

Research Paper

Statistical Metrics for the Temporal Acoustics of Durationally Contrastive Vocalics: A Proposal Tested with Data from Arabic and Japanese

Yahya ALDHOLMI

*Department of Linguistics, College of Language Sciences
King Saud University*P.O. Box 87907, Riyadh 11652, Saudi Arabia; e-mail: yaldholmi@ksu.edu.sa*(received June 4, 2023; accepted November 21, 2023; published online February 29, 2024)*

Previous research has utilized the *duration ratio* and occasionally the *duration difference* as single-value metrics to measure and compare the temporal acoustics of durationally contrastive vocalics (short vs. long vowels), which allow researchers to reduce two values (short and long) to one, but express a *relationship* instead of representing the vocalic duration values directly. The duration ratio may even be misleading when comparing two languages or dialects, as it is possible to exhibit a similar ratio but differ in durational acoustics, or vice versa. The current study proposes two alternative statistical metrics: a duration metric and a difference metric. The duration metric is an intermediate (mean-like) value between the duration of the short and long vocalics, and the difference metric is a \pm value that can be added to or subtracted from the duration metric to obtain the duration of long or short vocalics. We conduct a production experiment on Arabic and Japanese vocalics and analyze the data using both traditional measures and the proposed metrics. The findings show that the proposed metrics better predict the language from which the vocalic duration values were obtained. Such results suggest that the proposed metrics are better candidates for measuring and comparing the temporal acoustics of vocalics.

Keywords: vocalic duration; duration difference; duration ratio; duration contrast; metrics.



Copyright © The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The durationally contrastive vocalics¹ (i.e., short and long vowels) in natural and synthetic speech have been investigated in research on both native vs. non-native and normal vs. impaired production and perception in languages with durationally contrastive vocalics, such as Arabic. Most of the early research took a theoretical perspective and focused on the observation that, as in other Semitic languages such as Hebrew, a vocalic duration is inherently phonemically contrastive in Arabic (e.g., HARRIS, 1942; CANTINEAU, 1956; FERGUSON, 1957; COWAN, 1970). Subsequent studies employed experimental or obser-

vational methods to examine the temporal acoustics of Arabic vocalics acoustically, perceptually, or acoustically-perceptually for both normal and impaired speakers (e.g., AL-ANI, 1970; MITLEB, 1984; ALGHAMDI, 1998; TSUKADA, 2009; 2011; 2012a; 2012b; 2013; AMIR *et al.*, 2014; KALALDEH, 2018; ALDHOLMI *et al.*, 2021). This phenomenon has been explored in Arabic varieties including Iraqi (HASSAN, 1981), Jordanian (MITLEB, 1984), Saudi, Sudanese, Egyptian (e.g., ALGHAMDI, 1998), Moroccan (AL-TAMIMI, BARKAT-DEFRADAS, 2003), Libyan (AHMED, 2008), Cairene (KOTBY *et al.*, 2011), Palestinian (SAADAH, 2011), Syrian (ALMBARK, HELLMUTH, 2015), and Modern Standard Arabic (MSA) (KALALDEH, 2018; ALDHOLMI, 2022), as well as other varieties spoken in non-Arab regions such as the Kasseem/Kfar Barra/Jaljulia dialects in Israel (AMIR *et al.*, 2012), Galilee and the Triangle (Muthallath) regions in Israel (AMIR *et al.*, 2014), and Malaysia (ABD ALMISREB *et al.*, 2016).

¹Note that “vocalics” and “temporal acoustics” are used as two general terms throughout the paper. The former refers to vowels regardless of length and type, and the latter refers to all durational aspects regardless of tool (e.g., duration difference, duration ratio...). Specific terms are used instead when applicable.

Most of the aforementioned studies have documented and characterized the temporal acoustics of Arabic vocalics and the short-to-long duration ratio thereof both interdialectally and cross-dialectally. Other studies sought to compare and contrast the Arabic native production of vocalic duration with that of other Semitic languages that exhibit vocalic durational contrast, such as Hebrew (e.g., AMIR *et al.*, 2012), and non-Semitic languages that either feature durational contrast, such as Japanese and Thai (TSUKADA, 2009), or do not, such as English (MITLEB, 1984). Other studies examined non-native production and perception of vocalic duration by speakers of Arabic as a second language (L2) whose native language (L1) was Japanese (TSUKADA, 2012a; 2012b), English (FLEGE, PORT, 1981; LABABIDI, PARK, 2014), Korean (HONG, SARMAH, 2009), or Hebrew (ZALTZ, SEGAL, 2021).

