Appendix

B, Table B1). Estimated marginal means were higher in the middle position than in the end position, and the Tukey-adjusted pairwise contrast was significant (Appendix B, Tables B2–B3). The fixed-effects estimates show the same pattern: relative to the middle position (intercept), jitter was significantly lower at the end (Figure 18; Table 24).

Table 24 Fixed effects for jitter (local %) in wh-questions (reference level: middle)

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	1.291	0.140	9.235	5.196	0
fixed	positionend	-0.567	0.076	-7.509	107.835	0

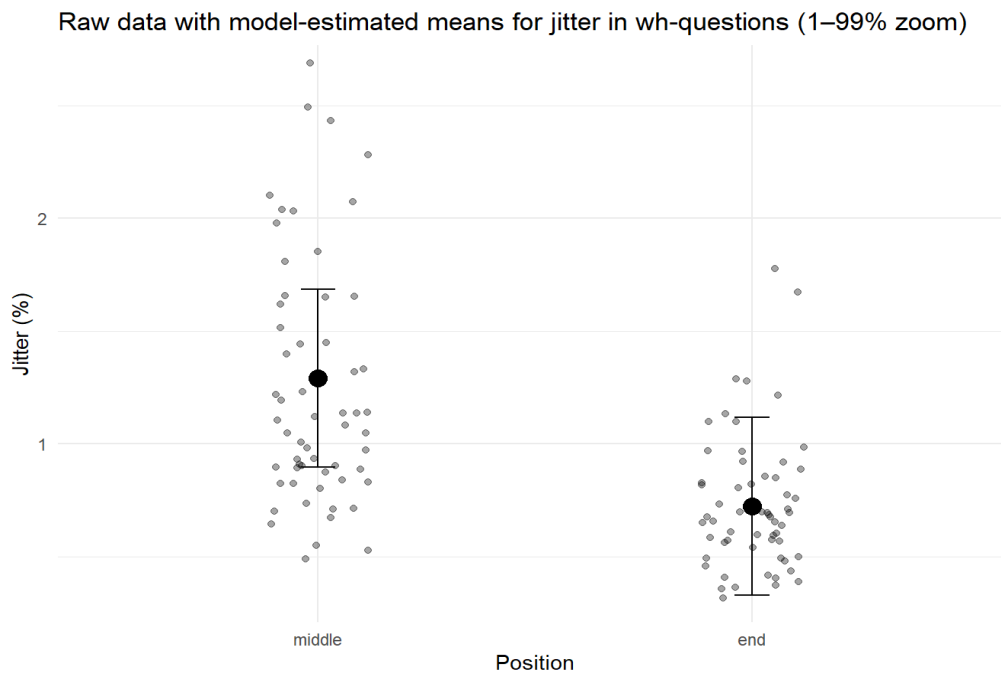


Figure 18 Raw jitter values with model-estimated means by position in wh-questions (1–99% zoom)

5.3.3 Shimmer (local %)

Shimmer (local %) was analyzed using the same modeling structure. The omnibus test indicated evidence for a Position effect ($p = .002$; Appendix B, Table B1). Estimated marginal means were higher in the middle position than in the end position, and the Tukey-adjusted pairwise contrast was significant (Appendix B, Tables B2–B3). In the fixed-effects table, the end position shows a significant decrease relative to the middle position (Figure 19; Table 25).

Table 25 Fixed effects for shimmer (local %) in wh-questions (reference level: middle)

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	7.470	0.711	10.502	4.217	0.000
fixed	positionend	-1.465	0.469	-3.125	107.737	0.002

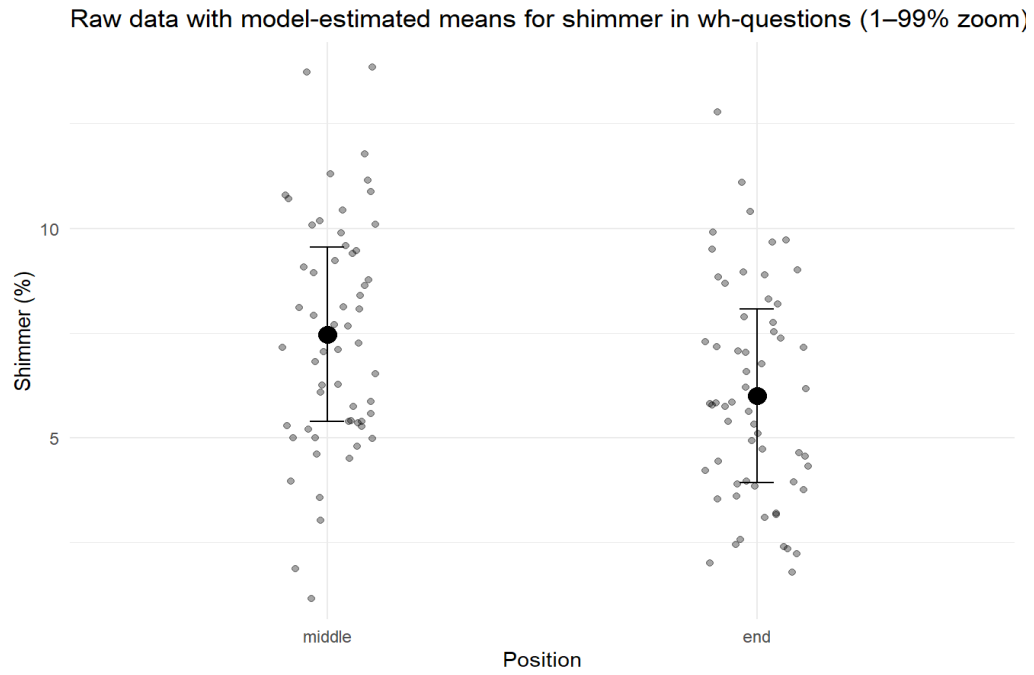


Figure 19 Raw shimmer values with model-estimated means by position in wh-questions (1–99% zoom)

5.3.4 Harmonics-to-noise ratio (HNR, dB)

HNR (dB) was modeled as a function of Position with random intercepts for speaker and target word. The omnibus test indicated evidence for a Position effect ($p < .001$; Appendix B, Table B1). Estimated marginal means were lower in the middle position than in the end position, and the Tukey-adjusted pairwise contrast was significant (Appendix B, Tables B2–B3). The fixed-effects estimates confirm a significant increase at the end position relative to the middle (Figure 20; Table 26).

Table 26 Fixed effects for HNR (dB) in wh-questions (reference level: middle)

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	14.585	1.389	10.498	10.297	0
fixed	positionend	1.599	0.419	3.814	107.952	0

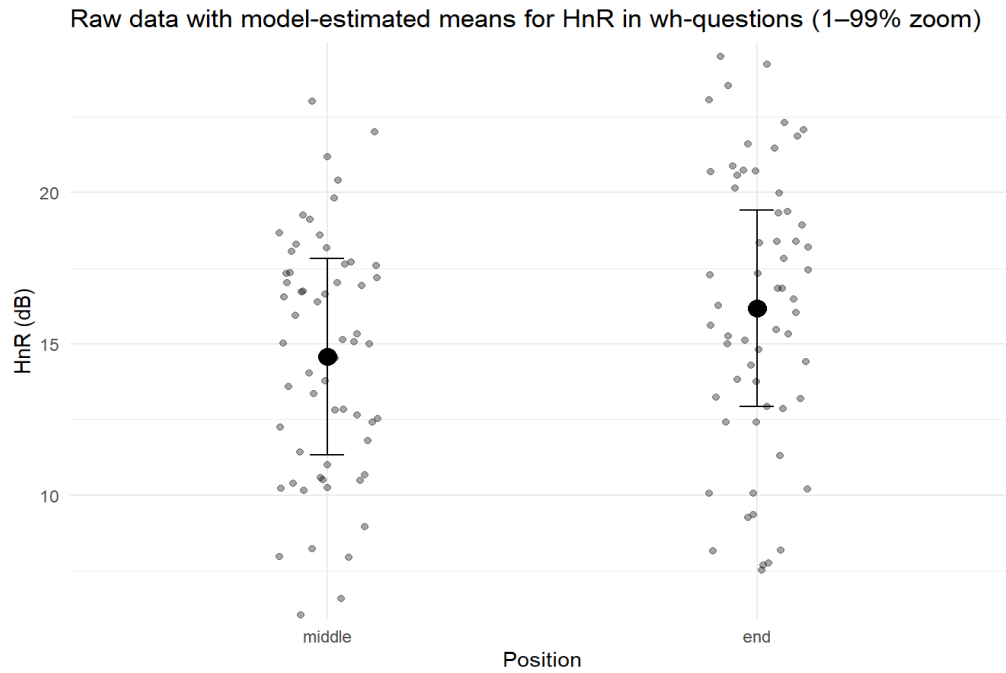


Figure 20 Raw HNR values with model-estimated means by position in *wh*-questions (1–99% zoom)

5.3.5 Summary

For *wh*-questions in this dataset, tonal annotation yielded a uniform low boundary tone (L%) and low pitch-accent labels (L and L*). Position was associated with systematic differences in acoustic correlates: jitter and shimmer were lower in the final than medial position, whereas HNR was higher in the final position.

5.4 Yes/no questions

This subsection reports the main study results for yes/no questions in Damascene Syrian Arabic. Analyses are presented (i) descriptively for the AM/ToBI-style tonal labels used in this dissertation and (ii) inferentially for three acoustic correlates of voice production: jitter, shimmer, and harmonics-to-noise ratio (HNR), modeled as a function of target-word position: beginning, middle, and end.

5.4.1 Intonational labels in yes/no questions

Across the analyzed yes/no-question tokens ($N = 180$), the boundary-tone tier contained a single label (H%; $n = 180$). The pitch-accent tier contained two labels: H* ($n = 120$) and H ($n = 60$). Within the present dataset, yes/no questions therefore pattern with a consistent phrase-final high boundary tone (H%) and a high pitch accent realized as either H* or H (Table 27; Figure 21).

Table 27 Counts of boundary tones and pitch accents in yes/no questions

type	category	label	n
yes-no question	boundaryTone	H%	180
yes-no question	pitchAccent	H*	120
yes-no question	pitchAccent	H	60

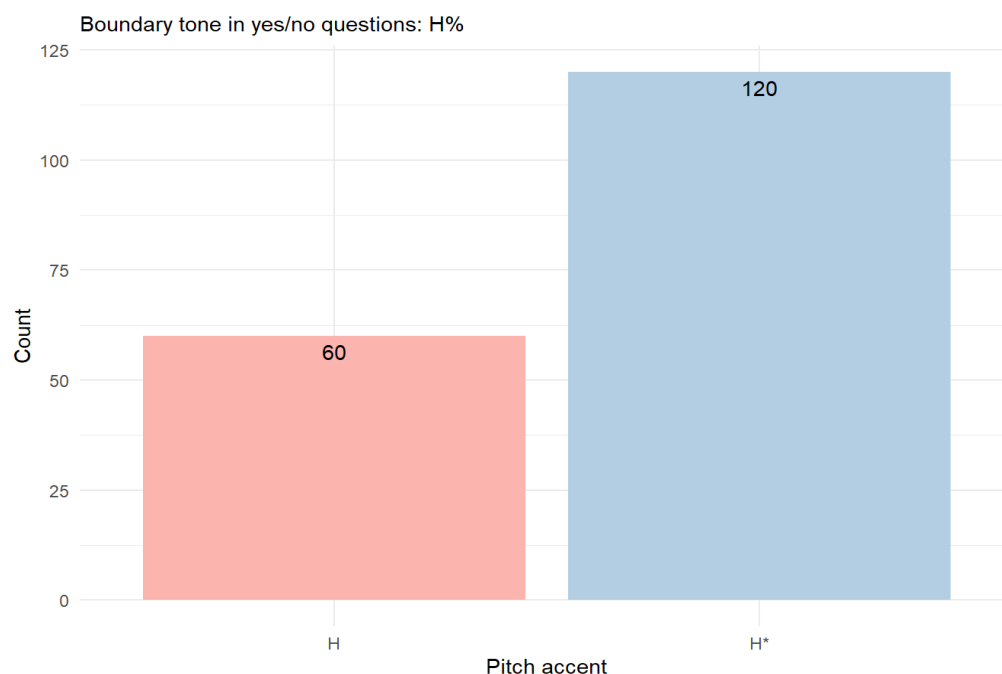


Figure 21 Distribution of pitch accents in yes/no questions (boundary tone uniformly H%)

Figures 22–24 show Praat picture-window figures for yes/no-question tokens with the target word in initial, medial, and final position. Each figure shows the displayed F0 contour and the aligned TextGrid annotation tiers used in this study: Tone, VQ, Syllable, Word, and Gloss, illustrating the alignment of tonal events with the target syllable and the segmentation used to define target-word position. The Tone tier contains AM/ToBI-style intonational labels, such as H* and H%. These figures are included to document the annotation procedure for yes/no questions and to provide token-level evidence corresponding to the tonal patterns summarized in the preceding tables and figures.

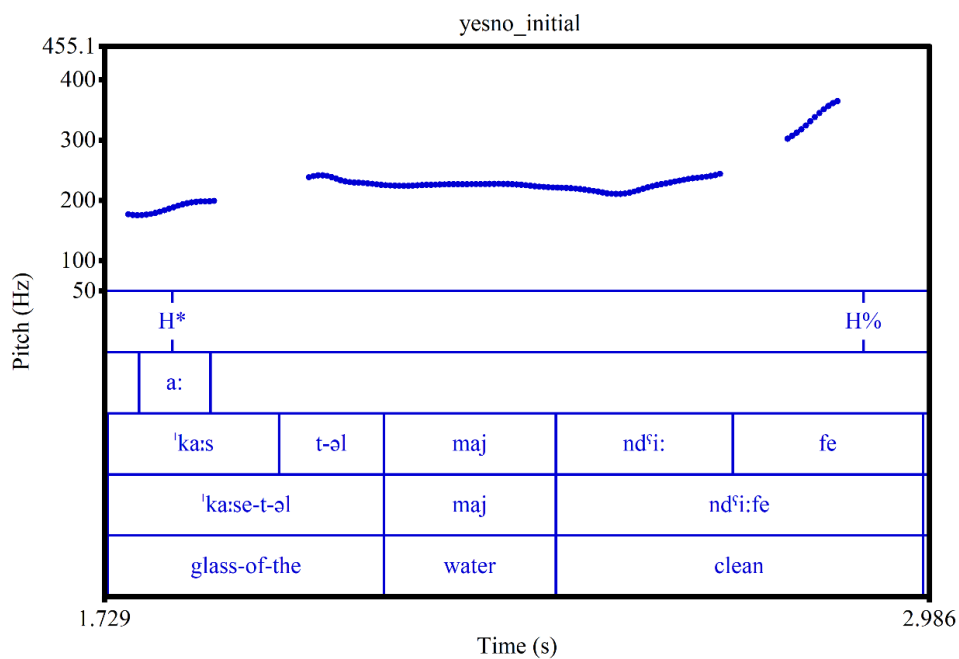


Figure 22 Yes/no questions (initial): F0 + tiers (H*, H%)

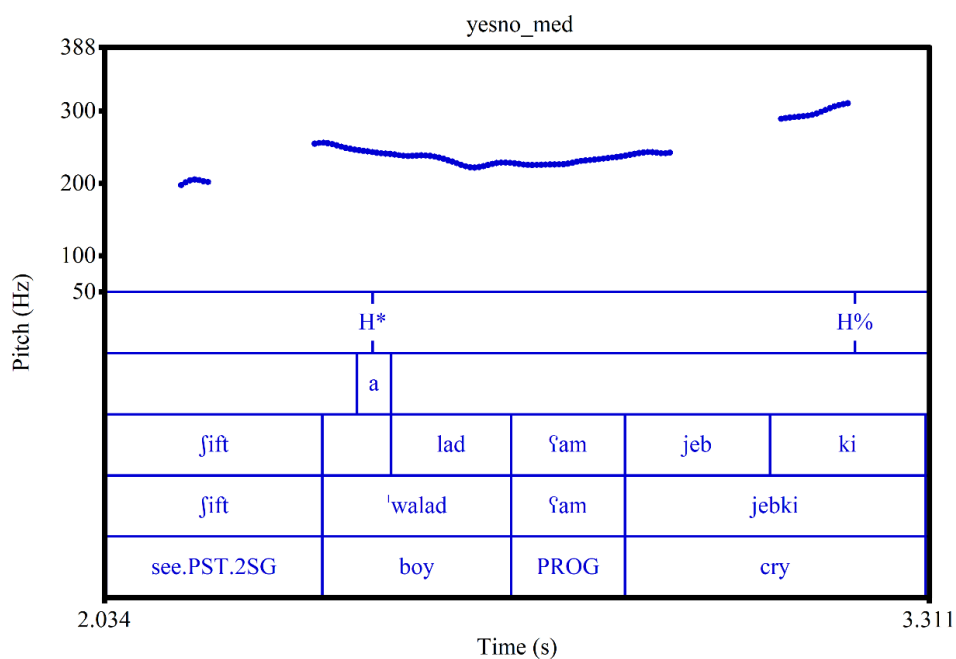


Figure 23 Yes/no questions (medial): F0 + tiers (H*, H%)

