Humanities and Social Sciences Communications

Article in Press https://doi.org/10.1057/s41599-025-06454-8

Perceptual boundary of vowel quantity: a perceptual study of synthesized Arabic vowels

Received: 16 June 2024

Accepted: 18 December 2025

Cite this article as: Alfaifi, A. Perceptual boundary of vowel quantity: a perceptual study of synthesized Arabic vowels. *Humanit Soc Sci Commun* (2026). https://doi.org/10.1057/s41599-025-06454-8

Abdullah Alfaifi

Perceptual boundary of vowel quantity: A perceptual study of synthesized Arabic vowels

Abdullah Alfaifi
Department of English Language & Literature, College of Arts & Letters, University of Bisha, Bisha, Asir, Saudi Arabia
ahalfifi@ub.ed.sa

Abstract

This study investigates how native speakers of Najdi and Cairene Arabic perceive the contrast between short and long vowels, focusing on the role of vowel duration. Forty participants (20 per dialect) completed a forced-choice identification task using synthesized minimal pairs with systematically varied vowel durations. Mixed-effects logistic regression models were used to analyze how duration and dialect influenced responses. Duration was a strong predictor of "long" responses across all vowel pairs, but sensitivity to duration varied by dialect and vowel. For [i], Cairene listeners categorized vowels as "long" at shorter durations (84 ms) and showed steeper perceptual slopes than Najdi listeners, who required longer durations (96 ms) and had more gradual responses. For [a], both dialects showed a shared boundary at 101 ms, though Cairene speakers again responded more categorically. For [u], boundary differences were small (Najdi = 100 ms, Cairene = 110 ms) and not statistically significant. These findings suggest that while Arabic speakers rely on duration to distinguish vowel quantity, perceptual calibration is influenced by dialect and vowel quality. The study highlights the value of controlled synthesis and mixed-effects modeling for examining subtle variation in phonemic perception across dialects.

1. Introduction

In Arabic, there exists a phonemic contrast in segment quantity, which manifests in both consonants, specifically between singletons and geminates, and vowels, distinguishing between long and short vowels (de Jong & Zawaydeh, 2002). An intriguing question that emerges in this context is how the concepts of "long" and "short" are quantified as measurable time intervals, especially considering that languages seem to exhibit variations in how they time the duration of long and short vowels (Lehiste, 1970: 34). To better understand this phonemic distinction, it is important to consider the role of vowel quality and duration in Arabic.

In Arabic specifically, many production studies have reported that long vowels are roughly 1.5 to 3 times the duration of short vowels, depending on linguistic context and stress (e.g., Al-Ani, 1970; Alghamdi, 1998; de Jong & Zawaydeh, 2002). On average, long vowels might be approximately 1.65 times as long as short vowels. However, it remains to be understood how listeners perceptually distinguish these lengths – that is, what duration a listener considers "long" vs "short" and whether this perceptual boundary is stable or varies among native speakers of a single Arabic dialect, or among different dialects.

One factor that could influence vowel length perception is dialectal background. Arabic is a dialect-rich language, and its regional varieties (or dialects) can differ in pronunciation, prosody, and timing patterns. Most prior research on Arabic vowel length has focused on a single variety (often Modern Standard Arabic or one specific dialect), examining the acoustic characteristics of the length contrast. As a result, it is not yet known whether speakers of different Arabic dialects perceive the long/short vowel distinction in the same way. In theory, because the long vs. short vowel contrast is phonemic across Arabic dialects, one might expect that the perceptual processing of this contrast is similar for different dialect speakers. However, dialects could differ in subtler timing norms or vowel quality cues that accompany length, potentially leading to different perceptual calibration. For instance, dialects differ in overall speech rate and vowel realizations, which might shift how duration cues are interpreted. Specifically, we ask whether listeners from two different Arabic dialects (Najdi vs. Cairene) perceive the short/long vowel distinction at the same durational threshold, or if dialect-specific timing norms shift this perceptual boundary. We focus on two dialects, Najdi and Cairene Arabic, to represent distinct Arabic dialect regions.

Najdi Arabic is spoken in central Saudi Arabia, particularly in the Najd region (Ingham, 1994). Cairene Arabic is the dialect of Cairo, Egypt, and is a major variety of Egyptian Arabic (Haeri, 1997). These two dialects were chosen for comparison because they are linguistically and geographically distant, which provides a strong test for dialectal influence on perception. Both dialects have the same phonemic vowels as MSA: three short vowels /a/, /i/, and /u/ and three long vowels /a:/, /i:/, and /u:/. However, the dialects differ in other phonological aspects. For example, the Classical Arabic phoneme /q/ is pronounced as [g] in Najdi Arabic but as [ʔ] (glottal stop) in Cairene Arabic, and the two dialects have distinct prosodic patterns and intonation. Najdi Arabic is often described as somewhat more conservative in phonology (Ingham 1994, p. 6), while Cairene Arabic has certain innovations such as the realization of some short vowels and the number of long vowels compared to other Arabic dialects (Watson, 2002, p. 10). These differences might indirectly affect how vowel length is produced and perceived through differences in typical speaking rate or stress patterns (see Hamdi et al., 2004; Ghazali et al., 2002). It is important to clarify that in this paper, the term "dialect" is used to mean a regional spoken variety of Arabic, as opposed to MSA. All our participants are native speakers of either Najdi or Cairene Arabic in daily life, and all have also learned and been exposed to MSA in formal settings (e.g. education, media). Thus, they are bidialectal in the sense of using both their colloquial dialect and MSA. We selected participants who primarily speak their native dialect and have not been extensively exposed to the other dialect, to ensure that any perceptual differences observed can be attributed to dialectal background rather than individual bilingualism in the other variety.

To our knowledge, there has been no prior experimental study that directly and systematically compares the perceptual vowel-length boundary in speakers of different Arabic dialects, even though production studies have documented dialect differences in vowel acoustics. By examining Najdi and Cairene listeners, the study aims to determine whether the same acoustic duration is perceived equivalently by both groups (indicating a shared perceptual boundary), or whether their perceptual category boundary for vowel length differs, which may indicate dialect-specific calibration of what counts as a long vowel. This has implications for our understanding of phonemic category plasticity and whether exposure to different speech timing patterns in one's dialect can shift fundamental perceptual thresholds.

This paper aims to investigate the perceptual boundaries between short and long vowels in Najdi and Cairene Arabic. We predict that native speakers of these dialects will exhibit similar perceptual boundaries for vowel length, despite any phonological differences between the two varieties. This prediction is based on the fact that Najdi and Cairene Arabic share the same phonemic vowel length contrast; thus, we initially assumed that fundamental perceptual processing of vowel quantity would be similar across dialects, despite minor differences in speaking rate or vowel quality. The structure of the paper is as follows: the current section reviews the relevant literature on the acoustic properties of Arabic vowels; Section 2 outlines the methodology; Section 3 reports the results; Section 4 offers a discussion of the findings; and Section 5 concludes with a summary of the key outcomes and their implications.

1.1. Phonemic Distinction as the Result of Quantity Differences

For Modern Standard Arabic (henceforth MSA), it is generally accepted that the vowel inventory contains three long and three short variants: /i:/ and /i/, /u:/ and /u/, /a:/ and /a/, and that the distinction between long and short vowels does not pertain to variations in vowel quality but rather differences in the temporal duration of vowel articulation (Ryding, 2005, p. 25; McCarus, 2008, p. 239). However, with regard to the Arabic low vowels /a/ and /a:/, there exists some variation in scholarly perspectives regarding the nature of the distinction between long and short vowels. Al-Ani (1970, p. 75) states that these vowels not only differ in their temporal duration but also exhibit distinctions in vowel quality, while Ryding (2005, p. 25) states that the difference is purely durational. These differing perspectives on the nature of the distinction between long and short vowels in Arabic highlight the need for a comprehensive examination of the acoustic properties of Arabic vowels.

1.2. Acoustic Properties of Arabic Vowels

The acoustic properties of Arabic vowels play a crucial role in distinguishing between different vowel categories and contribute to the overall phonological system of the language. These properties include formant frequencies, duration, fundamental frequency, and intensity.

Arabic has six main vowel phonemes: three short vowels (/a/, /i/, /u/) and three long vowels (/a:/, /i:/, /u:/) (Al-Ani, 1970; Ryding, 2005; Al-Tamimi, 2007). The distinction between short and long vowels is phonemic in Arabic, meaning that the length of the vowel can change the meaning of a word (Al-Ani, 1970; Alghamdi, 1998). For example, the word *kataba* means "he wrote", while *ka:taba* means "he corresponded" (Huthaily, 2003). While it is often said that Modern Standard Arabic (MSA) has six core vowel phonemes, it is also common to include diphthongs such as /aj/ and /aw/ in descriptions of MSA. Many dialects, however, Cairene Arabic included, have partially or fully monophthongized these diphthongs into mid vowels (often transcribed as /eː/ or /oː/) (Philippa, et al, 2017). Such mergers mean that the vowel inventory in practice can differ substantially from the textbook MSA set, which introduces further complexity when comparing short and long vowels across Arabic dialects. These dialectal developments also highlight the fact that length contrasts can intersect with important changes in vowel quality. What is phonemically a short-long pair in one variety may be realized in another as two distinct monophthongs or as mid versus high vowels. Since this study is concerned specifically with the effect of duration on the perception of /a/, /aː/, /i/, /iː/, /u/, and /uː/, we set aside diphthongs and their dialectal reflexes.