To this end, researchers have utilized both the *duration difference* and the *duration ratio* in dialect and language comparisons. For instance, some Arabic dialects, specifically Jordanian, have been reported to have a short-to-long duration ratio of 0.65 (MITLEB, 1984), while others have demonstrated a considerably smaller ratio, for example, 0.39 in Palestinian (SAADAH, 2011). This discrepancy may not necessarily be due to the use of different stimuli or methods but rather due to actual interdialectal variations, as the duration ratio does *not* truly convey much about vocalic duration in one dialect or another, nor does it allow for a clear cross-dialect or cross-language comparison.

The duration ratio does not directly express vocalic duration in time units (e.g., ms); instead, it shows only how large or small a value is in relation to another value. That is, the duration ratio of 100 to 200 ms is exactly the same as that of 200 to 400 ms (0.5 in both), which makes this measure unhelpful when comparing two language varieties. The duration difference only shows the quantitative relationship between two given vocalics as short in duration and long in duration, rather than reflecting the actual duration acoustics of the segments under investigation. In addition, sometimes, the short version becomes too long or the long version becomes too short, which results in a negative duration difference value when calculating the difference for each minimal pair (e.g., 70 – 100 = –30 ms). Considering that individual value (rather than the overall mean difference), the difference between two positive values should be any nonnegative value (including zero), but a negative duration ratio value will be uninterpretable. There are a few potential solutions to this particular issue, but each has its own problems. For instance, we could transform and normalize data to be at or above zero, but this would increase the overall mean duration difference.

Hence, this study proposes two statistical metrics to allow for *direct* comparison between different vari-

eties in terms of vocalic duration. The first section provides background and describes the two metrics, the duration metric and the difference metric, that can be used instead of the duration difference or duration ratio values reported in previous studies. The two metrics are illustrated using available data from relevant literature. In the second section, a production experiment is conducted to test the two alternative metrics, using data from Arabic and Japanese as two languages that have been repeatedly compared and contrasted in the literature (e.g., TSUKADA, 2009) because they share similar durationally contrastive vocalics (e.g., ALD-HOLMI, 2022).

2. Traditional measures vs. proposed metrics

2.1. Traditional measures

As reported in some previous studies, the traditional method for obtaining a short-long duration ratio divides the mean duration of the short vowels by that of the long vowels. For instance, MITLEB (1984) reported a ratio of 0.65, calculated as (Eq. (1)):

$$\begin{aligned} \text{ratio} &= \frac{\text{mean short vowels}}{\text{mean long vowels}}, \\ 0.65 &= \frac{83 \text{ ms}}{128 \text{ ms}}. \end{aligned} \quad (1)$$

In some cases, the duration difference is reported instead of the duration ratio. The duration difference is simply the difference between the mean duration of the long vowels and that of the short vowels, as shown in Eq. (2). MITLEB (1984) reported a duration difference of 45 ms:

$$\begin{aligned} \text{difference} &= \text{mean long vowels} - \text{mean short vowels}, \\ 45 &= 128 \text{ ms} - 83 \text{ ms}. \end{aligned} \quad (2)$$

The duration ratio is sometimes reported in qualitative rather than numerical form. For instance, TSUKADA (2011) stated that “long [Arabic] vowels are twice as long as their short counterparts” (p. 989), while “long Japanese vowels tend to be more than twice as long as their short counterparts” (p. 990). Regardless, both the duration ratio and the duration difference depend on the range of the two values, specifically the mean short vowel duration and the mean long vowel duration, which on their own are insufficient to precisely quantify the vocalic duration in a given dialect. For example, suppose that in one Arabic dialect the mean duration of two short vocalics (65 + 75 ms) is 70 ms while the mean duration of two long vocalics (165 + 175 ms) is 170 ms, and in another Arabic dialect the mean duration of two short vocalics (115 + 125 ms) is 120 ms while the mean duration of two long vocalics (285 + 295 ms) is 290 ms. In both scenarios, the ratio is

approximately 0.41, but the difference is 100 ms in the first and 170 ms in the second. Using a duration ratio value makes the two dialects seem similar, but the range of the values and the difference do not.