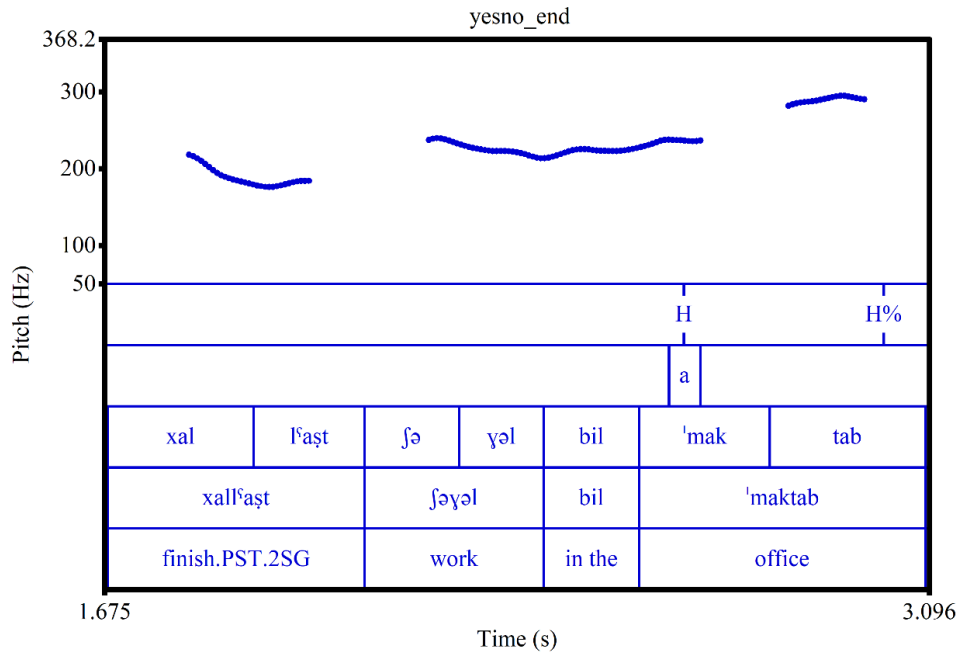


Figure 24 Yes/no questions (final): F0 + tiers (H, H%)

5.4.2 Jitter (local %)

Jitter (local %) was modeled using a linear mixed-effects model with Position as a fixed effect and random intercepts for speaker (code) and target word. The omnibus test did not indicate evidence for a Position effect ($p = .620$; Appendix C, Table C1). Estimated marginal means were similar across positions, and Tukey-adjusted pairwise contrasts were non-significant (Appendix C, Tables C2–C3). Consistent with this, relative to the beginning position (reference level), neither the middle position nor the end position differed reliably (Figure 25; Table 28).

Table 28 Fixed effects for jitter (local %) in yes/no questions (reference level: beginning)

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	0.966	0.093	10.440	14.027	0.000
fixed	positionmiddle	-0.082	0.120	-0.686	166.196	0.494
fixed	positionend	-0.114	0.120	-0.949	166.196	0.344

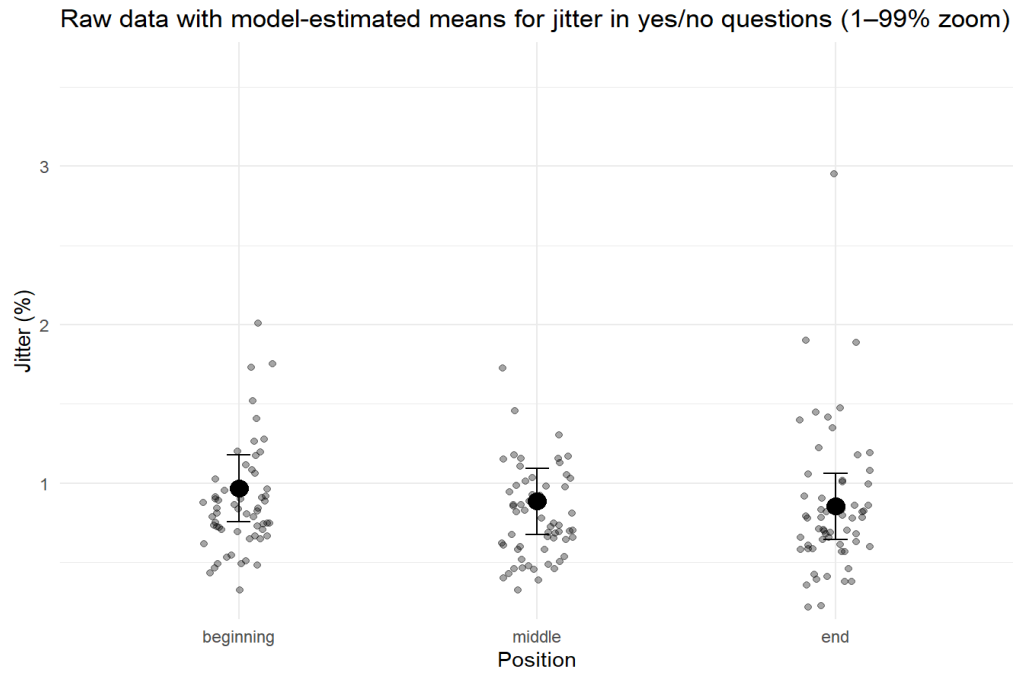


Figure 25 Raw jitter values with model-estimated means by position in yes/no questions (1–99% zoom)

5.4.3 Shimmer (local %)

Shimmer (local %) was analyzed with the same modeling structure. The omnibus test indicated no evidence for a Position effect ($p = .345$; Appendix C, Table C1). Estimated marginal means were broadly comparable across positions, and all Tukey-adjusted pairwise contrasts were non-significant (Appendix C, Tables C2–C3). Relative to the beginning position, the fixed-effects estimates did not indicate reliable differences at the middle or end positions (Figure 26; Table 29).

Table 29 Fixed effects for shimmer (local %) in yes/no questions (reference level: beginning)

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	9.026	1.192	7.569	6.415	0.000
fixed	positionmiddle	-0.988	0.684	-1.445	167.872	0.150
fixed	positionend	-0.629	0.684	-0.920	167.872	0.359

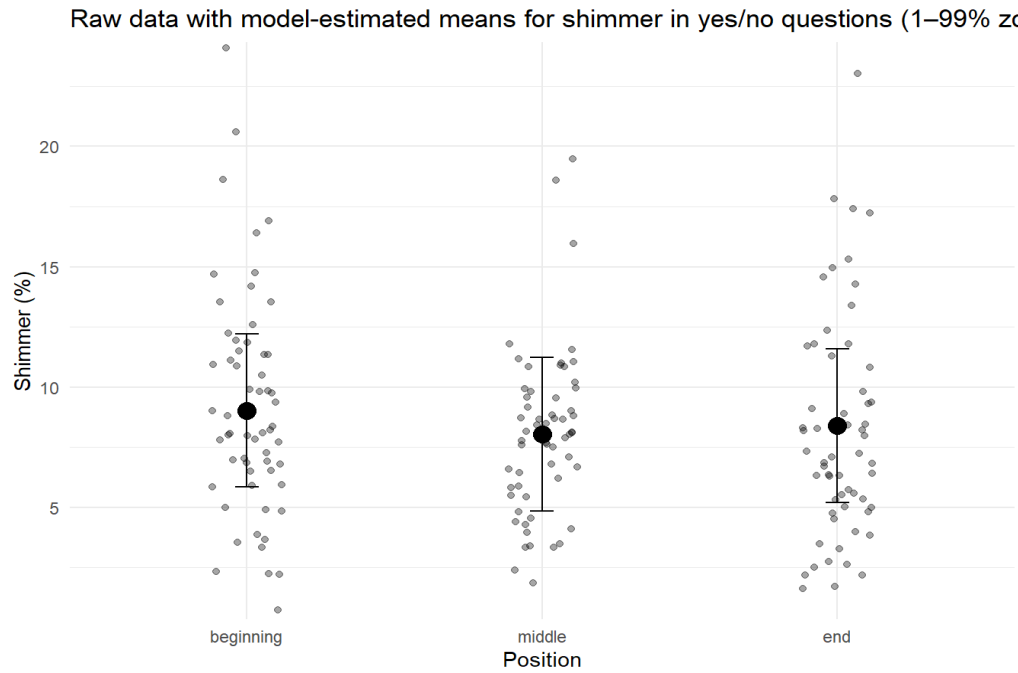


Figure 26 Raw shimmer values with model-estimated means by position in yes/no questions (1-99% zoom)

5.4.4 Harmonics-to-noise ratio (HNR, dB)

HNR (dB) was modeled as a function of Position with random intercepts for speaker and target word. The omnibus test indicated evidence for a Position effect ($p = .002$; Appendix C, Table C1). Estimated marginal means increased across positions (beginning: $M = 13.956$; middle: $M = 14.626$; end: $M = 15.506$; Appendix C, Table C2). The fixed-effects estimates show that the end position was associated with higher HNR than the beginning position ($p < .001$; Table 30; Figure 27). After Tukey adjustment, the beginning-end contrast remained significant ($p = .001$), whereas the beginning-middle and middle-end contrasts were not reliable (Appendix C, Table C3).

Table 30 Fixed effects for HNR (dB) in yes/no questions (reference level: beginning)

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	13.956	1.542	9.050	7.590	0.000
fixed	positionmiddle	0.670	0.431	1.555	167.973	0.122
fixed	positionend	1.550	0.431	3.594	167.973	0.000

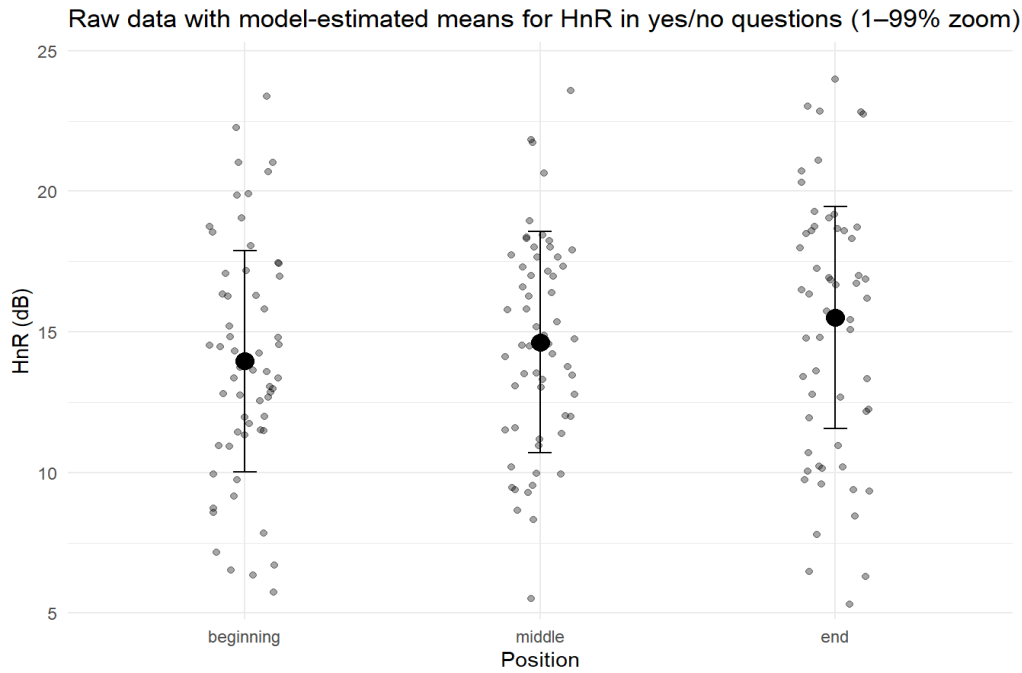


Figure 27 Raw HNR values with model-estimated means by position in yes/no questions (1–99% zoom)

5.4.5 Summary

For yes/no questions in this dataset, tonal annotation yielded a uniform high boundary tone (H%) and two pitch-accent labels (H* and H). Mixed-effects analyses indicated no reliable effect of target-word position on jitter or shimmer. By contrast, HNR showed a position-related increase, driven primarily by higher HNR in the end than in the beginning position.

Taken together, the analyses reported in this chapter show that position-related differences in acoustic voice-quality measures are not distributed uniformly across sentence modalities in the main-study corpus. Statements showed no robust positional differentiation, *wh*-questions showed the clearest medial-to-final shift across jitter, shimmer, and HNR, and yes/no questions showed a narrower effect, with HNR increasing toward the utterance end while jitter and shimmer remained comparatively stable. These findings establish the prosodic baseline for the dissertation’s main dataset, but they also motivate a further question: whether the observed modality-sensitive patterns are homogeneous across speakers or are additionally conditioned by speaker-related variation. Because the main-study corpus includes a gender-balanced speaker sample and the same modality-specific mixed-effects framework can be extended to include Gender and Gender $\times$ Position, the next chapter re-examines the corpus from that perspective. Its aim is not to replace the results presented here, but to refine them by asking whether the periodicity- and

perturbation-related patterns identified in Chapter 5 remain stable across male and female speakers, or whether some of them emerge more strongly in particular speaker groups and prosodic contexts.

Chapter Six: Gender Effects in the Main-Study Corpus

6.1 Introduction

Chapter Five established the main prosodic baseline of the dissertation by showing how sentence modality and target-word position condition intonational structure and acoustic voice measures in the main-study corpus. The results demonstrated that positional effects were not distributed uniformly across sentence types: statements showed little robust positional differentiation, *wh*-questions showed the clearest medial-to-final shifts across jitter, shimmer, and HNR, and yes/no questions showed a narrower pattern, with HNR increasing toward the utterance end while jitter and shimmer remained comparatively stable.

The present chapter extends that analysis by asking whether those modality- and position-related patterns remain stable once speaker gender is taken into account. More specifically, it examines whether the perturbation- and periodicity-related effects identified in Chapter Five are homogeneous across male and female speakers, or whether some of them emerge more strongly in particular speaker groups and prosodic contexts. In addition, because acoustic voice measures in connected speech are known to be sensitive to local pitch and intensity conditions, this chapter incorporates token-level F0 and token-level intensity into the statistical models in order to refine the interpretation of any apparent gender effects.

This chapter, therefore, does not reopen the broader theoretical background of the dissertation, nor does it repeat the full methodological account provided in Chapter Three. Instead, it builds directly on the main-study design and dataset already introduced there and re-examines the same corpus from a more specific analytical perspective. As in Chapter Five, the results are presented separately for statements, *wh*-questions, and yes/no questions, since the positional structure is not fully symmetrical across sentence types. The chapter concludes with a brief summary of what the gender-based reanalysis adds to the main-study findings before the dissertation moves to the general discussion chapter.

6.2 Specific analytical procedure

This chapter reanalyzes the main-study corpus presented in Chapters Three and Five from the perspective of gender-related variation in acoustic voice measures. The participants, materials, recording procedure, annotation workflow, and base acoustic extraction protocol are therefore not repeated here in full; they are identical to those described in Chapter

Three for the main study. In brief, the analysis draws on the same read-speech corpus of statements, *wh*-questions, and yes/no questions, based on three target words produced by ten speakers and annotated at the level of the target-word vowel interval for the extraction of jitter, shimmer, and harmonics-to-noise ratio (HNR).

The chapter-specific extension lies in the statistical modeling. Whereas Chapter Five focused on modality- and position-related differences in the main corpus, the present chapter asks whether those patterns remain stable once Gender is included and whether the interpretation of any apparent gender effects changes when local acoustic covariates are modeled directly. To that end, token-level mean F0 and token-level mean intensity were extracted from the same segmented vowel interval and incorporated into the mixed-effects analyses as continuous covariates. For modeling purposes, both covariates were centred and scaled.

As in Chapter Five, the data were analyzed separately by sentence type rather than in a single omnibus model, because the positional structure is not fully symmetrical across modalities: *wh*-questions do not contain an utterance-initial target-word position. For each sentence type and each dependent measure (jitter, shimmer, HNR), linear mixed-effects models were fitted with Gender, Position, and their interaction ($\text{Gender} \times \text{Position}$) as the main predictors of interest, with random intercepts for Speaker and Target Word. Token-level F0 and intensity were included as additional fixed covariates. Models were fitted by maximum likelihood in R using *lme4*, with inferential tests obtained via *lmerTest* and follow-up estimated marginal means and pairwise comparisons computed with *emmeans* where needed. In the post hoc comparisons and interaction plots, the covariates were held at their mean values.

Sensitivity checks excluding intensity were also examined in order to assess the stability of the estimated gender effects. However, the primary reported models retain both F0 and intensity, since the aim of the reanalysis is not simply to test for gender differences in isolation, but to determine whether the modality- and position-sensitive patterns identified in Chapter Five remain robust once local acoustic conditions are taken into account.

For clarity, the main text reports only the final models that include token-level F0 and intensity as covariates. Earlier unadjusted models are not discussed in detail here, but they are retained in the Appendix for transparency and comparison with the final analytical version of the chapter.