One of the most important acoustic properties of vowels is their formant frequencies. Formants are the resonant frequencies of the vocal tract that are produced when air passes through it during speech (Kent & Read, 2002). The first two formants (F1 and F2) are particularly important in determining the quality of a vowel sound (Ladefoged & Johnson, 2011). F1 is inversely related to vowel height, while F2 is related to the degree of backness or frontness of the tongue (Huthaily, 2003).

Several studies have investigated the formant frequencies of Arabic vowels. Al-Tamimi (2007) conducted a study on Jordanian Arabic and found that the mean F1 values for the short vowels [a], [i], and [u] were 619 Hz, 386 Hz, and 429 Hz, respectively. For the long vowels [aː], [iː], and [uː], the mean F1 values were 677 Hz, 290 Hz, and 316 Hz, respectively. The mean F2 values for the short vowels were 1513 Hz ([a]), 1753 Hz ([i]), and 1169 Hz ([u]), while for the long vowels, they were 1369 Hz ([aː]), 2202 Hz ([iː]), and 869 Hz ([uː]).

In another study, Alghamdi (1998) investigated the vowels of Saudi Arabic and found similar patterns. The mean F1 values for the short vowels [a], [i], and [u] were 695 Hz, 400 Hz, and 450 Hz, respectively, while for the long vowels [aː], [iː], and [uː], they were 705 Hz, 285 Hz, and 300 Hz. The mean F2 values for the short vowels were 1590 Hz ([a]), 1980 Hz ([i]), and 1175 Hz ([u]), and for the long vowels, they were 1520 Hz ([aː]), 2320 Hz ([iː]), and 800 Hz ([uː]).

These studies demonstrate that Arabic vowels exhibit distinct formant patterns that help distinguish between different vowel qualities. Short vowels generally have higher F1 values than their long counterparts, indicating a lower tongue position. The F2 values show that [i] and [iː] are front vowels, [u] and [uː] are back vowels, and [a] and [aː] are central vowels (Al-Tamimi, 2007; Alghamdi, 1998). These formant patterns play a crucial role in vowel identification and contribute to the overall distinctiveness of Arabic vowels. While formant frequencies are primary acoustic cues for vowel quality, they work in conjunction with other acoustic properties to create the complete acoustic profile of Arabic vowels.

Other acoustic properties have been studied in Arabic vowels. One such property is the fundamental frequency (F0), which is related to the perceived pitch of a vowel (Khattab, 2002). Al-Tamimi (2007) found that the mean F0 values for Arabic vowels ranged from 120 Hz to 260 Hz, depending on the speaker’s gender and the vowel's position within a word.

Another acoustic property that has received attention is the intensity of Arabic vowels. Intensity refers to the amount of energy in a sound wave and is related to the perceived loudness of a vowel (Ladefoged & Johnson, 2011). Al-Ani (1970) found that Arabic vowels exhibited different intensity levels depending on their quality and position within a word, with low vowels ([a] and [aː]) generally having higher intensity than high vowels ([i], [iː], [u], and [uː]).

The acoustic properties of Arabic vowels can also be influenced by the surrounding consonants. The phenomenon of coarticulation, where the articulation of a sound is affected by the preceding or following sounds, has been observed in Arabic (Embarki et al., 2011). For example, the formant frequencies of a vowel may be influenced by the place of articulation of the adjacent consonants, leading to variations in vowel quality (Khattab, 2002).

Another important acoustic property of Arabic vowels, and central to this study, is their duration. The duration of Arabic vowels can be affected by various factors, such as stress, syllable structure, and speaking rate (Alghamdi, 1998). In general, stressed vowels tend to be longer than unstressed vowels, and vowels in open syllables (CV) are longer than those in closed syllables (CVC) (Al-Ani, 1970). Speaking rate also influences vowel duration, with faster speech rates leading to shorter vowel durations (Mitleb, 1984). Several studies have investigated the duration of Arabic vowels as pronounced by native speakers from various Arabic dialects. While there can be variations in vowel duration among native speakers and across dialects, the general consensus is that long vowels have a longer duration than their short counterparts.

In Iraqi Arabic, Al-Ani (1970, p. 75) reported that long vowels in connected speech are at least twice as long as short vowels, with short vowels typically ranging from 100–150 ms and long vowels ranging from 225–350 ms. Interestingly, in isolation, all three long vowels have a fixed duration of exactly 600 ms, while short vowels maintain a consistent duration of exactly 300 ms. Fathi and Qassim also found that long vowels are approximately twice as long as short vowels in Iraqi Arabic, with [iː] being 2.2 times longer than [i], [aː] being 1.9 times longer than [a], and [uː] being 1.8 times longer than [u].

In Jordanian Arabic, Mitleb (1984, p. 231) reported that the short-long vowel average ratio was 0.65 (i.e. long vowels are about 1.54 times longer than short vowels). De Jong & Zawaydeh (2002) found that the duration of long vowels was approximately 115 ms for unstressed vowels and about 145 ms for stressed long vowels, while the duration of short vowels was about 50 ms if unstressed and about 65 ms if stressed. Kalaldeh (2018) provided specific durations for each vowel in Jordanian Arabic, with [iː] at 131 ms, [i] at 45 ms, [aː] at 125 ms, [a] at 60 ms, [uː] at 138 ms, and [u] at 45 ms.

For Saudi Arabic speakers, Flege (1979, p. 64) found that the ratio of long to short vowels was about 2 to 1. Alghamdi (1998) identified three long-short pairs of vowels in Najdi (and Southern) Arabic, with durations of 275 ms for [iː], 116 ms for [i], 294 ms for [aː], 128 ms for [a], 304 ms for [uː], and 116 ms for [u]. In Syrian Arabic, Almbark and Hellmuth (2015) found that the duration of long vowels is approximately one and a half times longer than short vowels.

Lastly, in Egyptian Arabic, Norlin (1987) measured the duration of the six vowels and found the following values: [iː] is 147 ms, [i] is 73 ms, [aː] is 171 ms, [a] is 95 ms, [uː] is 159 ms, and [u] is 85 ms.

These studies demonstrate that while the exact durations and ratios of long to short vowels may vary across Arabic dialects, the overall pattern remains consistent, with long vowels having a longer duration than their short counterparts. This difference in vowel duration is a crucial aspect of the phonological systems of these dialects, contributing to the distinction between short and long vowel phonemes.

Additionally, other sources like Ryding (2005) and McCarus (2008), which do not specify the dialectal background of the speakers, reported that long vowels are twice as long as short vowels. McCarus (2008) further adds that unstressed word-final long vowels are only half as long as long vowels in stressed syllables. What these studies collectively highlight is that there is some variation in Arabic dialects when it comes to the duration of long vowels. Most of these studies, as noted by Newman and Verhoeven (2002), were done in a laboratory setting, which can occasionally result in somewhat unrealistic findings as in Al-Ani (1970) who noted that the duration of long vowels was precisely double that of short vowels.

1.3. Current Study

While extensive research has been conducted on the acoustic properties of Arabic vowels, there remains a significant gap in our understanding of how these properties, particularly duration, influence native speakers' perception. Previous studies have primarily focused on production data, leaving questions about perceptual boundaries largely unexplored. The distinction between short and long vowels in Arabic is a crucial aspect of the language's phonology, as vowel length can change the meaning of words. While previous studies have investigated various acoustic properties of Arabic vowels, such as formant frequencies, intensity, and coarticulation effects, one area that has received less attention is the perceptual boundary between short and long vowels. In other words, the point at which a vowel's duration causes it to be perceived as either short or long by listeners has not been thoroughly explored.

Building upon the existing research on Arabic vowel duration, this study aims to further investigate the perceptual boundaries between short and long vowels in Arabic. By focusing on the role of vowel duration while controlling for other acoustic variables, we seek to gain a deeper understanding of how native Arabic speakers perceive the distinction between short and long vowels.

The current study aims to address this gap in the literature by presenting the results of a perceptual experiment focusing on the role of vowel duration in the perception of Arabic vowels as short or long. The study involves native Arabic speakers from two distinct dialects: Cairene Arabic and Najdi Arabic.

This study acknowledges that the difference between short and long vowel pairs in Arabic is not solely based on duration. These vowel pairs also exhibit other internal characteristics, such as differences in formant frequencies, which contribute to their distinct qualities. However, the current study specifically focuses solely on the effect of vowel duration on the perception of vowels as short or long, while controlling for other acoustic variables. By isolating the role of duration in vowel perception, this study seeks to provide insights into the perceptual mechanisms underlying the distinction between short and long vowels in Arabic. The findings of this research contribute to a deeper understanding of Arabic phonetics and have implications for cross-linguistic studies of vowel perception. Furthermore, the inclusion of speakers from two distinct Arabic dialects in the study offers insights into the potential differences in vowel perception across regional varieties of Arabic.

2. Methodology

This section presents a description of the stimuli, participants, and data collection and analysis procedures.

2.1. Stimuli

The stimuli consisted of three MSA minimal pairs containing the long vowels [iː], [aː], and [uː] and their short counterparts [i], [a], and [u]. The words were selected based on their phonological structure (CVC) and the presence of the target vowels in a stressed syllable. For consistency, all selected words were monosyllabic, making all vowels stressed. The original recordings of these words were produced by a native speaker of Saudi Arabic, with the long vowels ranging in duration from 151 ms to 176 ms and the short vowels ranging from 72 ms to 87 ms.

The speaker who produced the initial recordings for this study is a native speaker of Saudi Arabic, though not from the Najd region. Importantly, these words were recorded in Modern Standard Arabic (MSA), a register often heard in news media and formal contexts across the Arab world. Although MSA aims to minimize distinct local dialectal features, some subtle phonetic coloring from the speaker's regional background may still be present. Nevertheless, using MSA helps reduce strong dialect-specific prosody or consonantal realizations that could unduly favor one group of listeners over another.