2.2. Proposed metrics

The proposed duration difference is similar to the formant spacing – compact-diffuse (C-D) measure used in some studies in which the first formant ($F1$, a smaller value) is subtracted from the second formant ($F2$, a larger value) (e.g., [BLOMGREN *et al.*, 1998](#); [KENT, VORPERIAN, 2018](#)). Although computing the C-D value has a different purpose, namely, to evaluate tongue elevation (e.g., [JAKOBSON *et al.*, 1963](#)), it reduces the two values into a single value that can be used for statistical description and inference. Another formant spacing value is the so-called grave-acute (G-A) measure ([KENT, VORPERIAN, 2018](#)), which describes tongue advancement ([JAKOBSON *et al.*, 1963](#); [BLOMGREN *et al.*, 1998](#)). The G-A value has been computed according to Eq. (3), where X = each individual vocalic, and n = the total number of data points (vocalics):

$$\text{G-A} = \frac{1}{n} \sum \frac{(F1 \text{ of vocalic } X + F2 \text{ of vocalic } X)}{2}. \quad (3)$$

This method can form the basis of a new, alternative metric that can be used to describe the vocalic duration and the vowel difference in languages where the vocalic duration is contrastive. The proposed metric can be calculated by the given equation (Eq. (4)):

$$\text{duration metric} = \frac{1}{n} \sum \frac{(\text{short} + \text{long})}{2}. \quad (4)$$

The output provides us with one value that lies between the original value of the short vowel and that of the long vowel, but it should better inform us about how short or long the two contrastive vocalics are in a given dialect or language. To illustrate this, consider the previous two scenarios, calculated as (a) and (b) for convenience. Note that we treat the mean durations as single data points for two individual vocalics:

$$\text{a) duration metric} = \frac{1}{2} \left(\frac{65+165}{2} + \frac{75+175}{2} \right) = 120 \text{ ms,}$$

$$\text{b) duration metric} = \frac{1}{2} \left(\frac{115+285}{2} + \frac{125+295}{2} \right) = 205 \text{ ms.}$$

The two obtained values indicate that the first dialect has notably smaller short and long vocalic durations than the second dialect. In other words, the overall duration of vocalics in the second dialect is approximately 42% longer than that in the first dialect. Neither the duration ratio, which is identical in both dialects (0.41), nor the duration difference, which always depends on the distance between the short and long vowel durations, will provide a unified metric that

allows for a direct comparison between the two dialects or languages. Nevertheless, the proposed duration metric here still does not show how far the duration value is from the original short and long durations. Hence, one further step is needed, which is to calculate the difference metric (Eqs. (5) and (6)):

$$\begin{aligned} \text{difference metric} &= \text{duration metric} \\ &\pm (\text{duration metric} - \text{short vocalic}) \end{aligned} \quad (5)$$

or

$$\begin{aligned} \text{difference metric} &= \text{duration metric} \\ &\pm (\text{long vocalic} - \text{duration metric}). \end{aligned} \quad (6)$$

Note that Eqs. (5) and (6) provide the exact value. Consider the vowel difference computed for the aforementioned scenarios:

$$\text{a) difference metric} = 120 \text{ ms} \pm \left(\frac{65+75}{2} = 70 \right) = 50 \text{ ms,}$$

$$\text{b) difference metric} = 205 \text{ ms} \pm \left(\frac{115+125}{2} = 120 \right) = 85 \text{ ms.}$$

The \pm value is the difference metric that we can add to or subtract from the vowel duration metric to obtain the duration of the short vocalic or of the long vocalic. In the first scenario, $120 \text{ ms} \pm 50 \text{ ms} = 70$ or 170 to yield the durations of the short and long vocalics, respectively. The difference metric shows that the difference between the short and long vowels is smaller in the first dialect than in the second dialect. The same applies to the second scenario.

Thus, the duration metric provides us with one value that represents both short and long vocalics. This cannot be achieved via the traditional duration difference (where the short duration is subtracted from the long duration) because the short and long vowels can have large values (e.g., 200 and 250 ms, respectively), but the duration difference, which will be 50 ms in this case, cannot be used to calculate the exact duration of either vocalic. Similarly, two smaller values for short and long vowels (e.g., 50 and 110 ms) can have a larger duration difference, calculated here as 60 ms, but this value also indicates nothing about the duration of the short and long vocalics. The proposed duration metric *does* provide information about how long the short and long vocalics are. To illustrate this with a real-world example, we analyze data from [TSUKADA \(2011\)](#).