6.3 Statements

6.3.1 Jitter (local %)

In statements, jitter showed no significant effects once Gender, Position, token-level F0, and token-level intensity were modeled together. The coefficient for males was small and non-significant ($\beta = 0.070$, $SE = 0.279$, $p = 0.803$), and neither the middle position ($\beta = 0.008$, $p = 0.959$) nor the end position ($\beta = 0.017$, $p = 0.933$) differed from the beginning. The Gender $\times$ Position interaction was also non-significant, and neither F0 nor intensity emerged as a significant covariate. Thus, in statements, jitter showed no reliable dependence on gender, position, or their interaction under the covariate-controlled model. See Tables 31–32 for details.

Table 31 Fixed effects for Jitter% in statements after controlling for F0 and intensity

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	0.929	0.175	5.324	110.696	0.000
fixed	genderM	0.070	0.279	0.250	175.000	0.803
fixed	positionmiddle	0.008	0.163	0.052	177.728	0.959
fixed	positionend	0.017	0.204	0.085	179.683	0.933
fixed	f0_z	-0.138	0.103	-1.341	179.285	0.182
fixed	intensity_z	-0.069	0.079	-0.876	65.399	0.384
fixed	genderM $\times$ positionmiddle	-0.339	0.227	-1.492	176.676	0.138
fixed	genderM $\times$ positionend	-0.151	0.232	-0.650	176.902	0.517

Table 32 ANOVA for Fixed effects for Jitter% in statements after controlling for F0 and intensity

Term	NumDF	DenDF	F value	p value
gender	1	164.356	0.172	0.679
position	2	177.195	1.041	0.355
f0_z	1	179.285	1.797	0.182
intensity_z	1	65.399	0.767	0.384
gender $\times$ position	2	176.873	1.120	0.328

6.3.2 Shimmer (local %)

In statements, shimmer did not show significant main effects of gender or position. Males were not significantly different from females at the reference position ($\beta = 1.005$, $SE = 3.085$, $p = 0.747$), and neither the middle position ($\beta = -0.213$, $p = 0.876$) nor the end position ($\beta = 1.711$, $p = 0.362$) differed reliably from the beginning. The overall Gender $\times$ Position interaction was not significant in the Type III test, although the Male $\times$ End coefficient remained significant ($\beta = -4.549$, $SE = 1.970$, $p = 0.022$), indicating that the

gender contrast was reduced and reversed at utterance end. Neither F0 nor intensity was significant. Taken together, these results suggest only limited and position-specific gender differentiation in shimmer for statements. See Tables 33–34 for details.

Table 33 Fixed effects for Shimmer% in statements after controlling for F0 and intensity

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	8.614	1.939	4.443	25.021	0.000
fixed	genderM	1.005	3.085	0.326	27.470	0.747
fixed	positionmiddle	-0.213	1.362	-0.156	171.459	0.876
fixed	positionend	1.711	1.872	0.914	134.083	0.362
fixed	f0_z	-0.855	1.161	-0.736	30.461	0.467
fixed	intensity_z	0.042	0.814	0.051	43.659	0.959
fixed	genderM $\times$ positionmiddle	-1.997	1.892	-1.056	167.760	0.293
fixed	genderM $\times$ positionend	-4.549	1.970	-2.309	177.226	0.022

Table 34 ANOVA for Fixed effects for Shimmer% in statements after controlling for F0 and intensity

Term	NumDF	DenDF	F value	p value
gender	1	20.215	0.198	0.661
position	2	158.573	0.827	0.439
f0_z	1	30.461	0.542	0.467
intensity_z	1	43.659	0.003	0.959
gender $\times$ position	2	172.315	2.678	0.072

6.3.3 Harmonics-to-noise ratio (HNR, dB)

In statements, HNR did not show a significant gender effect once token-level F0 and intensity were included in the model. The coefficient for males was negative but non-significant ($\beta = -1.127$, $SE = 1.943$, $p = 0.565$), and the Type III test for gender was likewise non-significant. Position also remained non-significant, and there was no reliable Gender $\times$ Position interaction. By contrast, F0 emerged as a significant positive predictor of HNR ($\beta = 1.764$, $SE = 0.681$, $p = 0.011$), while intensity was not significant. Thus, in statements, the gender-related HNR pattern is not robust once local acoustic covariates are taken into account; instead, HNR is more clearly related to token-level F0. See Tables 35–36 and Figure 28 for details. The corresponding unadjusted model is provided in Appendix D for comparison.

Table 35 Fixed effects for HNR in statements (controlling for F0 and intensity)

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	14.551	1.528	9.521	13.662	0.000
fixed	genderM	-1.127	1.943	-0.580	36.470	0.565
fixed	positionmiddle	0.648	0.663	0.977	169.516	0.330
fixed	positionend	1.128	0.975	1.158	172.871	0.249
fixed	f0_z	1.764	0.681	2.589	94.409	0.011
fixed	intensity_z	-0.588	0.475	-1.239	152.026	0.217
fixed	genderM $\times$ positionmiddle	-0.498	0.917	-0.543	168.289	0.588
fixed	genderM $\times$ positionend	0.532	0.973	0.547	175.812	0.585

Table 36 ANOVA for fixed effects (HNR in statements, controlling for F0 and intensity)

Term	NumDF	DenDF	F value	p value
gender	1	27.821	0.411	0.527
position	2	174.329	1.631	0.199
f0_z	1	94.409	6.704	0.011
intensity_z	1	152.026	1.536	0.217
gender $\times$ position	2	171.553	0.620	0.539

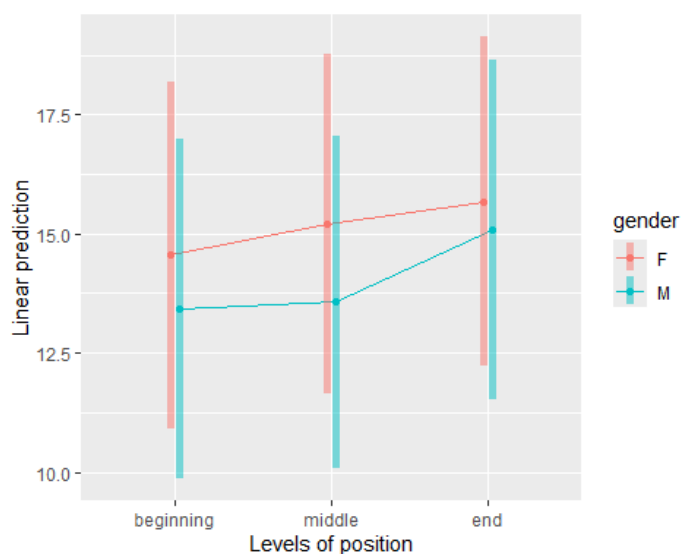


Figure 28 Estimated marginal means of HNR in statements by gender and position from the mixed-effects model including token-level F0 and intensity as covariates. Error bars indicate 95% confidence intervals.

6.4 Wh-questions

6.4.1 Jitter (local %)

In *wh*-questions, jitter showed a significant position effect and a significant Gender $\times$ Position interaction after controlling for F0 and intensity. Jitter was lower at the end than in the middle position ($\beta = -0.411$, $SE = 0.127$, $p = 0.002$), and the Male $\times$ End interaction

was also significant ($\beta = -0.359$, $SE = 0.150$, $p = 0.018$), indicating that the gender contrast was smaller at the utterance end than at the middle position. The gender coefficient at the middle position was positive but not significant in the refitted model ($\beta = 0.370$, $p = 0.135$), and neither F0 nor intensity was significant. These results show that the *wh*-question jitter pattern remains structured by prosodic position and by the interaction between gender and position, although the direct gender contrast is weaker than it appears in the unadjusted model. See Tables 37–38 and Figure 29 for details. The unadjusted version of this model is included in Appendix E as a supplementary analysis.

Table 37 Fixed effects for Jitter in wh-questions after controlling for F0 and intensity

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	1.118	0.191	5.845	13.591	0.000
fixed	genderM	0.370	0.243	1.524	44.122	0.135
fixed	positionend	-0.411	0.127	-3.247	115.484	0.002
fixed	f0_z	-0.040	0.098	-0.407	63.466	0.686
fixed	intensity_z	0.002	0.064	0.027	94.650	0.979
fixed	genderM $\times$ positionend	-0.359	0.150	-2.401	108.550	0.018

Table 38 ANOVA for fixed effects for Jitter in Wh-questions after controlling for F0 and intensity

Term	NumDF	DenDF	F value	p value
gender	1	34.874	0.751	0.392
position	1	108.075	35.552	0.000
f0_z	1	63.466	0.165	0.686
intensity_z	1	94.650	0.001	0.979
gender $\times$ position	1	108.550	5.767	0.018

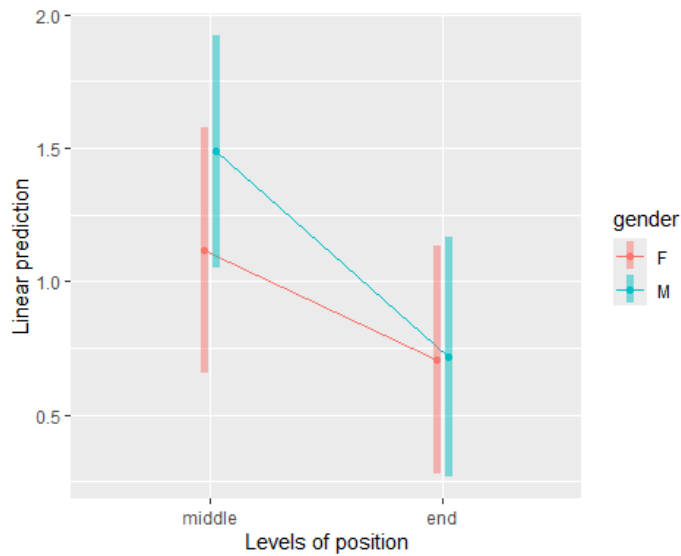


Figure 29 Estimated marginal means of jitter in *wh*-questions by gender and position from the mixed-effects model including token-level *F0* and intensity as covariates. Error bars indicate 95% confidence intervals.

6.4.2 Shimmer (local %)

In *wh*-questions, shimmer showed no significant effects in the covariate-controlled model. Gender was non-significant ($\beta = 1.652$, $SE = 1.281$, $p = 0.200$), position was non-significant ($\beta = -0.674$, $p = 0.381$), and the Gender $\times$ Position interaction was also non-significant. Neither *F0* nor intensity emerged as a significant covariate. Accordingly, the earlier position effect suggested by the unadjusted model is not retained once *F0* and intensity are included. See Tables 39–40 for details. For the corresponding unadjusted model, see Appendix F.

Table 39 Fixed effects for Shimmer in *wh*-questions after controlling for *F0* and intensity

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	6.349	0.953	6.666	16.440	0.000
fixed	genderM	1.652	1.281	1.290	119.431	0.200
fixed	positionend	-0.674	0.766	-0.880	117.511	0.381
fixed	f0_z	0.697	0.519	1.342	119.288	0.182
fixed	intensity_z	0.190	0.361	0.527	118.378	0.599
fixed	genderM $\times$ positionend	-0.401	0.955	-0.420	117.023	0.675

Table 40 ANOVA for fixed effects for Shimmer in *Wh*-questions after controlling for F0 and intensity

Term	NumDF	DenDF	F value	<i>p</i> value
gender	1	118.834	1.674	0.198
position	1	117.700	2.223	0.139
f0_z	1	119.288	1.801	0.182
intensity_z	1	118.378	0.277	0.599
gender × position	1	117.023	0.176	0.675

6.4.3 Harmonics-to-noise ratio (HNR, dB)

In *wh*-questions, HNR continued to show a strong position effect and a significant positive association with F0. HNR was significantly higher at the end than in the middle position ($\beta = 2.692$, $SE = 0.720$, $p < 0.001$), and F0 was also a significant positive predictor ($\beta = 1.585$, $SE = 0.595$, $p = 0.009$). Intensity was not significant, and the Gender × Position interaction was not significant. The coefficient for males at the middle position was negative and marginal in the coefficient table ($\beta = -3.067$, $p = 0.064$), while the overall Type III test indicated a significant gender effect. See Tables 41–42 for details. The corresponding unadjusted model is provided in Appendix G for comparison.

Table 41 Fixed effects for HNR in *wh*-questions after controlling for F0 and intensity

Effect	Term	Estimate (β)	SE	t value	df	<i>p</i> -value
fixed	(Intercept)	15.758	1.407	11.199	10.394	0.000
fixed	genderM	-3.067	1.603	-1.913	36.448	0.064
fixed	positionend	2.692	0.720	3.739	117.520	0.000
fixed	f0_z	1.585	0.595	2.665	108.268	0.009
fixed	intensity_z	-0.363	0.378	-0.961	119.161	0.338
fixed	genderM × positionend	-0.750	0.818	-0.917	107.195	0.361

Table 42 ANOVA for fixed effects for HNR in *Wh*-questions after controlling for F0 and intensity

Term	NumDF	DenDF	F value	<i>p</i> value
gender	1	29.268	5.342	0.028
position	1	117.745	16.472	0.000
f0_z	1	108.268	7.101	0.009
intensity_z	1	119.161	0.924	0.338
gender × position	1	107.195	0.841	0.361

Follow-up contrasts showed that the female–male HNR difference was not significant in the middle position but remained significant at the end position. Thus, the gender-related HNR pattern in *wh*-questions is more limited and position-sensitive than the unadjusted

results suggested, whereas utterance-final position and F0 remain the clearest predictors. See Tables 43–44 and Figure 30 for details.

Table 43 Simple effects of Gender within each Position (HnR in Wh-questions, controlling for F0 and intensity)

position	contrast	Estimate	Std. Error	df	t value	p value
middle	F - M	3.067	1.733	49.245	1.769	0.083
end	F - M	3.817	1.607	41.064	2.376	0.022

Table 44 Pairwise comparisons of Position within each Gender (HnR in Wh-questions, controlling for F0 and intensity)

gender	contrast	Estimate	Std. Error	df	t value	p value
F	middle - end	-2.692	0.748	122.451	-3.599	0.000
M	middle - end	-1.942	0.707	120.553	-2.747	0.007

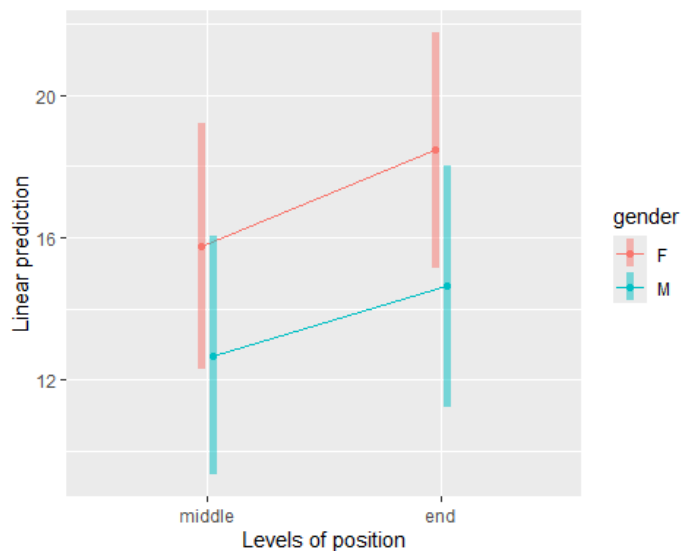


Figure 30 Estimated marginal means of HNR in wh-questions by gender and position from the mixed-effects model including token-level F0 and intensity as covariates. Error bars indicate 95% confidence intervals.

6.5 Yes/no questions

6.5.1 Jitter (local %)

In yes/no questions, jitter showed no significant effects once Gender, Position, F0, and intensity were modeled together. Gender was non-significant ($\beta = 0.087$, $SE = 0.272$, $p = 0.749$), position was also non-significant, and the Gender $\times$ Position interaction did not reach significance. Neither F0 nor intensity was significant. As in statements, the covariate-

controlled jitter model for yes/no questions provides no evidence for systematic gender- or position-based differentiation. See Tables 45–46 for details.

Table 45 Fixed effects for Jitter in Yes/No questions after controlling for F0 and intensity

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	0.924	0.154	5.996	65.755	0.000
fixed	genderM	0.087	0.272	0.322	76.462	0.749
fixed	positionmiddle	0.132	0.167	0.786	170.159	0.433
fixed	positionend	-0.025	0.176	-0.142	178.255	0.887
fixed	f0_z	-0.150	0.107	-1.399	125.711	0.164
fixed	intensity_z	-0.049	0.065	-0.754	74.886	0.453
fixed	genderM $\times$ positionmiddle	-0.378	0.237	-1.598	169.694	0.112
fixed	genderM $\times$ positionend	-0.236	0.237	-0.996	170.251	0.321

Table 46 ANOVA for fixed effects for Jitter in Yes/No questions after controlling for F0 and intensity

Term	NumDF	DenDF	F value	p value
gender	1	47.591	0.241	0.625
position	2	175.903	0.653	0.522
f0_z	1	125.711	1.957	0.164
intensity_z	1	74.886	0.568	0.453
gender $\times$ position	2	169.562	1.300	0.275

6.5.2 Shimmer (local %)

In yes/no questions, shimmer showed no significant effects of gender, position, or their interaction. However, unlike the other shimmer models, token-level intensity emerged as a significant positive predictor ($\beta = 1.312$, $SE = 0.489$, $p = 0.009$), while F0 remained non-significant. This indicates that shimmer variation in yes/no questions is more closely related to token-level intensity than to gender or utterance position. See Tables 47–48 for details.