Because participants in this study included both Najdi Arabic and Cairene Arabic speakers, it is possible that a dialect alignment or mismatch might affect perception, particularly if any regional features had been audible. However, our stimuli were specifically manipulated only in their vowel durations, with other factors (e.g., formant transitions, consonant production) left unaltered or minimized. In this sense, the effect of dialect familiarity is likely limited, as we did not systematically exploit dialect-specific cues. Still, it cannot be entirely ruled out that Najdi speakers, in principle, might have found the MSA stimuli slightly more familiar than the Cairene speakers.

To create the stimuli for the perception experiment, the durations of the long vowels in the original recordings were systematically manipulated. The duration of each long vowel was reduced in increments of 10 milliseconds until it matched the duration of its corresponding short vowel. This process resulted in the creation of twenty-eight stimuli, as shown in Table 1.

Table 1. Vowel Durations in Stimuli

Word	Vowel Duration	Gloss	Word	Vowel Duration	Gloss
/tˤi:b/	176 ms	*Scent*	/tˤib/	84 ms	*Medicine*
/ta:m/	151 ms	*Complete*	/tam/	87 ms	*Done*
/ku:b/	160 ms	*Cup*	/kub/	72 ms	*Pour*

All stimuli were embedded in carrier phrases and produced without any specific emphasis, ensuring that they were unfocused in the context of the carrier phrase. This approach allowed for the evaluation of the perception of these synthesized vowels by native Arabic speakers. As shown in the following figure, the durations of the long and short vowels in the current study are very similar to the majority of studies which measured the duration of vowels in Arabic dialects.

In this study, all vowel durations were measured manually by the author using Praat (Boersma & Weenink, 2024). Each token was segmented by inspecting both the waveform and the spectrogram to determine vowel boundaries. Specifically, the onset of a vowel was marked at the first observable periodic waveform and visible formant structure following the release or frication of the preceding consonant. The offset was determined at the point where periodicity ended or immediately before the closure for the following consonant became evident.

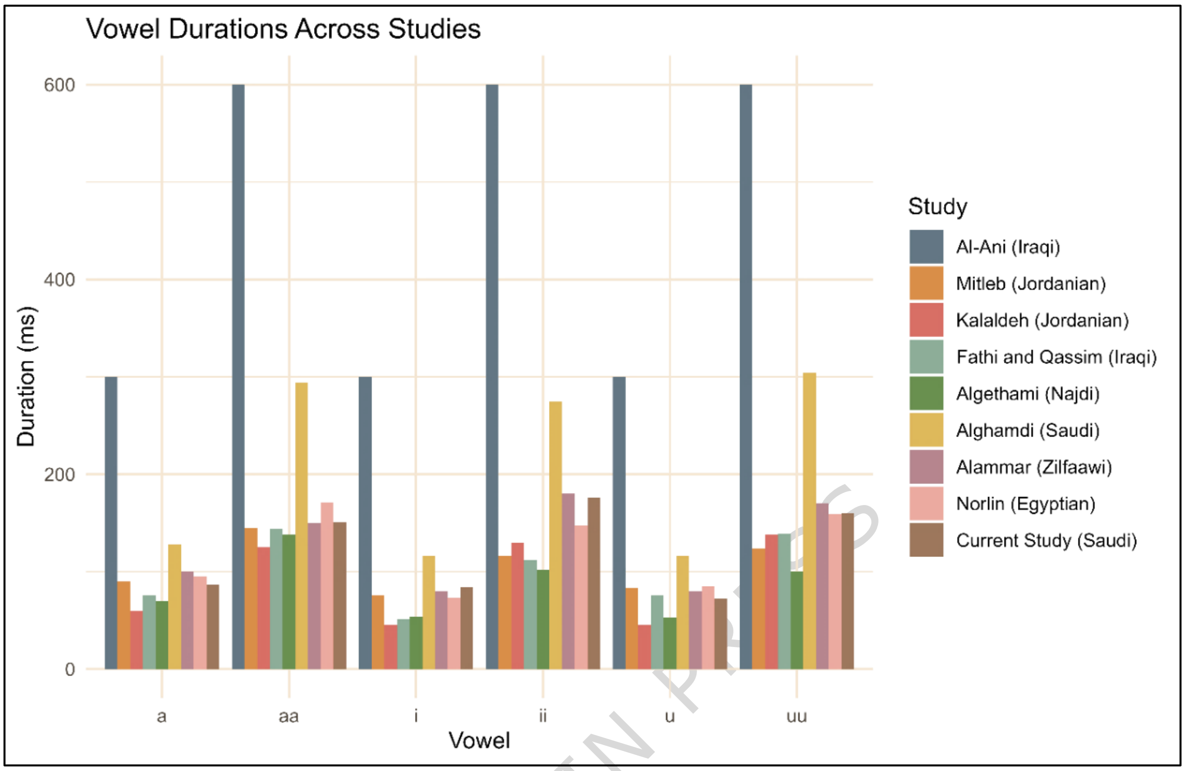

Figure 1. *Durations of Arabic vowels as reported in Al-Ani (1970), Mitleb (1984), Kalaldeh (2018), Fathi & Qassim (2020), Algethami (2023), Alghamdi (1998), Alammar (2023), Norlin (1987), and the durations in the current study.*

To generate the 28 stimuli, there were two options at hand. The first option was to systematically clip 10ms from the beginning, middle, or end of the vowel; or by removing excised vowel pitch periods as in Al-Bannai (1995). The second option was to modify the duration of the vowel by increasing its tempo. That is, increasing the tempo of a long vowel in order to reduce its duration, rather than physically clipping a fixed duration. The tempo-based method was preferred for several reasons outlined below

First, increasing the tempo of the vowel compresses the vowel in time while maintaining its overall characteristics, such as formant transitions, pitch contours, and spectral shape. This ensures that the shortened vowel still sounds like the original vowel, albeit with a shorter duration. On the other hand, physically clipping a fixed duration from the vowel may result in abrupt changes or loss of important vowel characteristics, which would make it sound unnatural or distorted. Additionally, by increasing the tempo, we can achieve a consistent and proportional reduction in vowel duration across the entire vowel segment. This means that the relative timing and structure of the vowel are largely preserved, even though the overall duration is shortened. In contrast, clipping a fixed duration from the vowel may result in inconsistent or disproportionate changes to the vowel's internal structure, depending on the specific vowel and its duration. Second, increasing the tempo allows for relatively smooth transitions between the shortened vowel and the surrounding segments. The vowel is compressed in a way that maintains the continuity and flow of the speech signal. Clipping a fixed duration, especially from the center of the vowel, may introduce abrupt or unnatural transitions, leading to audible artifacts or discontinuities in the resulting audio.

The script gradually increased the tempo of the long vowels, resulting in a systematic reduction of their duration by approximately 10 ms in each step. This process was repeated until the duration of the long vowel matched or closely approximated the duration of its short counterpart. The script ensured that the vowel compression maintained the overall characteristics of the vowel while avoiding abrupt changes or distortions in the audio signal. For example, the original duration of [uː] was 160 ms and the original duration of [u] was 72 ms. The target durations are 150 ms, 140 ms, 130 ms, 120 ms, 110 ms, 100 ms, 90 ms, 80 ms, and 70 ms. The tempo factor change was calculated by dividing the original duration with the target duration (160ms / 150ms = 1.0667). The percentage of tempo increase was calculated using the formula (tempo increase = (1.0667 - 1) * 100% = 6.67%), which means that in order to reduce the duration of [uː] from 160 ms to 150 ms, the long vowel's tempo is increased by 6.67%, and so on, as shown in Table 2.

Table 2. Percentage of Tempo Change and the Resulting Durations

[uː] = 160 ms, [u] = 72 ms		[aː] = 151 ms, [a] = 87 ms		[iː] = 176 ms, [i] = 84 ms	
Duration	Tempo Change	Duration	Tempo Change	Duration	Tempo Change
70 ms	122.22%	87 ms	73.56%	84 ms	109.52%
80 ms	100%	91 ms	66%	96 ms	83.33%
90 ms	77.78%	101 ms	49.50%	106 ms	66.04%
100 ms	60%	111 ms	36%	116 ms	51.72%
110 ms	45.45%	121 ms	24.79%	126 ms	39.68%
120 ms	33.33%	131 ms	15.27%	136 ms	29.41%
130 ms	23.08%	141 ms	7.09%	146 ms	20.55%
140 ms	14.29%			156 ms	12.82%
150 ms	6.67%			166 ms	6.02%

The tempo of the long vowels was gradually increased to reduce vowel duration using Audacity's Change Tempo effect (Audacity Team, 2024), which implements pitch-preserving time-stretching algorithms. These algorithms, based on Waveform Similarity Overlap-Add (WSOLA) and in high-quality mode, Subband Sinusoidal Modeling (SBSMS), perform synchronized time-stretching and pitch correction in a single processing step. This integrated approach is superior to applying pitch correction as a separate post-processing step, which would introduce additional artifacts and potentially distort formant frequencies.