Table 47 Fixed effects for Shimmer in Yes/No questions after controlling for F0 and intensity

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	8.583	1.555	5.518	8.032	0.001
fixed	genderM	0.168	1.796	0.094	56.801	0.926
fixed	positionmiddle	0.168	0.958	0.175	169.350	0.861
fixed	positionend	0.617	1.045	0.591	178.032	0.555
fixed	f0_z	-0.840	0.649	-1.293	166.584	0.198
fixed	intensity_z	1.312	0.489	2.682	103.649	0.009
fixed	genderM $\times$ positionmiddle	-1.940	1.355	-1.432	168.963	0.154

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	genderM $\times$ positionend	-0.713	1.358	-0.525	169.768	0.600

Table 48 ANOVA for fixed effects for Shimmer in Yes/No questions after controlling for F0 and intensity

Term	NumDF	DenDF	F value	p value
gender	1	40.879	0.186	0.668
position	2	175.541	1.226	0.296
f0_z	1	166.584	1.672	0.198
intensity_z	1	103.649	7.195	0.009
gender $\times$ position	2	168.763	1.054	0.351

6.5.3 Harmonics-to-noise ratio (HNR, dB)

In yes/no questions, HNR no longer showed a significant gender effect after controlling for F0 and intensity. The coefficient for males was small and non-significant ($\beta = -0.779$, $SE = 1.445$, $p = 0.595$), and the Type III test for gender was likewise non-significant. Position remained significant overall, with higher HNR at the end than at the beginning ($\beta = 1.290$, $SE = 0.601$, $p = 0.033$), while the middle position did not differ from the beginning. F0 was a strong positive predictor of HNR ($\beta = 2.564$, $SE = 0.382$, $p < 0.001$), whereas intensity was not significant. See Tables 49–50 for details. The unadjusted version of this model is included in Appendix H as a supplementary analysis.

Table 49 Fixed effects for HNR in yes/no questions after controlling for F0 and intensity

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	14.618	1.466	9.969	7.063	0.000
fixed	genderM	-0.779	1.445	-0.540	22.797	0.595
fixed	positionmiddle	-0.223	0.541	-0.412	168.499	0.681
fixed	positionend	1.290	0.601	2.146	172.761	0.033
fixed	f0_z	2.564	0.382	6.707	175.294	0.000
fixed	intensity_z	-0.453	0.307	-1.475	178.894	0.142
fixed	genderM $\times$ positionmiddle	0.866	0.765	1.132	168.347	0.259
fixed	genderM $\times$ positionend	-0.196	0.767	-0.255	168.663	0.799

Table 50 ANOVA for fixed effects for HNR in yes/no questions after controlling for F0 and intensity

Term	NumDF	DenDF	F value	p value
gender	1	19.917	0.158	0.695
position	2	171.004	4.208	0.016
f0_z	1	175.294	44.982	0.000
intensity_z	1	178.894	2.175	0.142
gender $\times$ position	2	168.288	1.114	0.330

Follow-up contrasts showed no significant gender differences at any position. These results indicate that in yes/no questions, HNR is shaped more clearly by utterance-final position and token-level F0 than by gender. See Tables 51–52 and Figure 31 for details.

Table 51 Simple effects of Gender within each Position (HnR in Yes/No questions, controlling for F0 and intensity)

position	contrast	Estimate	Std. Error	df	t value	p value
beginning	F - M	0.779	1.530	26.494	0.509	0.615
middle	F - M	-0.086	1.546	27.211	-0.056	0.956
end	F - M	0.975	1.580	29.468	0.618	0.542

Table 52 Pairwise comparisons of Position within each Gender (HnR in Yes/No questions, controlling for F0 and intensity)

gender	contrast	Estimate	Std. Error	df	t value	p value
F	beginning - middle	0.223	0.551	174.711	0.405	0.914
F	beginning - end	-1.290	0.615	179.165	-2.096	0.094
F	middle - end	-1.513	0.580	176.837	-2.610	0.026
M	beginning - middle	-0.643	0.552	174.220	-1.165	0.475
M	beginning - end	-1.094	0.561	175.566	-1.950	0.128
M	middle - end	-0.451	0.576	176.119	-0.783	0.714

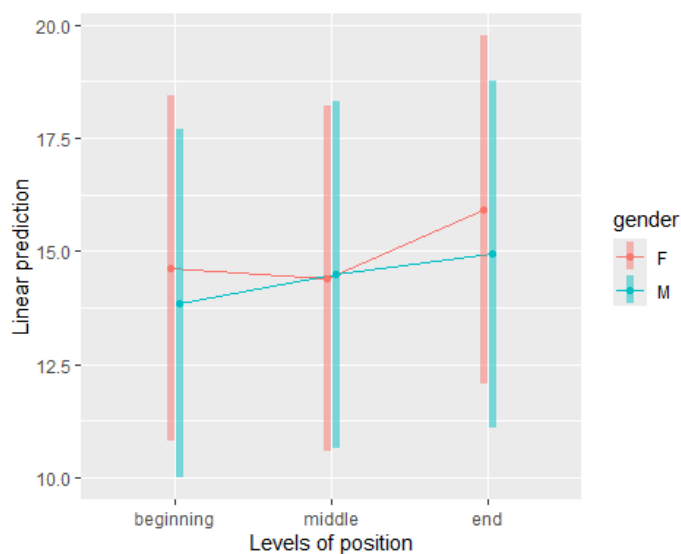


Figure 31 Estimated marginal means of HNR in yes-no questions by gender and position from the mixed-effects model including token-level F0 and intensity as covariates. Error bars indicate 95% confidence intervals.

6.6 Summary

This chapter re-examined the Chapter Five main-study corpus from the perspective of gender-related variation in jitter, shimmer, and HNR, while also incorporating token-level

F0 and intensity into the statistical models. The results show that the modality- and position-related patterns established in Chapter Five do not simply translate into a uniform male–female contrast. Instead, the reanalysis indicates that the strongest and most consistent predictor of HNR is token-level F0 rather than gender. In statements and yes/no questions, gender did not remain a reliable predictor of HNR once local acoustic covariates were included, while in *wh*-questions the gender-related pattern was weaker and less uniform across positions.

Jitter and shimmer likewise showed mainly context-dependent rather than global gender-related patterning. Jitter remained most clearly structured in *wh*-questions, where values decreased at the utterance end and the Gender $\times$ Position interaction persisted, although the direct gender contrast was weaker under the covariate-controlled models. Shimmer showed only limited evidence of position-dependent gender differentiation in statements, no reliable effect in *wh*-questions, and a closer association with token-level intensity in yes/no questions. Taken together, these findings suggest that acoustic voice measures in connected speech are highly sensitive to prosodic context and local acoustic realization.

More broadly, this chapter refines rather than replaces the main-study results reported in Chapter Five. The modality- and position-related patterns identified there remain central, but the present reanalysis shows that some effects initially suggestive of gender differentiation—especially in HNR—are better interpreted in relation to local acoustic covariation, particularly F0. This provides a more cautious basis for the broader dissertation interpretation developed in the following chapter, where the findings from Chapters Five and Six are brought together in relation to the dissertation’s main research questions.

Chapter Seven: General Discussion

7.1 Chapter overview

This chapter interprets the results reported in Chapters Five and Six, with the goal of clarifying what the observed patterns suggest about the intonation–phonation interface in Damascene Syrian Arabic. Chapter Five established the main prosodic baseline of the dissertation by showing how sentence modality and prosodic position structure intonational patterns and acoustic voice measures in the main-study corpus, while Chapter Six extended that analysis by asking whether those modality- and position-related patterns remain stable across male and female speakers once token-level F0 and intensity are taken into account. The discussion therefore proceeds by (i) summarizing the key empirical findings across both chapters in relation to the research questions, (ii) interpreting each major pattern with explicit links to the reported models and contrasts, (iii) situating those interpretations in the relevant phonetic and phonological literature, and (iv) outlining methodological constraints, broader implications, and testable directions for future work.

7.2 Summary of key findings

Four findings emerge from Chapters Five and Six. First, sentence type is associated with distinct tonal profiles. Statements predominantly display high-pitched accents (H*, !H*) and a low boundary tone (L%). *Wh*-questions predominantly display low accents (L, L*), with (L%.) Yes/no questions predominantly display high accents (H*, H) with a high boundary tone (H%). These modality-specific tonal patterns provide the prosodic frame within which the voice-quality results must be interpreted.

Second, prosodic position effects on voice-quality measures are sentence-type dependent. In statements, position does not show statistically reliable effects on jitter, shimmer, or HNR, with only weak numerical tendencies. In *wh*-questions, position is strongly associated with all three measures: compared to medial position, final position shows lower jitter, lower shimmer, and higher HNR in the baseline models, and the covariate-controlled reanalysis confirms that final position remains especially important for jitter and HNR. In yes/no questions, position effects are more limited: jitter and shimmer do not show reliable positional differences, but HNR is significantly higher in the final than in the initial position, and this effect remains in the gender-based reanalysis.

Third, the direction of significant effects is internally coherent across measures when effects occur. In *wh*-questions, the combination of lower perturbation (jitter/shimmer) and higher harmonicity (HNR) in the final position points to a single descriptive interpretation: phonation is more periodic and less noisy in the final position than in the medial position for this sentence type. The same broader tendency toward increased final-position harmonicity is also visible in yes/no questions, although there it is restricted primarily to HNR.

Fourth, the gender-based reanalysis shows that these prosodic patterns are not best interpreted as reflecting a uniform male–female contrast. Across sentence types, HNR is more consistently associated with token-level F0 than with gender. In statements and yes/no questions, gender does not remain a reliable predictor of HNR once acoustic covariates are taken into account, while in *wh*-questions, the gender-related pattern is weaker and less uniform across positions than the unadjusted models initially suggest. Jitter and shimmer likewise show mainly context-dependent rather than global gender patterning: jitter remains most clearly structured in *wh*-questions, and shimmer shows only limited position-dependent gender differentiation in statements while being more closely associated with intensity in yes/no questions. Taken together, these findings suggest that the intonation–phonation interface in Syrian Arabic connected speech is shaped not only by modality and prosodic position, but also by the interaction between local acoustic realization and speaker-related variation.

These findings jointly suggest that (i) tonal organization differs systematically by sentence modality, (ii) the phonatory dimension is not uniformly modulated by position across the grammar but becomes more detectable in particular tonal and modal configurations, especially *wh*-questions, and (iii) some patterns that might initially appear to reflect gender are better understood once token-level F0 and intensity are modeled directly. The following sections develop these points in relation to the dissertation’s research questions.

7.3 Conceptual and measurement background for interpreting the outcomes

Before interpreting the results theoretically, it is essential to be explicit about what the acoustic outcomes reflect. Jitter and shimmer quantify cycle-to-cycle irregularity in the acoustic waveform: jitter indexes short-term variability in fundamental period, while shimmer indexes short-term variability in amplitude. HNR (harmonics-to-noise ratio)

indexes the relative dominance of periodic (harmonic) energy over aperiodic (noise) energy; higher HNR typically corresponds to a cleaner, more periodic signal. These measures are widely used as correlates of overall voice quality, but their validity and stability depend on recording conditions, intensity, algorithmic settings, and whether the signal contains non-modal phonation that challenges periodicity assumptions (Maryn et al., 2009; Gerratt et al., 2016). They should therefore be treated as acoustic correlates of periodicity, stability, and noise rather than as direct categorical labels of phonation type.

A related point is methodological rather than theoretical: acoustic voice measures are not independent of the phonetic conditions under which they are measured. Perturbation measures, in particular, can be sensitive to vocal intensity, fundamental frequency, and to differences between sustained-vowel phonation and connected speech (Orlikoff & Kahane, 1991; Brockmann et al., 2008, 2011; Gerratt et al., 2016). This matters directly for the interpretation of the present results, because the main-study corpus consists of connected Syrian Arabic speech, where F0, intensity, segmental composition, and prosodic position all vary across the utterance. As a result, position effects in connected speech may reflect both genuine laryngeal adjustment and measurement-related differences associated with prosodic organization. This does not invalidate the measures, but it does require cautious interpretation, especially when contrasting null results with significant effects across sentence types.

The lack of consistent overall gender effects on jitter and shimmer supports the view that perturbation measures in connected speech are shaped more by local production conditions than by stable anatomical differences alone. Previous studies have reported mixed gender patterns in perturbation. Sørensen and Horii (1983), for example, found higher frequency and amplitude perturbation in female voices under some conditions, whereas Teixeira and Fernandes (2014) showed that jitter, shimmer, and HNR pattern differently across gender, vowel, and tone even in healthy voices. Similarly, Shabnam and Pushpavathi (2022) reported gender-related differences in multiparameter voice indices, but not a simple, uniform perturbation contrast that would generalize across all contexts. The present results are more in line with this latter perspective: gender-related differences in perturbation were limited, context-specific, and sensitive to the inclusion of acoustic covariates.

The reanalysis presented in Chapter Six further strengthens this point. Studies such as Dehqan et al. (2010) and Shabnam and Pushpavathi (2022) provided a reasonable basis for expecting higher HNR in female speakers. However, once token-level F0 and intensity are

included directly in the models, some effects that initially appear compatible with a gender-based interpretation become weaker or disappear. This is most evident for HNR: in the gender chapter, HNR is more consistently associated with token-level F0 than with gender, and the apparent gender contrast is no longer robust across sentence types once local acoustic covariates are controlled. This suggests that at least part of the variation captured by HNR in connected speech reflects local pitch-related and intensity-related conditions rather than a stable speaker-level voice-quality contrast. More broadly, the Chapter Six reanalysis confirms that these acoustic measures are best interpreted as context-sensitive correlates of speech production, not as fixed readouts of speaker identity or phonation category.

This interpretation is consistent with a broader methodological literature showing that HNR and related perturbation/noise measures are not acoustically inert. Gerratt et al. (2016) showed that sustained phonation and continuous speech do not yield interchangeable voice-quality measures, which is directly relevant here because the present study examines connected speech rather than sustained-vowel phonation. Likewise, Brockmann-Bauser et al. (2018) demonstrated that commonly used perturbation measures vary systematically with vocal intensity, and more recent modeling work has shown that widely used perturbation and noise measures are affected by both F0 and intensity across a broad parameter space (Baker et al., 2026). Taken together, these studies suggest that an apparent gender effect weakens once token-level F0 and intensity are explicitly controlled. Under this view, the earlier female>male HNR findings reported by Dehqan et al. (2010) and Shabnam and Pushpavathi (2022) are not necessarily incorrect, but they should be interpreted as task- and dataset-specific patterns rather than as universally stable properties of healthy male and female voices.

Finally, the present dissertation uses an AM/ToBI-style representational framework to describe intonation. In this view, sentence melody is modeled as sequences of discrete tonal events (pitch accents and boundary tones) aligned to prosodic structure rather than as purely gradient contours. This matters because it establishes a concrete prediction space: if phonation covaries with intonation, it may do so systematically with tonal events, with prosodic position, or with sentence modality (Silverman et al., 1992; Ladd, 2008; Jun, 2005). At the same time, the findings from Chapters Five and Six suggest that any such covariation must be interpreted through the measurement properties of the acoustic outcomes themselves. In other words, the relevant question is not simply whether

intonation and phonation co-occur, but how prosodic structure, local acoustic realization, and speaker-related variation jointly shape the measurable correlates of that relationship.

7.4 Discussion by research question

7.4.1 RQ1: How do prosodic features interact with voice quality measures in Syrian Arabic?

In this dissertation, the interaction between prosodic features and voice-quality measures is established in two linked steps. First, Chapter Five identifies sentence-type-specific tonal profiles: statements show predominantly high pitch accents (H^* , $!H^*$) with a low boundary tone ($L\%$), *wh*-questions show low accents (L , L^*) with ($L\%$), and yes/no questions show high accents (H^* , H) with a high boundary tone ($H\%$). Second, within each sentence type, mixed-effects models test whether the acoustic outcomes vary by prosodic position rather than including tonal labels as direct predictors. The present results, therefore, do not yet demonstrate a direct mapping from particular pitch-accent or boundary-tone labels to jitter, shimmer, or HNR. Instead, they show that the prosodic environments associated with different sentence types are also associated with different degrees of phonatory differentiation.

A motivated interpretation is that intonation and phonation are not independent in this dataset. Rather, the intonational environments defined by sentence modality, prosodic position, and their typical tonal realizations appear to be associated with different degrees of periodicity-related and perturbation-related modulation. This interpretation is compatible with work showing that pitch-related and phonatory dimensions can covary in production, and that non-modal phonation is often conditioned by local pitch range and prosodic context rather than by categorical tonal labels alone (Kuang, 2017; Gordon & Ladefoged, 2001; Esposito & Khan, 2020; Zhang, 2016). Studies of languages where tone and phonation are independently contrastive demonstrate that phonation and tonal implementation can still interact in their phonetic realization, implying that coupling can arise even when the phonological systems are distinct (Garellek & Keating, 2011). At the same time, the reanalysis of Chapter Six shows that some apparent gender-related effects, especially for HNR, become weaker or disappear once token-level F_0 and intensity are modeled directly. This means that at least part of the apparent intonation–phonation relation in connected speech must be interpreted through the local acoustic realization of the utterance, not solely through abstract tonal category.