To verify the effectiveness of the pitch preservation, we measured F0 values at the temporal midpoint of all manipulated stimuli using Praat (Boersma & Weenink, 2024), where pitch is typically most stable and least affected by consonantal transitions. The manipulated stimuli showed F0 values ranging from 109.6 minimum to 113.2 Hz maximum for [iː], 111.8-113.7 Hz for [aː], and 122-124 Hz for [uː], which represents a narrow range of pitch values across all experimental items. Crucially, participants made all perceptual judgments based solely on this coherent set of manipulated stimuli, after an initial familiarization phase with unmanipulated tokens. Since all experimental stimuli exhibited similar F0 values with minimal variation within each vowel category, and participants were not comparing manipulated to unmanipulated stimuli, any uniform pitch characteristics across the stimulus set would not affect the perception of relative vowel duration.

It is important to acknowledge that the phonetic difference between short and long vowels in Arabic involves more than just duration. For instance, phonemic short [i] is often articulated closer to [ɪ], making it more centralized relative to its long counterpart [iː]. However, the current study intentionally focused on manipulating only vowel duration, while eliminating or minimizing variation in other phonetic features. We aimed to examine whether, and how, listeners rely on duration alone when making perceptual judgments about vowel length, while holding other factors (e.g. formant transitions, F0, and spectral shape) constant as much as possible. By doing so, we were able to isolate the specific contribution of duration to the short-long vowel contrast. Nonetheless, we fully recognize that these additional qualitative differences between short and long vowels (particularly centralization in short vowels) likely play a significant role in everyday speech perception.

2.2. Participants

Forty native Arabic speakers (20 Najdi Arabic speakers and 20 Cairene Arabic speakers) were recruited for this study. Participants were selected based on the following criteria: (1) being native speakers of either Najdi or Cairene Arabic, (2) having no reported history of hearing or speech disorders, and (3) being within the age range of 18-40 years. The sample consisted of an equal number of male and female participants in each dialect group.

2.3. Data Collection Procedure

Data collection was conducted utilizing a Phonic audio survey (Infillion, 2024). Following the acquisition of requisite authorization from the research committee, all audio files were uploaded to Phonic, and the corresponding link was distributed to the participants. Prior to starting the trials, each participant received an orientation session to become familiar with the procedures.

The audio files were presented to participants in a pre-randomized fixed sequence. This means that the order of the stimuli was randomized prior to the commencement of the experiment, and this randomized sequence remained consistent across all participants, as shown in Table 3.

Table 3. Order of Stimuli as Presented to the Participants

[iː] Original Duration = 176 ms	[aː] Original Duration = 151 ms	[uː] Original Duration = 160 ms
156 ms	131 ms	120 ms
84 ms	87 ms	130 ms
166 ms	111 ms	80 ms
116 ms	121 ms	70 ms
126 ms	91 ms	90 ms
106 ms	141 ms	140 ms
96 ms	101 ms	100 ms
136 ms		150 ms
146 ms		110 ms

The choice to present the stimuli in a pre-randomized fixed sequence was made to ensure that any potential order effects would be consistent across all participants.

The data collection proceeded in two phases. The initial phase of the data collection process required the participants to listen to the minimal pairs in their original, unaltered form (i.e., the original recordings without any changes in the tempo) and categorize them as either different or the same. As expected, all minimal pairs were judged to be different at this stage. This phase of the survey was structured to allow participants to listen to each stimulus multiple times, should they want to do so.

This phase of the experiment was necessary for three reasons. First, by having participants judge the original, unaltered minimal pairs, the researcher can confirm that the chosen stimuli are indeed perceived as different by the participants. Second, this phase also helps to screen participants and ensure that they have the necessary auditory discrimination skills to complete the task. Third, listening to the original minimal pairs allows participants to become acquainted with the nature of the stimuli.

The second phase of the experiment involved participants listening to the audio files and judging whether they heard the word with the long vowel or the word with the short vowel. Participants were allowed to listen to each audio file as many times as they wished before making their judgment. However, the survey was designed in such a way that once participants moved on to the next stimulus, they were not able to go back and listen to or change their response to the previous stimulus. This approach ensured that participants' judgments were based on their initial perception of each audio file and prevented any potential bias that could arise from re-evaluating previous stimuli in light of subsequent ones. The collected responses were fed into an Excel file and used for analysis in R (R Core Team, 2024), and specifically the packages "ggplot2" (Wickham, 2016) and "tidyverse" (Wickham et al., 2019).

It is important to note that while the initial phase used unmanipulated recordings for familiarization, all perceptual boundary judgments in the main experimental phase were based exclusively on the manipulated stimuli. This design ensures that any acoustic characteristics introduced by the tempo manipulation process (such as minor pitch variations) were constant across all experimental trials and therefore could not systematically bias the perception of vowel duration. Participants were effectively making relative duration judgments within a coherent acoustic context, similar to how vowel length contrasts are perceived in natural speech across different speakers and prosodic contexts

The perceptual boundary was calculated with a code that finds the vowel duration where the proportion of responses for the short and long vowels is closest to being equal by minimizing the absolute difference between the proportions of responses for the short and long vowels at each vowel duration. Based on the responses, the calculation identifies the point where the perception of the vowel sound is most evenly split between the short and long variants.

Mixed-effects logistic regression models were fitted using the lme4 package (Bates et al., 2015) with the binary response (long vs. short) as the dependent variable and log-transformed duration (mean-centered) as the primary predictor. Log(Duration) was mean-centered by subtracting the dataset's mean log-duration (∼113 ms) so that the intercept in each model corresponds to the log-odds of a "long" response at the average vowel duration. Models included random intercepts for both subjects and items, and by-subject random slopes for duration to account for individual differences in sensitivity to durational cues. Dialect was included as a fixed effect along with its interaction with duration. Statistical significance was assessed using likelihood ratio tests and model comparison.

3. Results

3.1. Individual Differences in Duration Sensitivity (Mixed-effects Models)

Mixed-effects logistic regression models were fit separately for each vowel using the glmer function in R. The models included log-transformed vowel duration *(logDuration)*, Dialect (sum-coded: Cairene vs. Najdi), and their interaction as fixed effects. Random intercepts were included for both subjects and items, and by-subject random slopes for *logDuration* were specified to model individual variability in durational sensitivity. Cairene listeners served as the reference level for the Dialect contrast, allowing dialect differences to be interpreted as deviations from the Cairene baseline.

Table 4 presents the fixed effects from the mixed-effects logistic regression models predicting the probability of a "long" response, with separate models run for each vowel. As expected, log vowel duration was a strong and highly significant predictor in all three models: longer durations were associated with higher likelihoods of a "long" response. For [i] and [a], the interaction between *logDuration* and Dialect was significant and negative, indicating that Najdi listeners showed shallower slopes than Cairene listeners. In other words, they were less sensitive to duration increases when identifying vowels as long. For [u], no significant dialectal interaction was found. These results suggest that vowel perception is not only duration-dependent, but that dialectal background appears to modulate sensitivity to durational cues.

Table 4. Fixed-effects results of mixed-effects logistic regression models predicting "long" vowel responses, by vowel. Intercepts represent log-odds at mean duration (113.39 ms) with Cairene as reference. Log(duration) was mean-centered; *p*-values are Bonferroni-corrected.

Vowel	Predictor	$\hat{\beta}$	SE	*z*	*p* (Bonf.)
[i]	(Intercept)	**1.478**	**0.223**	**6.627**	< .001
	Mean-centered log(Duration)	18.617	4.437	4.195	< .001
	Dialect = Najdi	**0.693**	**0.315**	**2.201**	**.084**
	Mean-centered log(Duration) × Dialect	−13.549	3.711	−3.651	< .001
[u]	(Intercept)	**−0.068**	**0.201**	**−0.338**	**1.000**
	Mean-centered log(Duration)	8.961	2.398	3.737	< .001
	Dialect = Najdi	**0.584**	**0.284**	**2.056**	**.120**
	Mean-centered log(Duration) × Dialect	1.794	1.237	1.450	**.434**
[a]	(Intercept)	**0.231**	**0.189**	**1.222**	**.666**
	Mean-centered log(Duration)	26.713	5.363	4.981	< .001
	Dialect = Najdi	**0.267**	**0.267**	**1.000**	**.951**
	Mean-centered log(Duration) × Dialect	−17.976	4.253	−4.227	< .001

As shown in Table 4, log(Duration) had a large positive coefficient for each vowel, reflecting that longer vowel durations greatly increased the odds of a "long" response (all $p < .001$ after correction). The negative Dialect × logDuration coefficients for [i] and [a] indicate that the slope of the duration effect was significantly shallower for Najdi listeners than for Cairene listeners in those vowel contexts. There was no significant interaction for [u]. It is also worth noting that the Dialect (Najdi) main effect for [i] was positive in the mean-centered model, but did not reach significance after Bonferroni correction ($p = .084$). This coefficient suggests that, at the average vowel duration, Najdi listeners were slightly more likely to respond "long" than Cairene listeners for [i], but this difference is not robust. No significant main effect of Dialect was found for [u] or [a] (after correction, $p = .120$ and $p = .951$, respectively). In summary, the dialect differences in perception manifested primarily in the interaction with duration (i.e. differences in how sharply listeners transitioned from "short" to "long" as duration increased).

Figures 2-4 below illustrate variability in listeners' sensitivity to vowel duration, plotted by dialect for each vowel. Steeper slopes indicate greater reliance on durational cues, while shallower slopes indicate reduced sensitivity.

3.1.1. Vowel [i].

For [i], Figure 2 shows that Cairene listeners exhibited consistently **steeper slopes**, reflecting sharper perceptual transitions between short and long categories. Najdi listeners, by contrast, had **more gradual slopes**, indicating reduced sensitivity to duration.

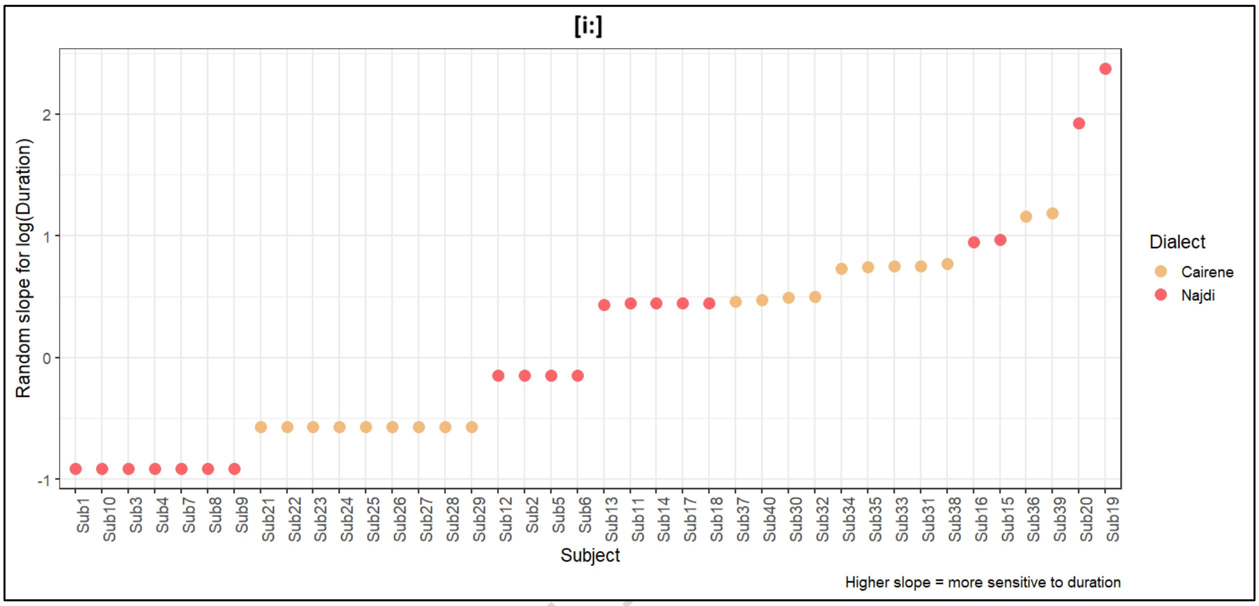

Figure 2. Listener-specific slopes for log(Duration) for vowel [iː], separated by dialect. Higher slopes indicate greater sensitivity to duration.

This aligns with the model's significant Dialect × *logDuration* interaction: duration influenced both groups' perception, but Cairene listeners showed more categorical behavior along the continuum.

3.1.2. Vowel [u].

For [u], Figure 3 reveals relatively steep slopes in both dialect groups, consistent with robust use of duration in distinguishing [u] from [uː].

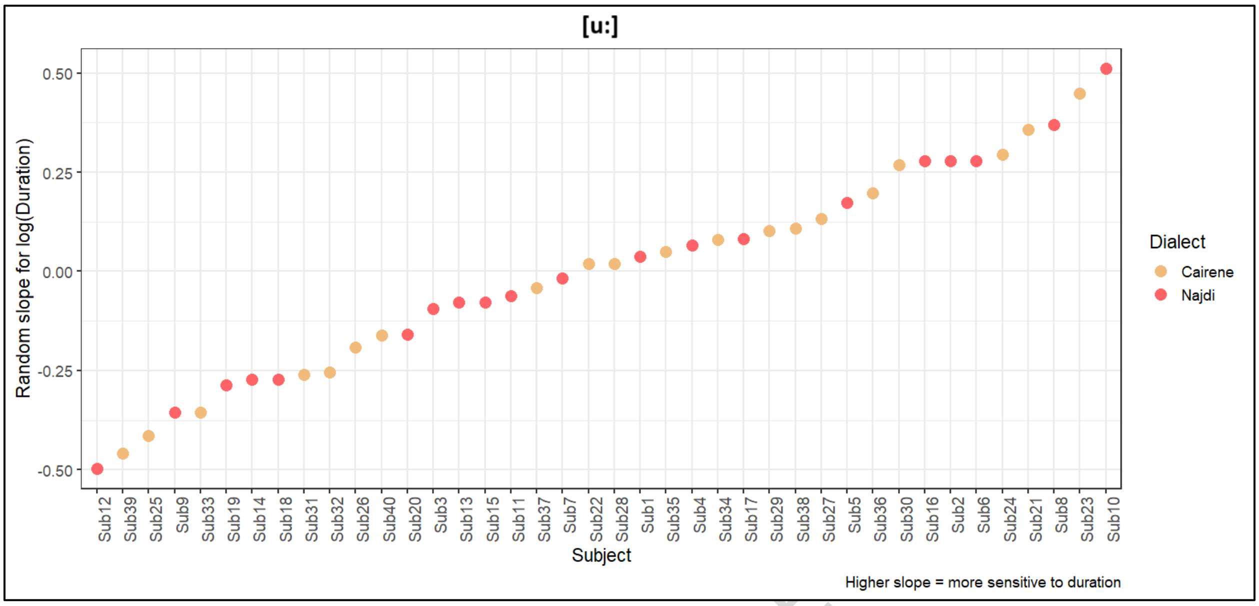

Figure 3. Listener-specific slopes for log(Duration) for vowel [u:], separated by dialect. Higher slopes indicate greater sensitivity to duration.

While Cairene and Najdi speakers did not differ significantly in slope magnitude, individual variability was somewhat greater among Cairene listeners. The mixed-effects model confirmed that no significant Dialect × Duration interaction was present for [u], indicating perceptual alignment across groups for this vowel.

3.1.3. Vowel [a].

For [a], both groups showed categorical responses, but Figure 4 shows that Cairene listeners had steeper identification slopes than Najdi listeners. This was confirmed by the significant negative Dialect × *logDuration* interaction, indicating that Najdi speakers had a more gradual shift in perception across durations. Although the average boundary was the same (101 ms), Cairene listeners were more sharply tuned to durational differences.

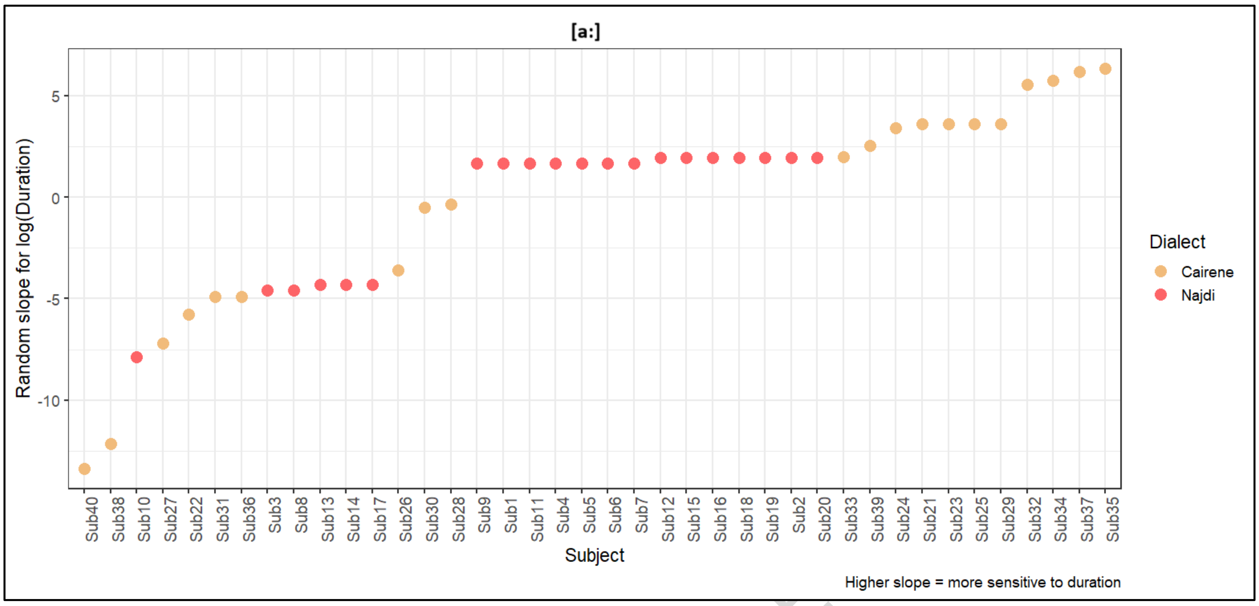

Figure 4. Listener-specific slopes for log(Duration) for vowel [aː], separated by dialect. Higher slopes indicate greater sensitivity to duration.

3.2. Descriptive Analysis of Perceptual Boundaries

The figures presented in this section compare the collective perceptual boundaries between Najdi and Cairene Arabic speakers for each vowel pair: [a] vs. [aː], [i] vs. [iː], and [u] vs. [uː]. Each plot displays how the perception of vowel length evolves across varying durations.

The x-axis in each plot represents the vowel duration in milliseconds, covering the full continuum tested in the experiment. The y-axis reflects the number of “long” responses out of 20 per duration point, with each line tracking identification trends across the stimulus set. Solid lines represent Najdi speakers, and dashed lines represent Cairene speakers. The complementary nature of the lines, often appearing as mirror images, is due to the binary forced-choice task: each duration has exactly 20 responses, split between “short” and “long.”

Similar response curves across dialects suggest shared perceptual calibration; divergent curves imply dialect-specific boundary locations or sensitivity.