This point is especially important for HNR. In the baseline Chapter Five results, HNR appears to pattern quite strongly with sentence type and position, especially in *wh*-questions and yes/no questions. However, the gender chapter shows that HNR is more consistently associated with token-level F0 than with gender across sentence types, and that the apparent gender contrast is not robust once local acoustic covariates are controlled. This does not weaken the claim that prosody and phonation are related; rather, it sharpens it. The emerging picture is that the intonation–phonation interface in this dataset is mediated by the interaction of prosodic structure and local acoustic realization, not by a simple one-to-one mapping from tonal labels to stable phonatory settings (Gerratt et al., 2016; Dehqan et al., 2010; Shabnam & Pushpavathi, 2022; Garellek, 2022).

From this perspective, the present findings are better understood as evidence that HNR in connected speech is strongly conditioned by local acoustic realization. This interpretation aligns more closely with work that treats acoustic voice measures in continuous speech as context-sensitive correlates of phonation rather than as fixed voice-quality categories (Gerratt et al., 2016; Garellek, 2022). Accordingly, the present data do not justify maintaining HNR as the most reliable gender marker in the study. Rather, they show that HNR is the measure most clearly shaped by acoustic covariation, especially by F0.

At the same time, the strength and direction of pitch–phonation covariation is not universal: languages differ in whether tone and phonation are tightly coupled, loosely covarying, or largely independent, and even within a language, the mapping can vary by prosodic context and communicative function (Gussenhoven, 2004; Ladd, 2008). This contextual dependence fits broader accounts of prosody as a multi-cue system, in which meanings such as interrogativity may be signalled through coordinated combinations of intonation, timing, and voice quality rather than through F0 alone (Hirschberg, 2004). Within Arabic specifically, sentence modality is a primary driver of intonational structure and function; this motivates treating pitch-accent and boundary-tone environments as plausible conditioning contexts for voice-quality covariation, while remaining explicit about what the present statistical models do and do not test directly (Rifaat, 2021).

Two alternatives should remain explicit. First, position effects may reflect differences in tonal implementation rather than differences in tonal category. Even when the phonological labels are the same (for example, L% in both statements and *wh*-questions), the phonetic realization of those tonal events may differ across sentence type and position through timing, scaling, or alignment differences (Ladd, 2008). Because neither Chapter Five nor

Chapter Six includes tonal labels as direct predictors or quantifies tonal implementation directly, the present analyses cannot distinguish this account from a modality-driven phonation account.

Second, position effects may reflect prosodic-phonetic correlates that influence measurement, especially in connected speech (Turk & Shattuck-Hufnagel, 2007; Paschen et al., 2022). Jitter and shimmer are sensitive, though to different degrees, to F0, intensity, segment duration, and boundary-related timing variation (Brockmann et al., 2011). Likewise, periodicity/noise indices such as HNR can be affected by factors including F0, intensity, and segment duration (Sampaio et al., 2020). The results from Chapter Six reinforce this by showing that some effects weaken when token-level F0 and intensity are introduced directly into the models. Accordingly, the present dissertation provides evidence that voice-quality differentiation varies across the prosodic environments associated with sentence type and position, but it does not yet establish a direct phonological mapping from ToBI labels to categorical phonation types.

In this dataset, the clearest evidence for an intonation–phonation association appears in the environment characterizing *wh*-questions, where final position shows lower jitter and higher HNR than medial position, and where the final-position pattern remains robust even after local acoustic covariates are modeled. By contrast, statements show little robust positional differentiation, and yes/no questions show a narrower pattern largely confined to HNR. Taken together, these results suggest that prosodically conditioned voice-quality variation is real in Syrian Arabic, but that its interpretation must remain sensitive to the interaction between tonal structure, sentence modality, and acoustic realization.

7.4.2 RQ2: Does prosodic position condition voice quality?

Prosodic position remained crucial for understanding how these measures behaved across sentence types, but its effects were not equally strong for all measures. Instead, the position voice-quality relation is modality-dependent, with the clearest and most internally consistent positional pattern occurring in *wh*-questions. Chapter Five already established this asymmetry, and Chapter Six refines it by showing that the main positional effects remain, but that some of their apparent speaker-related interpretations change once token-level F0 and intensity are modeled directly.

In statements, target-word position does not reliably condition jitter, shimmer, or HNR. The Chapter Five models showed no robust positional effects, and the Chapter Six

reanalysis confirms that this remains true once Gender, F0, and intensity are included. This suggests that, in the controlled read-speech conditions used here, declaratives maintain a relatively stable periodicity/noise profile across beginning, medial, and final target-word positions. This does not necessarily mean that phrase-edge phonatory variability is absent in declaratives. Rather, it suggests either that such variability is weak under the present task and measurement windows or that it is not reliably captured by jitter, shimmer, and HNR in this corpus (Redi & Shattuck-Hufnagel, 2001; Dallaston & Docherty, 2020; Irons & Alexander, 2016).

In *wh*-questions, by contrast, position clearly conditions voice-quality measures. Final position shows lower jitter and higher HNR than medial position, and this remains true in the Chapter Six reanalysis. This combination supports a single descriptive interpretation: phonation is more periodic and less noisy in the final position than in the medial position for *wh*-questions. Comparable modality-linked differences in utterance-final phonation have been reported elsewhere. For Mandarin, Chai (2019) reports that the utterance-final position in questions is more periodic (higher HNR) than non-final positions, contrasting with statements, where final position patterns as less periodic (lower HNR). For French, statement-final vowels are described as breathier/more often devoiced than question-final vowels, suggesting that questions can be more periodic in final position than statements, independent of the boundary-tone category (Smith, 1999). This matters for the present results because the *wh*-questions here end in L% (like the statements), yet final position still shows a shift toward greater periodicity relative to medial position; therefore, the effect points to modality-specific implementation even in low-boundary-tone contexts. Importantly, the reanalysis also shows that the apparent direct gender contrast in jitter and HNR is weaker once F0 and intensity are included, while the positional pattern itself remains strong. This is a key result, because it indicates that the *wh*-question effect is not primarily a male–female contrast; it is a prosodically conditioned positional effect that survives covariate control.

In yes/no questions, positional conditioning is more limited. Chapter Five showed a significant final-position increase in HNR but no reliable positional effects in jitter or shimmer, and Chapter Six confirms that this narrower pattern remains: HNR is still higher at the utterance end, while jitter and shimmer remain largely stable. This again argues against a broad categorical shift in phonation across all measures and instead suggests a more selective effect on periodicity/noise-related structure. The fact that the HNR effect

remains while the gender effect disappears once F0 and intensity are included further supports the view that utterance-final position in yes/no questions is associated with greater harmonicity, but not necessarily with a general shift in all acoustic correlates of phonation.

One possible explanation is that question-final material is produced with greater voicing stability, yielding lower perturbation and greater harmonicity. Another is that prosodic restructuring near the end of an interrogative contour changes the local acoustic conditions under which these measures are calculated. Both interpretations are consistent with AM-based work on Syrian Arabic intonation showing that sentence type and prosodic structure shape the realization of utterance-final material (Al Hasan & Mahanta, 2022; 2023), as well as with broader phonetic work linking phrase-final glottalization and creaky or non-modal voice to prosodic position (Redi & Shattuck-Hufnagel, 2001; Garellek, 2015; Keating et al., 2015). The present results therefore suggest that position effects in connected speech are tied not simply to linear place in the utterance, but to the interaction between sentence type, prosodic organization, and local acoustic realization.

At the same time, the absence of reliable shimmer effects in *wh*-questions and the emergence of intensity as a significant predictor in yes/no questions further underline that shimmer is closely tied to local acoustic conditions. Taken together, these findings suggest that perturbation measures in connected speech are best understood as context-sensitive correlates of speech production rather than as direct markers of stable gender-linked phonatory properties (Teixeira & Fernandes, 2014; Gerratt et al., 2016).

Thus, the answer to RQ2 is yes, but only in a modality-specific sense: prosodic position conditions voice-quality measures in Damascene Syrian Arabic, most clearly in *wh*-questions and more selectively in yes/no questions, while statements show little robust positional differentiation under the present task conditions.

7.4.3 RQ3: How does sentence modality shape voice quality metrics?

Taken together, the results show a clear asymmetry in how voice-quality measures pattern across modalities: *wh*-questions show the strongest and most internally consistent differentiation across measures, yes/no questions show a more selective pattern, and statements show the least robust position-linked differentiation. This asymmetry was already visible in Chapter Five and remains central after the Chapter Six reanalysis. What changes in the gender chapter is not the basic cross-modal ranking itself, but the

interpretation of some effects—especially HNR—which become more clearly tied to token-level acoustic covariation than to broad male–female differences.

A plausible interpretation is that sentence modality modulates how strongly phonatory cues are recruited, or at least how detectable they become, across the utterance. In this dataset, *wh*-questions show the clearest position-linked differentiation in voice-quality measures, while statements show the least. This supports a working hypothesis of modality-sensitive covariation between prosodic structure and acoustic voice-quality indices. At the same time, the Chapter Six results caution against treating these modality effects as purely structural or purely speaker-driven. In the reanalysis, the strongest and most consistent HNR predictor is token-level F0 rather than gender, and jitter/shimmer remain largely context-dependent. This means that modality effects in connected speech are best interpreted as the outcome of coordinated prosodic and acoustic implementation rather than as direct evidence for categorical phonation types linked to sentence class (Pierrehumbert & Hirschberg, 1990; Gussenhoven, 2004; Ladd, 2008; Xu, 2005).

Two mechanisms remain plausible. First, a phonological–pragmatic mechanism: interrogative meaning, particularly in *wh*-questions, may motivate cue enhancement via a combination of intonational structure and accompanying laryngeal/phonatory adjustments, without implying that any single acoustic index diagnoses a categorical phonation type. This is compatible with cue integration/enhancement perspectives in speech, where multiple properties can co-signal a communicative contrast. In their classic analysis of tune meaning, Pierrehumbert and Hirschberg (1990) give an explicit example in which the same sentence type receives different discourse interpretations depending on the tune. For the yes–no question ‘May I interrupt you’, they contrast H- H H% (treated as more ‘confirmation-like,’ with the speaker expecting permission) with L- H H% (more uncertainty about the response). Kingston and Diehl (1994) argue that cue covariation can be actively controlled because it can enhance contrasts. In their discussion of stop voicing, they explicitly propose that F0 perturbation can be an independently controlled articulation whose purpose is to enhance the voice contrast, rather than a purely automatic by-product.

Second, a phonetic-implementation mechanism: modalities may differ in how F0 targets are realized, and pitch–phonation covariation may therefore differ across modalities even when the phonological labels are not sufficient on their own to predict the acoustic outcome. Work on pitch–voice quality covariation supports the plausibility of such coupling. Kuang (2017) revisits Mandarin ‘creaky voice’ and shows (i) creak is not limited

to Tone 3, (ii) creak can accompany any low pitch target, and (iii) manipulating overall pitch range changes the likelihood of creak. The Chapter Six reanalysis strengthens the plausibility of the second mechanism, because once F0 and intensity are modeled directly, some apparent gender differences weaken while the modality-sensitive positional effects remain. In other words, the modality effect is robust, but its measurable expression is filtered through local acoustic realization.

A serious alternative remains that the *wh*-question subset differs structurally from the statement and yes/no subsets in ways that correlate with position, such as *wh*-word placement, focus distribution, or segmental environments around the target words. Prosodic edge effects, including phrase-final lengthening, may also change the local acoustic conditions under which perturbation and HNR are estimated. This alternative does not negate the modality pattern observed here, but it limits the strength of any claim that the effect is purely phonatory or purely tonal in origin. What the present dissertation can claim with confidence is narrower and more robust: sentence modality shapes the distribution of acoustic voice measures in Syrian Arabic, but it does so through the interaction of tonal structure, prosodic position, and local acoustic realization rather than through a simple categorical mapping between modality and phonation.

7.5 Methodological considerations and limitations

Jitter, shimmer, and HNR are informative acoustic indices, but they are not one-to-one diagnostics of categorical phonation types. Accordingly, the discussion is limited to what these measures support directly: relative periodicity (lower perturbation, higher HNR) versus relative noise or aperiodicity (higher perturbation, lower HNR), rather than claims about discrete breathy, modal, or creaky settings. This conservative interpretation is consistent with evidence that commonly used acoustic measures do not map uniquely onto perceived overall voice quality across contexts and tasks (Maryn et al., 2009). The inclusion of F0 and intensity in Chapter Six strengthens the analysis, but it does not change the basic point that the measures remain indirect correlates rather than categorical phonation labels.

Perturbation measures were developed largely in sustained-phonation contexts and can be less stable in continuous speech because running speech introduces intensity variation, segmental transitions, and prosodically driven timing differences. This does not invalidate their use here, but it increases the need to interpret significant effects as acoustic

covariation with prosodic position and sentence modality, not as direct evidence of laryngeal mode changes (Gerratt et al., 2016).

Prosodic position can systematically affect duration and temporal organization near boundaries, for example through phrase-final lengthening, and such temporal changes can alter the acoustic conditions under which perturbation and harmonicity are estimated. In other words, a positional effect in jitter, shimmer, or HNR may reflect boundary-adjacent timing and coordination differences in addition to phonatory adjustment. Chapter Six addresses part of this concern by including token-level F0 and intensity as covariates, which allows a more precise interpretation of effects that remain after covariate control. Even so, the current models still do not include all potentially relevant local predictors, such as segment duration or more direct spectral measures of phonation, so the results should still be interpreted as approximations to a more complex acoustic-phonetic process.

Because the *wh*-word occupies the clause-initial slot in the stimulus set, *wh*-questions were analyzed only in medial versus final position, whereas statements and yes/no questions were analyzed across initial, medial, and final positions. This limits strict comparability across sentence types and makes it harder to determine whether the *wh*-question pattern reflects a general boundary-adjacent effect or a specifically medial-to-final shift tied to *wh*-question prosodic structure. For this reason, both Chapter Five and Chapter Six analyze the data separately by sentence type rather than using a single omnibus model with a full sentence-type $\times$ position interaction. This means that cross-sentence-type comparisons should be interpreted more cautiously than within-sentence-type effects.

Although the design includes multiple speakers and repeated tokens, the segmental environments surrounding the measured vowel intervals cannot be matched perfectly across all positions and modalities. The main study reduced this variability through more tightly controlled materials and by including the target word as a random effect, but vowel quality itself was not modeled directly. In addition, the speaker sample remains relatively small and socially narrow, which limits how broadly the results can be generalized across Damascene Syrian Arabic more generally. The Chapter Six reanalysis adds an important layer by testing whether the modality- and position-related patterns remain stable across male and female speakers, but even there the interpretation remains constrained by the same lexical and corpus design. In at least one refitted model, the random-effect variance for Speaker was estimated near zero, indicating that recoverable between-speaker variance was limited for that specific outcome under the present materials and model structure. This

warns against strong claims about speaker-level variability beyond the reported fixed-effect patterns.

The Chapter Five models were intentionally structured within sentence type because *wh*-questions lack an initial-position level, and therefore directly support within-modality position statements and the modality-wise descriptive comparison reported in the results chapter. Chapter Six extends this modeling strategy by adding Gender, Gender $\times$ Position, and token-level F0 and intensity, allowing a more refined interpretation of the apparent voice-quality patterns. However, even this reanalysis does not yet provide a single unified test of sentence type $\times$ position $\times$ gender across the full dataset, nor does it include tonal labels as direct predictors. The current dissertation therefore supports robust within-modality claims and a careful cross-modal synthesis, but it does not yet resolve whether the observed effects are driven more by prosodic timing, or other local acoustic factors. That remains an important next step.

Taken together, these results provide strong evidence that intonation structure, sentence modality, prosodic position, and local acoustic realization interact in shaping jitter, shimmer, and HNR in controlled Damascene Syrian Arabic read speech. At the same time, the findings should not be treated as diagnostics of categorical phonation, nor as directly generalizable to all speaking styles or varieties without further replication under broader empirical conditions.

7.6 Broader implications

For the intonation–phonation interface, the combined results of Chapters Five and Six support a view in which acoustic voice-quality indices are not uniformly modulated by prosodic position across modalities in Damascene Syrian Arabic. Statements show little robust position-related differentiation across jitter, shimmer, or HNR. *Wh*-questions show the clearest and most internally coherent medial-to-final shift, with lower perturbation and higher harmonicity toward the utterance end. Yes/no questions show a narrower pattern, in which HNR increases toward the end of the utterance while jitter and shimmer remain comparatively stable.