3.1.4. [i] vs. [iː]

Figure 5 illustrates the perceptual responses to the [i] ~ [iː] contrast. For Cairene Arabic speakers, the perceptual boundary appears near 84 ms, with responses shifting sharply from “short” to “long” at that point. Najdi speakers show a boundary closer to 96 ms, with a slightly more gradual transition.

The data support these thresholds: at 84 ms, Cairene listeners responded “long” 14 times and “short” 6 times, while Najdi listeners gave 13 “short” and 7 “long” responses, indicating that 84 ms was still predominantly perceived as short in Najdi Arabic. At 96 ms, Najdi speakers responded "long" 12 times, showing a perceptual crossover around this duration.

Cairene speakers showed more categorical behavior, with a steep slope and little variation near the boundary. Najdi speakers, by contrast, required slightly more duration to consistently categorize the vowel as long, and their responses were more gradual, indicating shallower slope estimates in the model.

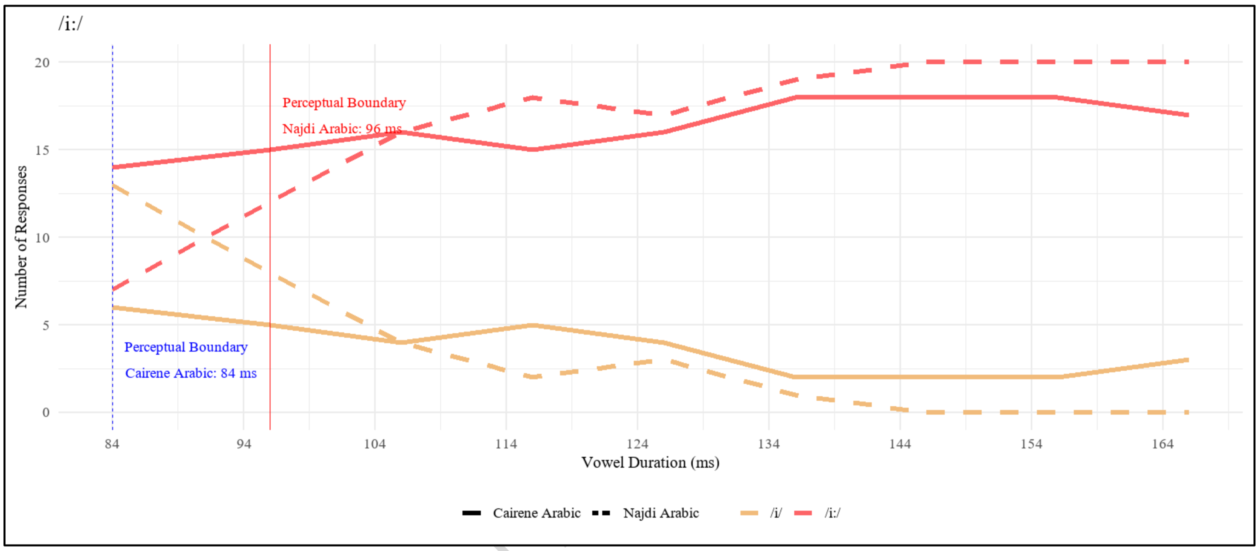

Figure 5. Perceptual identification curves for the [i]~[i:] vowel contrast for Cairene and Najdi speakers.

3.1.5. [a] vs. [a:]

Figure 6 shows the collective perceptual boundaries between the short vowel [a] and its long counterpart [a:] for Cairene Arabic and Najdi Arabic speakers. The perceptual boundary for both Cairene Arabic and Najdi Arabic is at 101 ms, indicating a strong similarity between the two dialects for this vowel pair.

The data confirm this finding, showing that at 101 ms, Cairene Arabic speakers identified the vowel as short 11 times and as long 9 times, while Najdi Arabic speakers identified it as short 10 times and as long 10 times, indicating a boundary near this duration where responses are approximately balanced.

Najdi speakers have a perceptual boundary very similar to Cairene speakers for [a] vs. [a:], around 100–110 ms. By around 121 ms, most listeners perceive the vowel as [a:] (long) consistently, whereas at shorter durations (below 91 ms) the majority of responses are [a] (short).

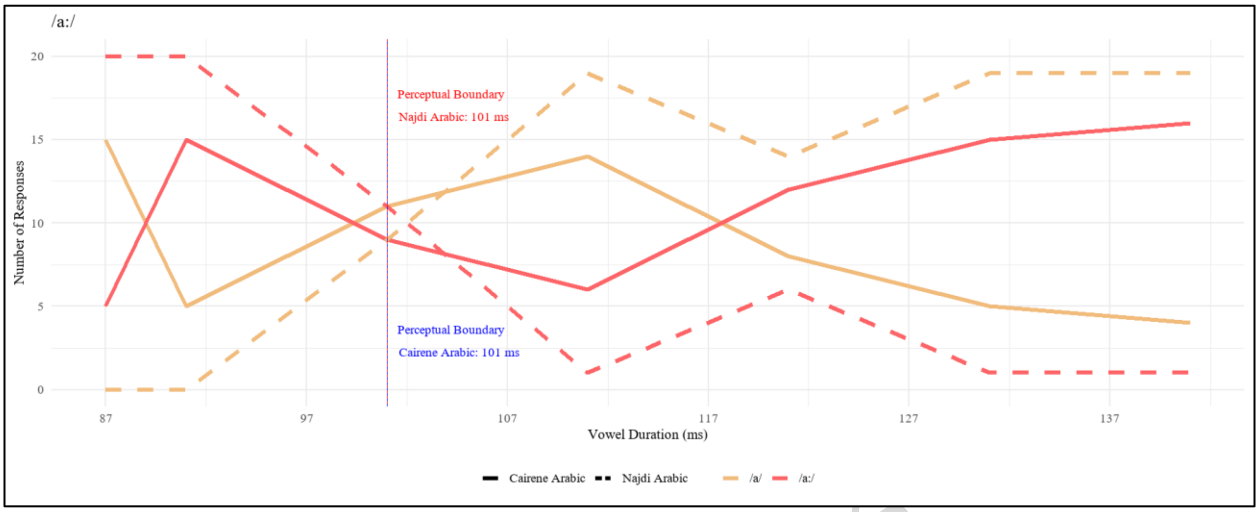

Figure 6. Perceptual identification curves for the [a]~[aː] vowel contrast for Cairene and Najdi speakers.

3.1.6. [u] vs. [uː]

Figure 7 presents the perceptual responses to [u] vs. [uː] across durations for Cairene and Najdi speakers. The estimated boundary for Cairene Arabic listeners is around 110 ms, while for Najdi Arabic it is slightly lower at 100 ms. This difference is numerically small and not statistically significant, indicating broad alignment between dialects for this vowel pair. At 100 ms, Cairene speakers categorized the vowel as short 13 times and long 7 times; at 110 ms, there were 9 "long" and 11 "short" responses, confirming a crossover near 110 ms. For Najdi listeners, the distribution at 100 ms was 11 "long" and 9 "short" responses, indicating a perceptual boundary in the 100–110 ms range. While Najdi speakers show slightly earlier categorization of [uː], both groups demonstrate categorical shifts over a narrow durational window, with limited dialectal divergence. Some individual variation remains: a few participants responded "long" to [uː] stimuli as short as 80 ms, while others required over 100 ms for consistent identification.

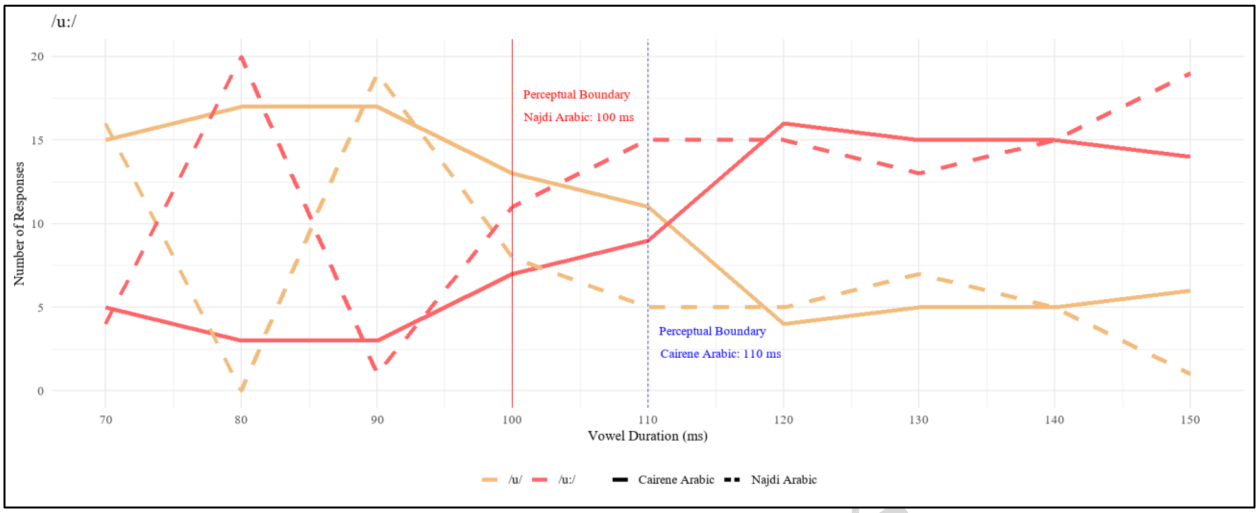

Figure 7. Perceptual identification curves for the [u]~[u:] vowel contrast for Cairene and Najdi speakers.

In summary, Figures 5–7 illustrate the perceptual boundaries between short and long vowels for Cairene and Najdi Arabic speakers. For [a] vs. [a:], both dialects show nearly identical boundary estimates at 101 ms, suggesting shared perceptual calibration for this low vowel. In contrast, a significant dialectal difference was found for [i] vs. [i:], with Cairene speakers categorizing shorter durations as "long" earlier than Najdi speakers (84 ms vs. 96 ms). For [u] vs. [u:], the estimated boundary for Cairene speakers (110 ms) was slightly higher than for Najdi speakers (100 ms), though this difference was not statistically significant.

For descriptive purposes, mean perceptual boundary locations were calculated for each vowel pair by dialect. These values represent the duration at which participants responded "long" on 50% of trials, derived from the fitted logistic models. Table 5 summarizes these estimates, supporting the conclusion that while both dialects exhibit robust sensitivity to duration, only the perceptual boundary for [i:] differs significantly by dialect, whereas boundaries for [a:] and [u:] are largely overlapping and statistically comparable.

Table 5. Estimated perceptual boundary locations (in ms) for each vowel pair, based on 50% "long" response thresholds.

Vowel pair	Najdi Mean Boundary (ms)	Cairene Mean Boundary (ms)
[i]~[i:]	96	84
[u]~[u:]	100	110
[a]~[a:]	101	101

4. Discussion

4.1. Individual variation in duration sensitivity

One of the central findings of this study is the pronounced individual variation in how Arabic listeners perceive vowel duration. Even though all participants are native speakers and thus have extensive experience with the long–short vowel contrast, their perceptual boundaries were not identical. Some listeners had an earlier boundary, labeling vowels as "long" at shorter durations, while others had a later boundary, requiring more acoustic length before hearing a long vowel. This result highlights an important point: phonetic category boundaries in our data are not fixed constants, even within a single speech community (though see Theodore et al., 2015, for evidence of boundary stability across talkers in a different perceptual domain). Cognitive and experiential differences between individuals, such as subtle differences in language exposure, personal cue weighting strategies, or even attentional factors, can lead to measurable differences in speech perception (see Kogan & Mora, 2022; Jasmin et al., 2021; Clayards, 2018). Our data thus add to a body of literature demonstrating that individuals often calibrate phonemic contrasts differently (e.g., see discussions in Yu 2021 on individual phonologization patterns, and in studies of categorical perception variability).

It is worth emphasizing that all listeners in our study understood the task and performed above chance – the variability we observe is a matter of degree, not a case of some listeners failing to perceive the length distinction. While categorical identification patterns are especially clear for [a] and [u], some participants, particularly Najdi listeners for [i], show more gradual transitions, which reflects individual differences in sensitivity to duration cues. The differences lie in exactly where and how sharply each person's identification curve transitions. Such differences have been observed in other phonemic contrasts such as in voice-onset-time boundaries for stop consonants (McMurray et al., 2008), or in tone perception (Francis et al., 2003), and they remind us that the human perceptual system allows some flexibility in category definitions. In our results, the steepness of the logistic slopes varied by listener – some had near-binary responses around the boundary, while others had a more gradual change – and the random-effect analysis confirmed that both intercepts and slopes differed significantly across individuals.

Why might individual Arabic listeners vary in this way? One possibility could be attention to secondary cues. Although we only manipulated duration, in natural speech long vs. short vowels can co-vary with quality differences, intensity, or contextual prosody. Listeners who subconsciously expect those additional cues might be slightly more conservative in judging "long" based on duration alone. On the other hand, listeners who rely almost exclusively on duration might treat any sufficiently longer vowel as long, even if the difference is small.

The findings here align with the notion that phonetic perception involves an interplay of general auditory abilities and language-specific experience (see Repp, 1982; Tsukada, 2012). All participants have the language-specific knowledge that vowel length is contrastive in Arabic, so none ignored the duration cue. But individual cognitive differences – such as how sharply one discriminates timing or how consistently one applies category labels – introduce variance around the group norm. From a practical perspective, this means that in studies of speech perception it is valuable to consider individual data rather than only averages. Important patterns can be obscured by averaging, and indeed, our analysis would have missed the rich variability and its implications if we had not used models that capture listener-specific effects. In summary, the individual differences in duration sensitivity observed in this study underscore that perceiving a "long" vs. "short" vowel is not a one-size-fits-all process.

4.2. Dialect-Specific Patterns in Vowel Quantity Perception

Our results provide evidence that dialect background influences vowel length perception in subtle but telling ways. Najdi and Cairene Arabic are both varieties of Arabic with the same phonemic long/short distinctions, yet we found systematic differences in how listeners from these dialects perceive the duration continuum for certain vowels. This finding is theoretically intriguing: it suggests that even when two dialects *phonologically* share a contrast, the *phonetic implementation* and perceptual weighting of that contrast can differ. Here we discuss each vowel pair in turn, followed by potential explanations.

For the [i] vs. [iː] contrast, Cairene listeners exhibit a lower durational threshold for identifying a vowel as [iː] than Najdi listeners. In other terms, a stimulus of 90 ms might already sound like "long [iː]" to a Cairene speaker, but the same stimulus may still sound like "short [i]" to a Najdi speaker. This pattern implies that Cairene listeners are somewhat more sensitive or attuned to short-duration cues for [iː] – they do not require as much absolute length to hear the vowel as long. Meanwhile, Najdi listeners seem to have a slightly stricter requirement: for them, the vowel needs to be held longer before it crosses the perceptual boundary into the [iː] category. One plausible reason is that Najdi Arabic may have relatively longer realizations of [iː] in everyday speech, or a larger ratio between [iː] and [i], such that Najdi listeners' internal standard for a "prototypical long [iː]" is higher. Conversely, Cairene Arabic could exhibit somewhat shorter long vowels (or a smaller ratio), conditioning listeners to accept a shorter vowel as long. Supporting this idea, previous research has noted that languages and dialects can differ in the exact phonetic realization of length contrasts (e.g. overall speech rate or vowel inherent duration). Our findings for [iː] fit this profile of dialectal calibration: both groups use duration, but what counts as "enough" duration for [iː] differs.

The [a] vs. [aː] contrast exhibited a statistically significant dialect difference in perceptual slope, even though the average boundary (around 101 ms) was similar across dialects. Cairene listeners showed a stronger durational sensitivity, meaning their responses shifted more sharply from "short" to "long" as duration increased. Najdi listeners exhibited shallower slopes, indicating a more gradual transition and reduced sensitivity to duration as a cue compared to Cairene speakers. This suggests that vowel quantity perception for [a] may be somewhat less categorical in Najdi Arabic, or that additional cues (such as vowel quality) are relied upon more variably by Najdi listeners. While the group-level averages were similar, the perceptual strategies differed measurably.

Finally, for the [u] vs. [uː] contrast, we did not find any meaningful difference between dialects. Both Najdi and Cairene listeners showed very similar perceptual boundaries for [uː], and identification curves that overlapped substantially. This suggests that, at least for this vowel, the two dialects are aligned in their use of duration cues. One reason might be that both dialects realize [u] vs. [uː] with comparable duration ratios and perhaps less influence of secondary cues. It is possible that high back vowels are realized more uniformly across these Arabic varieties, or simply that any minor differences were too small to detect given our sample size and stimulus set. Unlike [iː], where Cairene listeners identified long vowels earlier than Najdi listeners, [uː] showed no significant dialect difference in boundary location or slope. Both groups used duration similarly for this vowel. This asymmetry between [iː] and [uː] hints that dialect effects are vowel-specific.

4.3. Conclusion

This study examined how native speakers of Najdi and Cairene Arabic perceive the boundary between short and long vowels using duration-manipulated minimal pairs. For the [a] vs. [aː] contrast, both dialect groups exhibited nearly identical perceptual boundaries centered around 101 ms, suggesting that this low vowel length distinction is robust and stable across varieties.

In contrast, the [i] vs. [iː] pair showed a statistically significant dialectal difference, with Cairene listeners identifying vowels as long at shorter durations (84 ms) and displaying steeper perceptual slopes, indicating greater sensitivity to duration. Najdi listeners required longer durations to perceive [iː] and showed more gradual shifts in categorization. For [u] vs. [uː], perceptual boundaries were numerically different (Najdi = 100 ms; Cairene = 110 ms), but this difference was not statistically significant, and both groups exhibited similar categorical perception.

These findings contribute to our understanding of Arabic phonetics by demonstrating that while the long–short vowel contrast is clearly perceived in both dialects, the temporal threshold and perceptual slope vary depending on dialect and vowel quality. This supports the view that phonemic categories are not fixed, but are fine-tuned by dialectal experience and individual perceptual strategies.

Methodologically, the study demonstrates the usefulness of synthesized stimuli and mixed-effects modeling for isolating perceptual cues. At the same time, we acknowledge a key limitation of the design. All participants heard the stimuli in a single pre-randomized fixed sequence, rather than in multiple counterbalanced lists. Although this ensured that any order effects were consistent across participants and unlikely to create systematic dialect differences, it prevents us from fully controlling for potential sequence effects and limits the generalizability of the findings to other possible stimulus orders. Future studies should therefore implement multiple randomized lists distributed across participants to balance order influences more effectively.

Pedagogically, the findings highlight the importance of teaching fine durational contrasts in Arabic L2 contexts, especially for high vowels like [iː]. Beyond the classroom, future research should expand to additional dialects and explore how production–perception relationships shape phonemic boundaries. For example, do dialects that produce shorter long vowels also have listeners who accept shorter durations as long? Longitudinal studies could also investigate whether extended exposure to another dialect shifts perceptual boundaries.