This asymmetry between statements and questions is important, because it indicates that utterance-final effects are not automatic consequences of phrase-final location alone. Rather, they appear to depend on the intonational and communicative demands of the sentence type. The stronger final-position effects in questions are therefore best understood

as prosodically conditioned adjustments in voice stability and harmonicity. This interpretation is consistent with the broader claim that connected-speech voice measures reflect the interaction of segmental, prosodic, and acoustic factors rather than simple static speaker traits (Gerratt et al., 2016).

Taken together, these outcomes support the broader claim that sentence modality is a meaningful domain for investigating the intonation–phonation interface, but they also suggest that the interface is mediated by phonetic implementation and local acoustic realization rather than by a direct categorical mapping from ToBI labels to phonation types (Kuang, 2017; Xu, 2005).

The findings from Chapter Six sharpen this implication further. Once token-level F0 and intensity are incorporated into the models, some apparent gender-related effects, especially for HNR, become weaker or disappear, while the broader modality- and position-sensitive patterns remain. This suggests that the interface between intonation and phonation in connected speech cannot be understood solely in terms of abstract tonal structure or broad speaker categories. Instead, the measurable outcomes reflect the interaction of prosodic structure, local acoustic realization, and speaker-related variation. In this sense, the dissertation supports a view of prosody as a multi-cue system in which F0, timing, and voice-quality correlates work together, but not always in fixed or uniform ways across sentence types or speakers (Hirschberg, 2004; Gussenhoven, 2004; Gerratt et al., 2016).

For AM/ToBI analyses of Arabic prosody, the tonal distributions reported in Chapter Five provide a clear modality asymmetry under the current annotation conventions: statements show a uniform (L%) boundary tone with high pitch accents (H*, !H*), *wh*-questions show (L%) with low pitch accents (L, L*), and yes/no questions show a uniform (H%) boundary tone with high accents (H*, H). This supports the usefulness of AM-style representations for capturing modality-specific tonal structure in Arabic and aligns with work emphasizing the value of AM/ToBI approaches for Arabic intonation while also acknowledging dialectal and pragmatic complexity (Rifaat, 2021; Al Hasan & Mahanta, 2022). At the same time, the dissertation shows that tonal structure alone does not predict uniform voice-quality patterning: the strongest position-related differentiation in acoustic voice measures emerges in *wh*-questions, even though both statements and *wh*-questions share (L%) in the present dataset.

For methodology in Arabic prosody research, the dissertation highlights the importance of treating sentence modality as a central organizing factor when testing prosody–phonation relations. If the data were collapsed across modalities, the robust *wh*-question pattern and the more selective yes/no-question pattern would be harder to detect and interpret. More broadly, the dissertation illustrates a productive workflow for Arabic prosody research: transparent AM/ToBI annotation, explicit separation of modalities when the design differs structurally, and mixed-effects modeling that takes repeated speakers and lexical items into account. The gender chapter adds an important extension to this workflow by showing that local acoustic covariates such as F0 and intensity should be modeled directly when interpreting apparent gender effects or periodicity-related differences in connected speech. This is especially important for HNR, which in the present dataset proves more consistently sensitive to token-level F0 than to gender alone.

Finally, the dissertation has broader implications for how voice measures are interpreted in connected speech. The present results do not support treating jitter, shimmer, or HNR as simple, invariant readouts of speaker-specific phonatory settings. Rather, they show that these outcomes are shaped by the interaction of prosodic structure, utterance position, sentence modality, and local acoustic conditions. This treats connected-speech voice measures as context-sensitive correlates of speech production rather than as direct labels of categorical phonation. In that respect, the dissertation contributes not only to Syrian Arabic prosody, but also to the broader methodological discussion of how intonation–phonation relations should be modeled and interpreted in running speech.

7.7 Future research directions

Several next steps follow from the present findings. First, although Chapter Six strengthens the analysis by incorporating token-level F0 and intensity directly into the mixed-effects models, the interpretation of the position effects would be further sharpened by modeling additional local phonetic properties, especially segment duration and continuous F0 characteristics such as level, slope, and range within the target region. This would help determine whether the remaining position effects in HNR and perturbation measures persist once a broader set of measurement-relevant acoustic conditions is taken into account (Baker et al., 2023; Carvalho Sampaio et al., 2020; Maryn et al., 2009).

Second, a more direct evaluation of the intonation–phonation interface would become possible if pitch-accent type, boundary tone, and other tonal properties were included as

predictors wherever tonal variability permits. The current dissertation shows that modality and prosodic position matter, but it does not yet test directly whether particular tonal categories or specific aspects of tonal implementation predict variation in periodicity- and noise-related outcomes. Future models of this kind would move the discussion from modality-based inference to more explicit tests of how intonational structure conditions acoustic voice measures.

Third, interpretation would be strengthened by adding measures that capture other dimensions of voice quality, especially cepstral and spectral indices. Such measures would help determine whether effects that are most visible in HNR reflect a broader change in periodicity and noise structure or a narrower pathway tied to the behavior of specific acoustic measures. This would also allow a more robust comparison between perturbation-based outcomes and measures more commonly used in connected-speech voice evaluation (Maryn et al., 2009).

Fourth, the extent to which the modality-sensitive patterns observed here generalize beyond the current task should be tested using comparable materials elicited in semi-spontaneous, dialogic, or other less constrained speaking conditions. This is especially important for interrogatives, since the present results suggest that *wh*-questions and yes/no questions pattern differently from statements in ways that may depend not only on tonal structure but also on communicative demands and speaking style. Extending the analysis to less scripted speech would therefore help determine whether the asymmetries reported in this dissertation are stable properties of Damascene Syrian Arabic prosody or are partly shaped by the controlled read-speech design.

7.8 Conclusion

This dissertation examined the interaction between intonation and phonation in Damascene Syrian Arabic using an AM/ToBI framework for tonal annotation and acoustic voice-quality measures extracted from target-word vowel intervals in a controlled read-speech design. The core analysis focused on how sentence modality and prosodic position condition jitter, shimmer, and harmonics-to-noise ratio (HNR). At the same time, a later reanalysis of the same corpus examined whether those patterns remain stable across male and female speakers once token-level F0 and intensity are taken into account. In this way, the dissertation approached the intonation–phonation interface both from the perspective of prosodic structure and from the perspective of speaker-related and local acoustic variation.

With respect to prosodic structure, the results show a clear modality asymmetry in the tonal inventories observed under the current annotation conventions. Statements patterned with a uniform low boundary tone (L%) and high pitch accents (H*, !H*). *Wh*-questions also showed a uniform low boundary tone (L%) but were characterized by low pitch-accent labels (L, L*). Yes/no questions patterned with a uniform high boundary tone (H%) and high pitch accents (H*, H). These distributions establish that sentence modality is robustly reflected in the intonational patterns of the dataset and provide a principled basis for investigating whether and how acoustic voice measures covary with prosodic structure.

Turning to the acoustic outcomes, the central empirical generalization is that position effects in voice-quality measures are modality-sensitive rather than uniform. In statements, the mixed-effects analyses indicated little reliable positional differentiation. In *wh*-questions, the measures showed the clearest and most internally coherent medial-to-final shift, with lower perturbation and higher harmonicity toward the utterance end. Yes/no questions showed a narrower pattern, in which HNR increased toward the end of the utterance while jitter and shimmer remained comparatively stable. Taken together, these results show that prosodic position does not exert a single global effect across the grammar; rather, the distribution of acoustic voice measures differs by sentence modality, with the strongest and most consistent positional differentiation emerging in *wh*-questions and a more selective HNR-only pattern in yes/no questions.

The gender-based reanalysis further refines this picture. Across sentence types, jitter and shimmer remained mainly context-dependent rather than uniformly gender-related, and the strongest and most consistent predictor of HNR proved to be token-level F0 rather than gender. In statements and yes/no questions, gender did not remain a reliable predictor of HNR once F0 and intensity were included, while in *wh*-questions any gender-related HNR difference became weaker and less uniform across positions. These results suggest that some effects that might initially appear to reflect a stable male–female contrast are better understood in relation to local acoustic realization, especially pitch-related variation at the token level.

Taken together, the dissertation supports a cautious theoretical conclusion about the intonation–phonation interface in this dataset. The findings do not support a direct mapping from specific ToBI labels to categorical phonation types, nor do they support a uniformly stable gender-based contrast in voice quality across contexts. More plausibly, the results are consistent with an interface in which acoustic voice measures covary with prosodic

structure through phonetic implementation, including sentence-type-specific tonal realization, prosodic position, and local acoustic conditions such as F0 and intensity. Within this interpretation, modality remains a meaningful domain for investigating prosody–phonation covariation, but the measurable outcomes are best understood as context-sensitive correlates of connected speech rather than as deterministic phonation labels or static speaker traits. This conclusion is compatible with approaches that treat intonation as communicatively grounded and phonetically implemented, while allowing that the same tonal category can be realized with different phonetic detail across contexts (Ladd, 2008; Xu, 2005).

Finally, the dissertation has treated jitter, shimmer, and HNR as acoustic indices that can reflect periodicity and noise differences, while avoiding claims that they uniquely diagnose discrete phonation categories. This conservative approach is appropriate given evidence that relationships between common acoustic measures and perceived overall voice quality are complex and context-sensitive, and that voice quality is multidimensional rather than reducible to a single metric (Maryn et al., 2009). Within these bounds, the dissertation establishes a clear empirical baseline: in this Damascene Syrian Arabic dataset, sentence modality, prosodic position, and local acoustic realization jointly condition systematic variation in acoustic voice measures, with the most coherent positional differentiation occurring in *wh*-questions and the clearest covariate-sensitive reinterpretation emerging for HNR.

References

- Abdel-Jawad, H. R. (2013). Sociolinguistic variation in Arabic: A new approach. *Dialectologia*, (Special Issue), 95–122.
- Abdelli-Beruh, N. B., Wolk, L., & Slavin, D. (2014). Prevalence of vocal fry in young adult male American English speakers. *Journal of Voice*, 28(2), 185–190. <https://doi.org/10.1016/j.jvoice.2013.08.011>.
- Abramson, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, 21, 319–325.
- Aguilar, L., Roseano, P., Vanrell, M. M., de la Mota, C., & Prieto, P. (2024). *Sp_ToBI Training Materials*. <https://sp-tobi.upf.edu> (Retrieved May 5, 2025).
- Alfozan, A. I. (1989). *Assimilation in classical Arabic: A phonological study* (Doctoral dissertation, University of Glasgow). University of Glasgow.
- Al Hasan, M., & Mahanta, S. (2023). The Effect of Post-tonic Lengthening and F0 in Syrian Arabic Prosody. TAI 2023.
- Al Hasan, M., & Mahanta, S. (2022). The Intonational Phonology of Syrian Arabic: A Preliminary Analysis. *Speech Prosody* 2022.
- Al-Tamimi, J., & Khattab, G. (2018). Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops. *Journal of Phonetics*, 71, 306–325. <https://doi.org/10.1016/j.wocn.2018.09.010>
- Aoun, J. E., Benmamoun, E., & Choueiri, L. (2010). *The Syntax of Arabic*. Cambridge University Press.
- Arvaniti, A. (2015). *The autosegmental–metrical model of intonational phonology* (submitted version of a book chapter). University of Kent.
- Arvaniti, A., Ladd, D. R., & Mennen, I. (2000). *What is a starred tone? Evidence from Greek*. In M. Broe & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the lexicon* (pp. 119–131). Cambridge University Press. (Landing page) <https://www.sciencedirect.com/science/article/pii/S1566584400000064>.

- Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of “segmental anchoring” of F0: Evidence from German. *Journal of Phonetics*, 32(2), 177–197. [https://doi.org/10.1016/S0095-4470\(03\)00039-1](https://doi.org/10.1016/S0095-4470(03)00039-1)
- Awan, S. N., Bahr, R., Watts, S., Boyer, M., Budinsky, R., Bridge2AI Voice Consortium, & Bensoussan, Y. (2024). Validity of acoustic measures obtained using various recording methods including smartphones with and without headset microphones. *Journal of Speech, Language, and Hearing Research*, 67(6), 1712–1730. https://doi.org/10.1044/2024_JSLHR-23-00759.
- Awan, S. N., Bensoussan, Y., Watts, S., Boyer, M., Budinsky, R., & Bahr, R. H. (2025). Influence of recording instrumentation on measurements of voice in sentence contexts: Use of smartphones and tablets. *Frontiers in Digital Health*, 7, 1610772. <https://doi.org/10.3389/fdgth.2025.1610772>.
- Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice (2nd ed.)*. Singular/Thomson Delmar Learning.
- Baker, C. P., Brockmann-Bauser, M., Purdy, S. C., & Rakena, T. O. (2023). High and wide: An in silico investigation of frequency, intensity, and vibrato effects on widely applied acoustic voice perturbation and noise measures. *Journal of Voice*. Advance online publication. doi:10.1016/j.jvoice.2023.10.007.
- BAS (Ludwig-Maximilians-Universität München). (2025). SpeechRecorder software.
- Basbøll, H. (2005). *The Phonology of Danish*. Oxford University Press.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255–309. https://www.phon.ox.ac.uk/jpierrehumbert/publications/intonation_japanese_english.pdf.
- Beckman, M. E., & Ayers Elam, G. (1997). *Guidelines for ToBI labelling (Version 3.0)*. Ohio State University. https://www.ling.ohio-state.edu/research/phonetics/E_ToBI/.

- Bettega, S., & Morano, R. (Eds.). (2022). *The classification of Arabic dialects: Traditional approaches, new proposals, and methodological problems*. MDPI.
<https://doi.org/10.3390/books978-3-0365-6140-0>.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, 17, 97–110. University of Amsterdam.
https://www.fon.hum.uva.nl/paul/papers/Proceedings_1993.pdf.
- Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer [Computer software]*. University of Amsterdam.
- Borrie, S. A., & Delfino, C. R. (2017). Conversational entrainment of vocal fry in young adult female American English speakers. *Journal of Voice*, 31(4), 513.e25–513.e32.
<https://doi.org/10.1016/j.jvoice.2016.12.005>.
- Boudlal, A. (2004). *The stress system of Casablanca Moroccan Arabic (ROA-650)*. Rutgers Optimality Archive. <https://roa.rutgers.edu/files/650-0404/650-BOUDLAL-3-0.PDF>.
- Brockmann, M., Drinnan, M. J., Storck, C., & Carding, P. N. (2011). Reliable jitter and shimmer measurements in voice clinics: The relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task. *Journal of Voice*, 25(1), 44–53. <https://doi.org/10.1016/j.jvoice.2009.07.002>.
- Brockmann, M., Drinnan, M. J., Storck, C., & Carding, P. N. (2008). Voice loudness and gender effects on jitter and shimmer in healthy adults. *Journal of Speech, Language, and Hearing Research*, 51(5), 1152–1160.
<https://pubmed.ncbi.nlm.nih.gov/18664710/>.
- Bruce, G. (1977). *Swedish Word Accents in Sentence Perspective*. Lund University.
- Bruggeman, A., Roettger, T. B., & Grice, M. (2021). *Acoustic correlates of lexical stress in Moroccan Arabic*. (Author accepted manuscript).
<https://eprints.whiterose.ac.uk/id/eprint/167492/>.
- Bruggeman, A., Roettger, T. B., & Grice, M. (2021). *Acoustic correlates of lexical stress in Moroccan Arabic*. *Journal of the International Phonetic Association*.
<https://www.cambridge.org/core/journals/journal-of-the-international-phonetic->

association/article/acoustic-correlates-of-lexical-stress-in-moroccan-arabic/AAFCB5BF15932B86A97A5A38090A14B8.

- Carvalho Sampaio, M., Vaz Masson, M. L., de Paula Soares, M. F., Bohlender, J. E., & Brockmann-Bauser, M. (2020). Effects of fundamental frequency, vocal intensity, sample duration, and vowel context in cepstral and spectral measures of dysphonic voices. *Journal of Speech, Language, and Hearing Research*, 63(5), 1326–1339. doi:10.1044/2020_JSLHR-19-00049.
- Catford, J. C. (1977). *Fundamental problems in phonetics*. Edinburgh University Press.
- Chahal, D. (1999). *A preliminary analysis of Lebanese Arabic intonation*. Proceedings of the Australian Linguistic Society 1999. <https://www.als.asn.au/proceedings/als1999/chahal.pdf>.
- Chahal, D. (2001). Modeling the intonation of Lebanese Arabic using the autosegmental-metrical framework: A comparison with English (Doctoral dissertation, The University of Melbourne). (If you cite the dissertation directly, include your access source.).
- Chahal, D. (2003). *Acoustic correlates of three levels of prominence in Lebanese Arabic*. In Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003). https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/p15_2067.html.
- Chahal, D., & Hellmuth, S. (2014). *The intonation of Lebanese and Egyptian Arabic*. In S.-A. Jun (Ed.), *Prosodic Typology II: The phonology of intonation and phrasing* (pp. 365–404). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199567300.003.0013>.
- Chávez-Peón, M. E. (2010). *The interaction of metrical structure, tone and phonation types: Prosodic patterns in Quiavini Zapotec* (Doctoral dissertation, University of British Columbia). <https://open.library.ubc.ca/collections/ubctheses/24/items/1.0071253>.
- Choi, S. H., Lee, J., Sprecher, A. J., & Jiang, J. J. (2012). The effect of segment selection on acoustic analysis. *Journal of Voice*, 26(1), 1–7. <https://doi.org/10.1016/j.jvoice.2010.10.009>

- Chomsky, N. (1981). *Lectures on Government and Binding: The Pisa Lectures*. Foris.
- Corretge, Ramon (2012-2025). Praat Vocal Toolkit. <https://www.praatvocaltoolkit.com>
- Cowell, M. W. (1964). *A Reference Grammar of Syrian Arabic (Based on the Dialect of Damascus)*. Georgetown University Press.
- Cruttenden, A. (1997). *Intonation (2nd ed.)*. Cambridge University Press. (Publisher page) <https://www.cambridge.org/core/books/intonation/65C966822EB0F424699A2E2324FF3419>.
- Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge University Press. (Archive record) <https://archive.org/details/prosodicssystems0000crys>.
- Dallaston, K., & Docherty, G. J. (2020). The quantitative prevalence of creaky voice (vocal fry) in varieties of English: A systematic review. *PLOS ONE*, 15(3), e0229960. <https://pmc.ncbi.nlm.nih.gov/articles/PMC7065773/>.
- Davidson, L. (2021). The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses. (PDF).
- De Jong, K., & Zawaydeh, B. A. (1999). Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics*, 27(1), 3–22. <https://doi.org/10.1006/jpho.1998.0088>.
- De Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research*, 36(2), 254–266.
- Dell, F., & Elmedlaoui, M. (2002). *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Kluwer Academic Publishers. <https://link.springer.com/book/10.1007/978-94-015-9950-6>.
- Draxler, C., & Jänsch, K. (2004). SpeechRecorder – a universal platform independent multi-channel audio recording software. In Proceedings of LREC 2004.
- El Zarka, D. (2017). *Intonation in Arabic*. In The Oxford Handbook of Arabic Linguistics. Oxford University Press.
- Epstein, M. A. (2003). *Voice quality and prosody in English*. In Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003) (pp. 2405–2408).

- https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_2405.pdf.
- Esposito, C. M., & Khan, S. u. D. (2020). The cross-linguistic patterns of phonation types. *Language and Linguistics Compass*, 14, e12392. <https://doi.org/10.1111/lnc3.12392>.
- Ferguson, C. A. (1959). Diglossia. *Word*, 15(2), 325–340. <https://doi.org/10.1080/00437956.1959.11659702>.
- Fernandes, J. F. T., Freitas, D., Junior, A. C., & Teixeira, J. P. (2023). Determination of harmonic parameters in pathological voices—Efficient algorithm. *Applied Sciences*, 13(4), 2333. <https://doi.org/10.3390/app13042333>.
- Ferrand, C. T. (2002). Harmonics-to-noise ratio: An index of vocal aging. *Journal of Voice*.
- Fletcher, J., Grabe, E., & Warren, P. (2005). *Intonational variation in four dialects of English: The high rising tune*. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 390–409). Oxford University Press. <https://scispace.com/pdf/intonational-variation-in-four-dialects-of-english-the-high-4vkbnbps7b.pdf>.
- Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, 101(6), 3728–3740. <https://pubmed.ncbi.nlm.nih.gov/9193060/>.
- Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association*, 41(2), 185–205. <https://doi:10.1017/S0025100311000193>.
- Garellek, M., Keating, P., & Esposito, C. M. (2013). Voice quality and tone identification in White Hmong. *The Journal of the Acoustical Society of America*, 133(2), 1078–1089. https://pages.ucsd.edu/~mgarellek/files/Garellek_etal_2013_JASA.pdf.
- Garellek, M. (2014). Voice quality strengthening and boundary tones in English. *Journal of Phonetics*, 45, 75–91. <https://doi.org/10.1016/j.wocn.2014.03.003>.
- Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics*, 29(4), 383–406. <https://www.sciencedirect.com/science/article/pii/S0095447001901470>.

- Grabe, E. (2004). *Intonational variation in urban dialects of English spoken in the British Isles*. In P. Gilles & J. Peters (Eds.), *Regional variation in intonation* (pp. 9–31). Niemeyer. https://www.phon.ox.ac.uk/files/people/grabe/Grabe_Niemeyer.pdf.
- Grabe, E., Kochanski, G., & Coleman, J. (2005). The intonation of native accent varieties in the British Isles: Potential for miscommunication? In K. Dziubalska-Kołaczyk & J. Przedlacka (Eds.), *English Pronunciation Models: A Changing Scene* (pp. 311–337). Bern: Peter Lang.
- Grabe, E., & Post, B. (2002). *Intonational variation in the British Isles*. In *Proceedings of Speech Prosody 2002* (pp. 343–346). ISCA. https://www.isca-archive.org/speechprosody_2002/grabe02b_speechprosody.pdf.
- Grácz, T. E., & Markó, A. (2018). *Word-initial glottal marking in Hungarian as a function of articulation rate and word class*. In M. Gósy & T. E. Grácz (Eds.), *Challenges in analysis and processing of spontaneous speech* (pp. 75–98). Research Institute for Linguistics, Hungarian Academy of Sciences. <https://doi.org/10.18135/CAPSS.75>.
- Grice, M., Baumann, S., & Benz Müller, R. (2005). *German intonation in Autosegmental-Metrical phonology*. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford University Press. (PDF) https://ifl.phil-fak.uni-koeln.de/sites/linguistik/Phonetik/pdf-publications/2005/Grice_Baumann_Benzmueller-German_Intonation_in_Autosegmental-metrical_Phonology_2005.pdf.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press. <https://doi:10.1017/CBO9780511616983>.
- Gussenhoven, C. (2005). *Transcription of Dutch intonation*. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford University Press. (Record) https://www.researchgate.net/publication/229078278_Transcription_of_Dutch_Intonation.
- Habib, R. (2010). Rural migration and language variation in Hims, Syria. *Finnish Journal of Arabic Studies*, 12, 1–23.

- Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America*, 125(1), 425–441. <https://doi.org/10.1121/1.3021306>
- Hellmuth, S. (2006). Intonational Pitch Accent Distribution in Egyptian Arabic. PhD dissertation, SOAS.
- Hellmuth, S. (2014). *Investigating variation in Arabic intonation: The case for a multi-level corpus approach*. In *Perspectives on Arabic Linguistics* (pp. 63–89). https://eprints.whiterose.ac.uk/id/eprint/102522/1/Hellmuth_2014_PAL25_multilevel_corpus_postprint.pdf.
- Henton, C. G., & Bladon, R. A. W. (1988). Creak as a sociophonetic marker. In L. M. Hyman & C. N. Li (Eds.), *Language, speech, and mind: Studies in honour of Victoria A. Fromkin* (pp. 3–29). London: Routledge.
- Hirschberg, J. (2004). Pragmatics and intonation. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 515–537). Blackwell Publishing.
- Hirst, D. (1998). Intonation in British English. In D. Hirst & A. Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 56–77). Cambridge: Cambridge University Press.
- Hollien, H. (1987). “Old voices”: What do we really know about them? *Journal of Communication Disorders*, 20(4), 301–317.
- Huffman, M. K. (1987). Measures of phonation type in Hmong. *The Journal of the Acoustical Society of America*, 81(2), 495–504. <https://doi.org/10.1121/1.394951>.
- Irons, S. T., & Alexander, J. E. (2016). Vocal fry in realistic speech: Acoustic characteristics and perceptions of vocal fry in spontaneously produced and read speech. *The Journal of the Acoustical Society of America*, 140(4), 3397. doi:10.1121/1.4970891.
- Jalal Al-Tamimi. (2022). *JalalAl-Tamimi/Praat-VQ-Measurements: Praat VQ measurements (Version v2)*. Zenodo. <https://doi.org/10.5281/zenodo.7270191>.
- Jun, S.-A. (Ed.). (2005). *Prosodic typology: The phonology of intonation and phrasing*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199249633.001.0001>.

- Jun, S.-A. (2014). *Prosodic typology: By prominence type, word prosody, and macro-rhythm*. In S.-A. Jun (Ed.), *Prosodic Typology II*. Oxford University Press.
- Kager, R. (2009). *Arabic stress*. In *Encyclopedia of Arabic Language and Linguistics*. (PDF) https://research-portal.uu.nl/ws/files/127310010/Kager_2009_Encyclopedia_of_Arabic_Language_and_Linguistics.pdf.
- Keating, P., Garellek, M., & Kreiman, J. (2015). *Acoustic properties of different kinds of creaky voice*. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*. https://idiom.ucsd.edu/~mgarellek/files/Keating_etal_2015_ICPhS.pdf.
- Kelly, N. E. (2023). Phrase-final intonation adjustment in Lebanese Arabic. *Journal of the International Phonetic Association*.
- Khan, S. u. D. (2012). The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics*, 40(6), 780–795. <https://doi.org/10.1016/j.wocn.2012.07.001>.
- Khattab, G. (2007). A phonetic study of gemination in Lebanese Arabic. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 153–156). Saarbrücken: Universität des Saarlandes.
- Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, 70(3), 419–454. <https://doi:10.1353/lan.1994.0023>.
- Kiparsky, P. (2003). *Syllables and moras in Arabic*. In C. Féry & R. van de Vijver (Eds.), *The Syllable in Optimality Theory* (pp. 147–182). Cambridge: CUP.
- Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, 87(2), 820–857.
- Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America*, 142(3), 1693–1706. <https://doi:10.1121/1.5003649>.
- Kuang, J., & Liberman, M. (2018). Integrating voice quality cues in the pitch perception of speech and non-speech utterances. *Frontiers in Psychology*, 9, 2147. <https://doi.org/10.3389/fpsyg.2018.02147>.

- Kuang, J., & Keating, P. (2014). Vocal fold vibratory patterns in tense versus lax phonation contrasts. *The Journal of the Acoustical Society of America*.
- Kulk, F., Odé, C., & Woidich, M. (2003). The intonation of colloquial Damascene Arabic: a pilot study. *IFA Proceedings* 25, 15–20.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.
<https://doi.org/10.18637/jss.v082.i13>
- Ladd, D. R. (2004/2006). Segmental anchoring of pitch movements: Autosegmental association or gestural coordination? *Rivista di Linguistica / Italian Journal of Linguistics*, 18(1).
- Ladd, D. R. (2008). *Intonational phonology (2nd ed.)*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511808814>.
- Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. University of Chicago Press.
- Laver, J. (1980). *The phonetic description of voice quality*. Cambridge University Press.
- Lenth, R. V., & Piaskowski, J. (2025). *emmeans: Estimated marginal means, aka least-squares means (R package version 2.0.3)*. <https://rvlenth.github.io/emmeans/>
- Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., & Corthals, P. (2009). Acoustic measurement of overall voice quality: A meta-analysis. *The Journal of the Acoustical Society of America*, 126(5), 2619–2634. <https://doi:10.1121/1.3224706>.
- Mayo, C., Aylett, M., & Ladd, D. R. (1997). Prosodic transcription of Glasgow English: An evaluation study of GlaToBI. In *Proceedings of the ESCA Workshop on Intonation: Theory, Models and Applications* (pp. 231–234). Athens.
- McCarthy, J. J. (1979). On stress and syllabification. *Linguistic Inquiry* 10(3), 443–465.
- McCarthy, J. J. (1980). A note on the accentuation of Damascene Arabic. *Studies in the Linguistic Sciences*, 10, 77–98.
- Mertens, P. (2004). The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In *Proceedings of Speech Prosody 2004* (pp. 549–552). Nara, Japan.

- Müller, K., & Wickham, H. (2026). *tibble: Simple Data Frames* (R package version 3.3.1).
<https://tibble.tidyverse.org/>
- Munro, P., & Lopez, F. H. (1999). *Di'csyonaary X:tèe'n Dii'zh Sah Sann Lu'uc (San Lucas Quiavini Zapotec Dictionary)*. Chicano Studies Research Center, UCLA.
<https://wals.info/refdb/record/Munro-and-Lopez-1999>.
- Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.
- Nespor, M., & Vogel, I. (1986/2007). Prosodic phonology (1st ed. 1986; updated reprint 2007). Foris/Mouton. (Bibliographic entry)
<https://www.cambridge.org/core/journals/phonology/article/marina-nespor-irene-vogel-1986-prosodic-phonology/0E6E7AE47CD6617500777DD8C9F7B0BA>
- Nishio, M., & Niimi, S. (2008). Changes in speaking fundamental frequency characteristics with aging. *Folia Phoniatrica et Logopaedica*, 60(3), 120–127.
<https://doi.org/10.1159/000118510>.
- O'Connor, J. D., & Arnold, G. F. (1973). *Intonation of colloquial English: A practical handbook (2nd ed.)*. Longman.
https://books.google.com/books/about/Intonation_of_Colloquial_English.html?id=MjcmAQAAMAAJ.
- Paschen, L., Fuchs, S., & Seifart, F. (2022). Final lengthening and vowel length in 25 languages. *Journal of Phonetics*, 94, 101179.
<https://doi.org/10.1016/j.wocn.2022.101179>.
- Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology*, 27(3), 887–905.
https://doi.org/10.1044/2018_AJSLP-17-0009.
- Peters, B., Kohler, K. J., & Wesener, T. (2005). Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In K. J. Kohler, F. Kleber, & B. Peters (Eds.), *Prosodic Structures in German Spontaneous Speech (AIPUK 35a)* (pp. 143–184). Kiel: IPDS.

- Picanço, G. (2005). *Mundurukú: Phonetics, phonology, synchrony, diachrony* (Doctoral dissertation, University of British Columbia).
<https://open.library.ubc.ca/collections/ubctheses/831/items/1.0092097>.
- Pierrehumbert, J., & Hirschberg, J. (1990). *The meaning of intonational contours*. In *Intentions in Communication*. MIT Press.
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (PhD thesis). MIT. <https://dspace.mit.edu/handle/1721.1/16065>.
- Pulleyblank, D. (1986). *Tone in Yoruba*. Reidel.
- R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, 29(4), 407–429.
<https://doi.org/10.1006/jpho.2001.0145>.
- Rifaat, K. (2021). The intonation of Arabic: Review and hypotheses. In K. C. Ryding & D. Wilmsen (Eds.), *The Cambridge handbook of Arabic linguistics* (pp. 330–352). Cambridge University Press. <https://doi.org/10.1017/9781108277327.015>.
- Rizzi, L. (1982). *Issues in Italian Syntax*. Foris.
- Rosenberg, A. (2010). AuToBI: A tool for automatic ToBI annotation. In *Proceedings of Interspeech 2010* (pp. 146–149). Makuhari, Japan.
- Ryding, K. C. (2005). *A Reference Grammar of Modern Standard Arabic*. Cambridge University Press.
- SAGE Research Methods. (n.d.). Within-subjects design. SAGE.
- Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. MIT Press. <https://mitpress.mit.edu/9780262690980/phonology-and-syntax/>.
- Selkirk, E. (2011). *The syntax–phonology interface*. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The handbook of phonological theory* (2nd ed.).
<https://people.umass.edu/selkirk/pdf/Selkirk%202011%20TSPI.pdf>.
- Shetewi, O. (2023). Accommodation patterns in the speech of Arabic-speaking immigrants in Spain. *Languages*, 8(4), 236. <https://doi.org/10.3390/languages8040236>.

- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). *ToBI: A standard for labeling English prosody*. In Proceedings of the 1992 International Conference on Spoken Language Processing (ICSLP). (Reference record)
https://www.researchgate.net/publication/221492301_ToBI_A_standard_for_labeling_English_prosody.
- Steffman, J., Shattuck-Hufnagel, S., & Cole, J. (2022). The rise and fall of American English pitch accents: Evidence from an imitation study of rising nuclear tunes. In *Proceedings of Speech Prosody 2022* (pp. 857–861). Baixas: International Speech Communication Association. <https://doi.org/10.21437/SpeechProsody.2022-174>
- Sulaiman, M. (2016). The syntax of wh-questions in Syrian Arabic (Doctoral dissertation, Newcastle University).
- Sullivan, J. (2012). The why of Belfast rises. In B. Migge & M. Ní Chiosáin (Eds.), *New Perspectives on Irish English* (pp. 67–84). Amsterdam: John Benjamins.
<https://doi.org/10.1075/veaw.g44.04sul>
- Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis—Jitter, shimmer and HNR parameters. *Procedia Technology*, 9, 1112–1122.
<https://www.sciencedirect.com/science/article/pii/S2212017313002788>.
- Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry*, 30(2), 219–255. https://www.leibniz-zas.de/fileadmin/Archiv2019/mitarbeiter/truckenbrodt/1999_HT_LI.pdf
- Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35(4), 445–472.
<https://doi:10.1016/j.wocn.2006.12.001>.
- Van Teijlingen, E., & Hundley, V. (2001). *The importance of pilot studies*. Social Research Update, 35. <https://abdn.elsevierpure.com/en/publications/the-importance-of-pilot-studies-2/>.
- Watson, J. C. E. (2011). *Arabic dialects (general article)*. In S. Weninger, G. Khan, M. Streck, & J. C. E. Watson (Eds.), *The Semitic Languages: An International Handbook* (pp. 851–896). De Gruyter Mouton.