References

Alammar, A. (2023). An acoustic study of Zilfaawi Arabic vowels. *Journal of Arabic Sciences and Humanities*, *2*(1), 1–34.

Al-Ani, S. (1970). *Arabic phonology: An acoustical and physiological investigation*. The Hague: Mouton.

Al-Bannai, A. (1995). Perception of vowel length in Arabic. *The Journal of the Acoustical Society of America, 97*(5_Supplement), 3419. https://doi.org/10.1121/1.412459

Algethami, G. (2023). *Acoustic characterization of the Najdi Arabic vowel system*. In *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)* (pp. 3384–3386). https://doi.org/10.31235/osf.io/ap3qc

Alghamdi, M. (1998). A spectrographic analysis of Arabic vowels: A cross-dialect study. *Journal of King Saud University, 10*, 3-24.

Almbark, R., & Hellmuth, S. (2015). Acoustic analysis of the Syrian vowel system. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.), *Proceedings of ICPhS 2015 University of Glasgow*. https://eprints.hud.ac.uk/id/eprint/28138/

Al-Tamimi, J. (2007). Static and dynamic cues in vowel production: A cross dialectal study in Jordanian and Moroccan Arabic. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 541-544). Saarbrücken, Germany.

Audacity Team. (2024). *Audacity(R): Free Audio Editor and Recorder* [Computer program]. Version 3.4.1. Retrieved November 20th, 2023, from https://www.audacityteam.org/download

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

Boersma, P. & Weenink, D. (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.27, retrieved 27 February 2024 from http://www.praat.org/

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *Journal of the Acoustical Society of America*, *144*(3), EL172–EL175. https://doi.org/10.1121/1.5052025

de Jong, K., & Zawaydeh, B. A. (2002). Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics, 30*(1), 53-75.

Embarki, M., Yeou, M., Guilleminot, C., & Al Maqtari, S. (2011). An acoustic study of coarticulation in modern standard Arabic and dialectal Arabic: Pharyngealized vs.

non-pharyngealized articulation. In Z. Hassan & B. Heselwood (Eds.), *Instrumental studies in Arabic phonetics* (pp. 141-164). Amsterdam: John Benjamins.

Fathi, H. M., & Qassim, Z. R. (2020). An acoustic study of the production of Iraqi Arabic vowels. *Journal of Al-Frahids Arts*, *12*, 692–704.

Flege, J. (1979). Phonetic interference in second language acquisition. *Unpublished doctoral dissertation*, Indiana University.

Francis, A. L., Ciocca, V., & Chit Ng, B.K. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics, 65*(7), 119–133. https://doi.org/10.3758/BF03194832

Ghazali, S., Hamdi, R., & Barkat, M. (2002). *Speech rhythm variation in Arabic dialects.* In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody 2002* (pp. 331–334). Aix-en-Provence, France: International Speech Communication Association (ISCA).

Haeri, N. (1997). *Sociolinguistic market of Cairo* (1st ed.). Routledge. https://doi.org/10.4324/9780203038611

Hamdi, R., Barkat-Defradas, M., Ferragne, E., & Pellegrino, F. (2004). *Speech timing and rhythmic structure in Arabic dialects: A comparison of two approaches.* In *Proceedings of Speech Prosody 2004* (pp. 131–134). Nara, Japan: International Speech Communication Association (ISCA). https://doi.org/10.21437/Interspeech.2004-49

Huthaily, K. (2003). Contrastive phonological analysis of Arabic and English. *Unpublished Master's Thesis*, The University of Montana.

Infillion. (2024). *Phonic Audio Survey* [Website]. https://www.phonic.ai/

Ingham, B. (1994). *Najdi Arabic: Central Arabian.* Amsterdam: John Benjamins. https://doi.org/10.1075/loall.1

Jasmin, K., Sun, H., & Tierney, A. (2021). Effects of language experience on domain-general perceptual strategies. *Cognition*, *206*, 104481. https://doi.org/10.1016/j.cognition.2020.104481

Kalaldeh, R. (2018). Acoustic analysis of vowels in modern standard Arabic. *International Journal of Arabic-English Studies, 18*(2), 23-48. https://doi.org/10.33806/ijaes2000.18.1.2

Kent, R., & Read, C. (2002). *The acoustic analysis of speech* (2nd ed.). Albany, NY: Thomson Learning.

Khattab, G. (2002). VOT in English and Arabic bilingual and monolingual children. In D. Parkinson & E. Benmamoun (Eds.), *Perspectives on Arabic Linguistics XIII-XIV* (pp. 1-38). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.230.03kha

Kogan, V. V., & Mora, J. C. (2022). The effects of individual differences in native perception on discrimination of a novel non-native contrast. *Laboratory Phonology, 24*(1). https://doi.org/10.16995/labphon.6431

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics* (6th ed.). Boston, MA: Wadsworth, Cengage Learning.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

McCarus, E. (2008). Modern Standard Arabic. In K. Versteegh, M. Eid, A. Elgibali, M. Woidich, & A. Zaborski (Eds.), *Encyclopedia of Arabic Language and Linguistics* (Vol. 3, pp. 238–262). Brill.

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1609–1631. https://doi.org/10.1037/a0011747

Mitleb, F. M. (1984). Vowel length contrast in Arabic and English: A spectrographic test. *Journal of Phonetics, 12*(3), 229-235. https://doi.org/10.1016/S0095-4470(19)30879-4

Newman, D., & Verhoeven, J. (2002). Frequency analysis of Arabic vowels in connected speech. *Antwerp Papers in Linguistics, 100*, 77-86.

Norlin, K. (1987). A phonetic study of emphasis and vowels in Egyptian Arabic. *Lund University Department of Linguistics Working Papers, 30*, 1-119.

Philippa, K., Philippa, M., & Roeleveld, A. (2017). Monophthongization of ay/ai and aw/au: A Comparison between Arabic and Germanic Dialects. *Amsterdamer Beiträge zur älteren Germanistik*, 77(3-4), 616-636. https://doi.org/10.1163/18756719-12340095

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

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81–110. https://doi.org/10.1037/0033-2909.92.1.81

Ryding, K. (2005). *A reference grammar of modern standard Arabic*. Cambridge: Cambridge University Press.

Theodore, R. M., Myers, E. B., & Lomibao, J. A. (2015). Talker-specific influences on phonetic category structure. *Journal of the Acoustical Society of America, 138*(2), 1068–1078. https://doi.org/10.1121/1.4927489

Tsukada, K. (2012). Comparison of native versus nonnative perception of vowel length contrasts in Arabic and Japanese. *Applied Psycholinguistics*, *33*(3), 501–516. https://doi.org/10.1017/S0142716411000452

Watson, J.C.E. (2002). *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Yu, A. (2021). Toward an individual-difference perspective on phonologization. *Glossa: A Journal of General Linguistics, 6*(1), 14. https://doi.org/10.5334/gjgl.661

Author Contribution

Abdullah Alfaifi is the sole author of this paper and was responsible for all aspects of its creation, including the conception, design, data collection, analysis, interpretation, and manuscript preparation.

Ethical Approval

This study was reviewed and approved by the Scientific Research Unit at the College of Arts and Letters, University of Bisha (Institutional Review Board). All research procedures involving human participants were conducted in accordance with the Declaration of Helsinki and the institutional guidelines and regulations of the University of Bisha. Ethical approval for the study procedures (recruitment, informed consent, data collection, storage, and analysis) was granted under IRB number UB.24.1.8.ALSRU.1 on 8 January 2024. The approval covered all aspects of the research protocol reported in this manuscript. The study involved no more than minimal risk, ensured equitable selection of participants, and included appropriate measures to protect confidentiality and privacy.

Informed Consent

This study involved adult participants (18 years or older), and written informed consent was obtained from all individuals prior to participation. Written consent was obtained electronically by the principal investigator before participants started the data-collection sessions, between 2 February 2024 and 11 March 2024. Participants were provided with a written information sheet that explained that the project was a research study, described its aims and procedures, outlined any potential risks and benefits (which were minimal), and confirmed that participation was voluntary and that refusal or withdrawal would carry no negative consequences. The information sheet also explained how their data would be handled, including that their responses would be anonymized for analysis, that their anonymity would be assured in any reports or publications, and that any identifiable personal data would be kept confidential, and it stated that anonymized data may be submitted to an online repository and used for journal publication for research purposes only. Participants were informed of their right to withdraw from the study at any time and, where feasible, to request removal of their data after collection. By electronically signing the written consent form, participants indicated that they had read and understood the information provided and agreed to take part under these conditions.

Competing interests

The author declares no competing interests.

Data Availability

The datasets generated during and/or analyzed during the current study are available in the Figshare repository, at the following link: https://figshare.com/s/ae64d72461e07bc08050.

Acknowledgement

The author is grateful to the Deanship of Graduate Studies and Scientific Research at the University of Bisha for supporting this work through the Fast-Track Research Support Program.

Figure Legends

Figure 1 Legends:

Al-Ani (Iraqi)
Mitleb (Jordanian)
Kalaldeh (Jordanian)
Fathi and Qassim (Iraqi)
Algethami (Najdi)
Alghamdi (Saudi)
Alammar (Zilfaawi)
Norlin (Egyptian)
Current Study (Saudi)

Figures 2, 3, and 4:

Cairene

Najdi

Figure 5 Legends:

/i/
/i:/
Cairene
Najdi

Figure 6 Legends:

/a/
/a:/
Cairene
Najdi

Figure 7 Legends:

/u/
/u:/
Cairene
Najdi