- Watson, J. C. E. (2002). *The Phonology and Morphology of Arabic*. Oxford University Press.
- Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (2012). How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology*, 3, Article 580. <https://doi.org/10.3389/fpsyg.2012.00580>.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer. <https://ggplot2.tidyverse.org/>
- Wickham, H., & Bryan, J. (2025). *readxl: Read Excel Files* (R package version 1.4.5). <https://readxl.tidyverse.org/>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2026a). *dplyr: A Grammar of Data Manipulation* (R package version 1.2.0). <https://dplyr.tidyverse.org/>
- Wickham, H., Vaughan, D., & Girlich, M. (2026b). *tidyr: Tidy Messy Data* (R package version 1.3.2). <https://tidyr.tidyverse.org/>
- Wolk, L., Abdelli-Beruh, N. B., & Slavin, D. (2012). Habitual use of vocal fry in young adult female speakers. *Journal of Voice*, 26(3), e111–e116. <https://pubmed.ncbi.nlm.nih.gov/21917418/> PubMed.
- Xie, Y. (2025). *knitr: A General-Purpose Package for Dynamic Report Generation in R* (R package version 1.50.4). <https://yihui.org/knitr/>
- Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46(3–4), 220–251. <https://doi.org/10.1016/j.specom.2005.02.014>.
- Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics*, 38(3), 329–336.
- Yip, M. (2002). *Tone*. Cambridge University Press.
- Yuasa, I. P. (2010). *Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women?* *American Speech*, 85(3), 315–337. <https://read.dukeupress.edu/american-speech/article/85/3/315/5885/Creaky-Voice-A-New-Feminine-Voice-Quality-for>.

- Yumoto, E., Gould, W. J., & Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *The Journal of the Acoustical Society of America*, 71(6), 1544–1550. <https://pubmed.ncbi.nlm.nih.gov/7108029/>.
- 't Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge University Press. (Publisher page) <https://www.cambridge.org/core/books/perceptual-study-of-intonation/2A02786A87C48E69F1CEA80E867500EA>.
- Zhang, Z. (2016a). Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *The Journal of the Acoustical Society of America*, 139(4), 1493–1507. <https://doi:10.1121/1.4944754>.
- Zhang, Z. (2016b). Mechanics of human voice production and control. *The Journal of the Acoustical Society of America*, 140(4), 2614–2635. <https://doi:10.1121/1.4964509>.

Appendix A. Supplementary model outputs for §5.2

This appendix provides the omnibus tests, estimated marginal means (EMMs), and Tukey-adjusted pairwise contrasts referenced in §5.2.

Table A 1 Omnibus tests for position (statements)

Outcome	NumDF	DenDF	F	p
Jitter (local %)	2	177	2.691	.071
Shimmer (local %)	2	167.358	0.994	.372
HNR (dB)	2	167.953	0.938	.394

Table A 2 Estimated marginal means by position (statements)

Outcome	Position	EMM	SE	df	95% CI (lower)	95% CI (upper)
Jitter	beginning	0.885	0.092	27.905	0.696	1.075
Jitter	middle	0.760	0.092	27.905	0.571	0.950
Jitter	end	1.027	0.092	27.905	0.838	1.216
Shimmer	beginning	8.864	1.005	13.888	6.707	11.022
Shimmer	middle	7.745	1.005	13.888	5.588	9.903
Shimmer	end	8.965	1.005	13.888	6.807	11.122
HNR	beginning	14.255	1.392	7.503	11.008	17.502
HNR	middle	14.604	1.392	7.503	11.357	17.850
HNR	end	14.897	1.392	7.503	11.650	18.143

Table A 3 Tukey-adjusted pairwise position contrasts (statements)

Outcome	Contrast	Estimate	SE	df	t	p (Tukey)
Jitter	beg - mid	0.125	0.116	169.816	1.080	0.528
Jitter	beg - end	-0.142	0.116	169.816	-1.225	0.440
Jitter	mid - end	-0.267	0.116	169.816	-2.305	0.058
Shimmer	beg - mid	1.119	0.965	169.467	1.159	0.479
Shimmer	beg - end	-0.101	0.965	169.467	-0.104	0.994
Shimmer	mid - end	-1.219	0.965	169.467	-1.263	0.418
HNR	beg - mid	-0.349	0.472	169.977	-0.738	0.741
HNR	beg - end	-0.642	0.472	169.977	-1.360	0.364
HNR	mid - end	-0.293	0.472	169.977	-0.621	0.809

Appendix B. Supplementary model outputs for §5.3

This appendix provides the omnibus tests, estimated marginal means (EMMs), and Tukey-adjusted pairwise contrasts for the mixed-effects models reported in §5.3.

Table B 1 Omnibus tests for position (wh-questions)

Outcome	NumDF	DenDF	F	p-value
Jitter (local %)	1	107.835	56.392	< .001
Shimmer (local %)	1	107.737	9.763	.002
HNR (dB)	1	107.952	14.546	< .001

Table B 2 Estimated marginal means by position (wh-questions)

Outcome	Position	EMM	SE	df	95% CI
Jitter	middle	1.291	0.159	5.737	[0.897, 1.685]
Jitter	end	0.723	0.159	5.737	[0.329, 1.117]
Shimmer	middle	7.470	0.840	5.722	[5.391, 9.549]
Shimmer	end	6.005	0.840	5.722	[3.926, 8.084]
HNR	middle	14.585	1.467	10.594	[11.341, 17.829]
HNR	end	16.184	1.467	10.594	[12.939, 19.428]

Table B 3 Tukey-adjusted pairwise position contrasts (wh-questions)

Outcome	Contrast	Estimate	SE	df	t ratio	p-value
Jitter	mid - end	0.567	0.076	108.850	7.477	< .001
Shimmer	mid - end	1.465	0.471	108.763	3.112	.002
HNR	mid - end	-1.599	0.421	108.961	-3.797	< .001

Appendix C. Supplementary model outputs for §5.4

This appendix provides the omnibus tests, estimated marginal means (EMMs), and Tukey-adjusted pairwise contrasts for the mixed-effects models reported in §5.4.

Table C 1 Omnibus tests for position (yes/no questions)

Outcome	NumDF	DenDF	F	p-value
Jitter (local %)	2	166.196	0.480	.620
Shimmer (local %)	2	167.872	1.070	.345
HNR (dB)	2	167.973	6.499	.002

Table C 2 Estimated marginal means by position (yes/no questions)

Outcome	Position	EMM	SE	df	95% CI
Jitter	beginning	0.966	0.102	24.401	[0.756, 1.176]
Jitter	middle	0.884	0.102	24.401	[0.674, 1.094]
Jitter	end	0.853	0.102	24.401	[0.643, 1.063]
Shimmer	beginning	9.026	1.331	6.632	[5.842, 12.209]
Shimmer	middle	8.038	1.331	6.632	[4.854, 11.221]
Shimmer	end	8.397	1.331	6.632	[5.213, 11.580]
HNR	beginning	13.956	1.673	7.184	[10.021, 17.891]
HNR	middle	14.626	1.673	7.184	[10.691, 18.561]
HNR	end	15.506	1.673	7.184	[11.571, 19.441]

Table C 3 Tukey-adjusted pairwise position contrasts (yes/no questions)

Outcome	Contrast	Estimate	SE	df	t ratio	p-value
Jitter	beg – mid	0.082	0.120	168.750	0.683	.774
Jitter	beg – end	0.114	0.120	168.750	0.945	.613
Jitter	mid – end	0.031	0.120	168.750	0.262	.963
Shimmer	beg – mid	0.988	0.688	169.900	1.437	.324
Shimmer	beg – end	0.629	0.688	169.900	0.915	.632
Shimmer	mid – end	-0.359	0.688	169.900	-0.522	.861
HNR	beg – mid	-0.670	0.434	169.997	-1.546	.272
HNR	beg – end	-1.550	0.434	169.997	-3.573	.001
HNR	mid – end	-0.879	0.434	169.997	-2.027	.109

Appendix D. Supplementary model outputs for §6.3.3

Table D 1 Fixed effects for HNR in statements

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	16.773	1.127	14.880	6.990	0.000
fixed	genderM	-5.036	1.011	-4.982	17.936	0.000
fixed	positionmiddle	0.396	0.657	0.603	167.922	0.547
fixed	positionend	-0.100	0.657	-0.152	167.922	0.880
fixed	genderM $\times$ positionmiddle	-0.095	0.929	-0.102	167.922	0.919
fixed	genderM $\times$ positionend	1.483	0.929	1.597	167.922	0.112

Table D 2 ANOVA for fixed effects for HNR in statements

Term	NumDF	DenDF	F value	p value
gender	1	9.342	28.485	0.000
position	2	167.922	0.958	0.386
gender $\times$ position	2	167.922	1.816	0.166

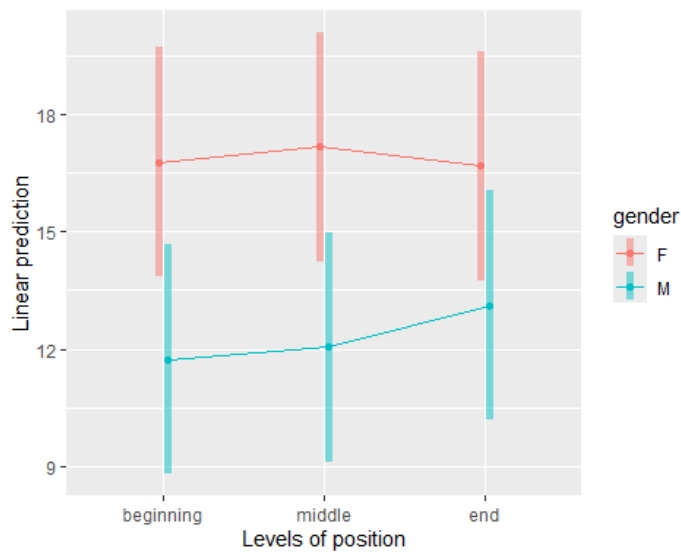


Figure D 1 Interaction plot of model-estimated means for HNR in statements

Appendix E. Supplementary model outputs for §6.4.1

Table E 1 Fixed effects for Jitter in wh-questions

Effect	Term	Estimate (β)	SE	t value	df	<i>p</i> -value
fixed	(Intercept)	1.071	0.149	7.173	6.936	0.000
fixed	genderM	0.439	0.139	3.148	17.727	0.006
fixed	positionend	-0.383	0.104	-3.687	107.824	0.000
fixed	genderM $\times$ positionend	-0.369	0.147	-2.511	107.824	0.014

Table E 2 ANOVA for fixed effects (Jitter in Wh-questions)

Term	NumDF	DenDF	F value	<i>p</i> value
gender	1	9.377	4.611	0.059
position	1	107.824	59.684	0.000
gender $\times$ position	1	107.824	6.304	0.014

Appendix F. Supplementary model outputs for §6.4.2

Table F 1 Fixed effects for Shimmer in wh-questions

Effect	Term	Estimate (β)	SE	t value	df	<i>p</i> -value
fixed	(Intercept)	7.127	0.791	9.005	6.326	0.000
fixed	genderM	0.685	0.724	0.947	26.441	0.352
fixed	positionend	-1.326	0.663	-2.001	107.730	0.048
fixed	genderM $\times$ positionend	-0.277	0.937	-0.296	107.730	0.768

Table F 2 ANOVA for fixed effects (Shimmer in Wh-questions)

Term	NumDF	DenDF	F value	<i>p</i> value
gender	1	9.306	0.984	0.346
position	1	107.730	9.771	0.002
gender $\times$ position	1	107.730	0.088	0.768

Appendix G. Supplementary model outputs for §6.4.3

Table G 1 Fixed effects for HNR in wh-questions

Effect	Term	Estimate (β)	SE	t value	df	<i>p</i> -value
fixed	(Intercept)	17.663	1.012	17.446	7.407	0.000
fixed	genderM	-6.157	0.964	-6.387	14.263	0.000
fixed	positionend	1.765	0.593	2.979	107.879	0.004
fixed	genderM $\times$ positionend	-0.332	0.838	-0.396	107.879	0.693

Table G 2 ANOVA for fixed effects for HNR in Wh-questions

Term	NumDF	DenDF	F value	<i>p</i> value
gender	1	9.426	53.053	0.000
position	1	107.879	14.562	0.000
gender $\times$ position	1	107.879	0.157	0.693

Table G 3 Simple effects of Gender within each Position (HNR in Wh-questions)

position	contrast	Estimate	Std. Error	df	t value	<i>p</i> value
middle	F - M	6.157	1.025	16.179	6.008	0
end	F - M	6.489	1.025	16.179	6.332	0

Table G 4 Pairwise comparisons of Position within each Gender (HNR in Wh-questions)

gender	contrast	Estimate	Std. Error	df	t value	<i>p</i> value
Female	middle - end	-1.765	0.598	109.919	-2.952	0.004
Male	middle - end	-1.433	0.598	109.919	-2.396	0.018

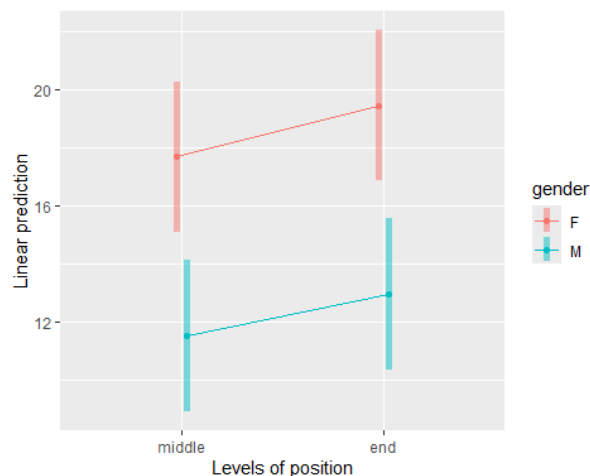


Figure G 1 Interaction plot of model-estimated means for HNR in wh-questions

Appendix H. Supplementary model outputs for §6.5.3

Table H 1 Fixed effects for HNR in Yes/No questions

Effect	Term	Estimate (β)	SE	t value	df	p-value
fixed	(Intercept)	16.826	1.323	12.717	6.188	0.000
fixed	genderM	-5.740	1.101	-5.212	14.522	0.000
fixed	positionmiddle	0.143	0.604	0.236	167.956	0.813
fixed	positionend	1.752	0.604	2.900	167.956	0.004
fixed	genderM $\times$ positionmiddle	1.055	0.854	1.235	167.956	0.219
fixed	genderM $\times$ positionend	-0.405	0.854	-0.474	167.956	0.636

Table H 2 ANOVA for fixed effects (HNR in Yes/No questions)

Term	NumDF	DenDF	F value	p value
gender	1	9.312	31.467	0.000
position	2	167.956	6.619	0.002
gender $\times$ position	2	167.956	1.556	0.214

Table H 3 Simple effects of Gender within each Position (HNR in Yes/No-questions)

position	contrast	Estimate	Std. Error	df	t value	p value
beginning	F - M	5.740	1.167	16.404	4.920	0.000
middle	F - M	4.685	1.167	16.404	4.016	0.001
end	F - M	6.145	1.167	16.404	5.267	0.000

Table H 4 Pairwise comparisons of Position within each Gender (HNR in Yes/No questions)

gender	contrast	Estimate	Std. Error	df	t value	p value
Female	beginning - middle	-0.143	0.611	172.053	-0.234	0.970
Female	beginning - end	-1.752	0.611	172.053	-2.865	0.013
Female	middle - end	-1.609	0.611	172.053	-2.632	0.025
Male	beginning - middle	-1.198	0.611	172.053	-1.959	0.126
Male	beginning - end	-1.347	0.611	172.053	-2.203	0.073
Male	middle - end	-0.149	0.611	172.053	-0.244	0.968

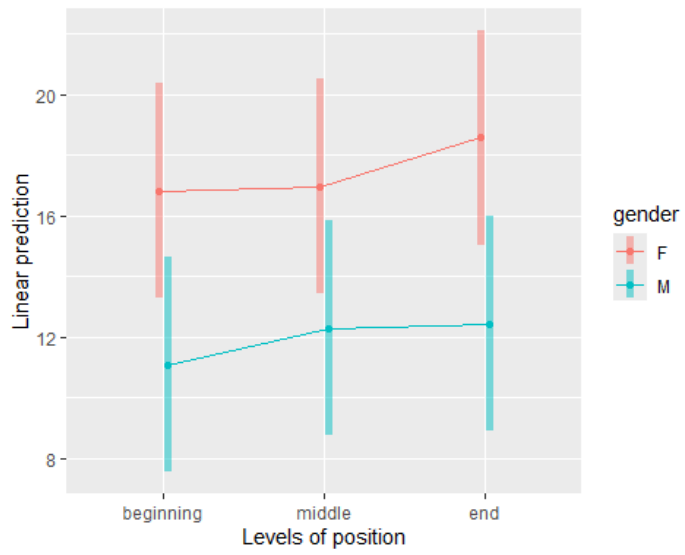


Figure H 1 Interaction plot of model-estimated means for HNR in yes/no questions