2.3. An example from Arabic and Japanese

The short vocalic /a/ in trial 1 has a relatively small duration ratio (0.37) compared to its long counterpart, which is below the lowest value reported in the literature on Arabic (0.39), while the short vocalic /u/ in trial 1 has a relatively larger duration ratio (0.51), which is above the frequently reported range (39–45) in the literature (e.g., [TSUKADA, 2011](#)). Nevertheless,

Table 1. Duration [in ms], difference [in ms], and ratio in Arabic (data adopted from TSUKADA, 2011, p. 992).

| Trial/item | Long vowels | Long vowels duration | Short vowels | Short vowels duration | Duration ratio | Duration difference | Short + long | Duration metric | Difference metric |
|-------------|-------------|----------------------|--------------|-----------------------|----------------|---------------------|---------------|-----------------|-------------------|
| 1 | /aa/ | 307.00 | /a/ | 114.00 | 0.37 | 193.00 | 421.00 | 210.50 | ±96.50 |
| 2 | /aa/ | 190.00 | /a/ | 96.00 | 0.51 | 94.00 | 286.00 | 143.00 | ±47.00 |
| 1 | /ii/ | 235.00 | /i/ | 103.00 | 0.44 | 132.00 | 338.00 | 169.00 | ±66.00 |
| 2 | /ii/ | 173.00 | /i/ | 87.00 | 0.50 | 86.00 | 260.00 | 130.00 | ±43.00 |
| 1 | /uu/ | 212.00 | /u/ | 109.00 | 0.51 | 103.00 | 321.00 | 160.50 | ±51.50 |
| 2 | /uu/ | 183.00 | /u/ | 88.00 | 0.48 | 95.00 | 271.00 | 135.50 | ±47.50 |
| Mean | All | 216.67 | All | 99.50 | 0.47 | 117.17 | 316.17 | 158.08 | ±58.58 |

Table 2. Duration [in ms], difference [in ms], and ratio in Japanese (data adopted from TSUKADA, 2011, p. 992).

| Trial/item | Long vowels | Long vowels duration | Short vowels | Short vowels duration | Duration ratio | Duration difference | Short + long | Duration metric | Difference metric |
|-------------|-------------|----------------------|--------------|-----------------------|----------------|---------------------|---------------|-----------------|-------------------|
| 1 | /aa/ | 190.00 | /a/ | 67.00 | 0.35 | 123.00 | 257.00 | 128.50 | ±61.50 |
| 2 | /aa/ | 166.00 | /a/ | 65.00 | 0.39 | 101.00 | 231.00 | 115.50 | ±50.50 |
| 1 | /ee/ | 182.00 | /e/ | 88.00 | 0.48 | 94.00 | 270.00 | 135.00 | ±47.00 |
| 2 | /ee/ | 159.00 | /e/ | 79.00 | 0.50 | 80.00 | 238.00 | 119.00 | ±40.00 |
| 1 | /ii/ | 195.00 | /i/ | 80.00 | 0.41 | 115.00 | 275.00 | 137.50 | ±57.50 |
| 2 | /ii/ | 176.00 | /i/ | 69.00 | 0.39 | 107.00 | 245.00 | 122.50 | ±53.50 |
| 1 | /oo/ | 181.00 | /o/ | 84.00 | 0.46 | 97.00 | 265.00 | 132.50 | ±48.50 |
| 2 | /oo/ | 156.00 | /o/ | 75.00 | 0.48 | 81.00 | 231.00 | 115.50 | ±40.50 |
| 1 | /uu/ | 179.00 | /u/ | 59.00 | 0.33 | 120.00 | 238.00 | 119.00 | ±60.00 |
| 2 | /uu/ | 152.00 | /u/ | 61.00 | 0.40 | 91.00 | 213.00 | 106.50 | ±45.50 |
| Mean | All | 173.60 | All | 72.70 | 0.42 | 100.90 | 246.30 | 123.15 | ±50.45 |

the actual durations of both short vocalics are 114 ms and 109 ms, respectively, which exhibit only a negligible difference ($114 - 109 = 5$ ms). The overall duration ratio (0.47) and the overall duration difference (117.17 ms) do not indicate the actual magnitude of the durations of short and long vocalics in Arabic. In contrast, the duration metric *does* show that Arabic short vocalics are generally shorter and Arabic long vocalics longer than 158.08 ms and that the distance between short or long vowels and this metric value is ±58.58 ms overall.

Table 2 shows a duration metric in Japanese of 123.15 ms and a difference metric of ±50.45 ms, suggesting that Japanese vocalics generally tend to be shorter than Arabic vocalics (123.15 vs. 158.08 ms, respectively) and that the difference (not the ratio) between short and long vocalics in Japanese is smaller than that in Arabic (50.45 vs. 58.58 ms, respectively). Inspection of the means for both Arabic and Japanese short and long vowels supports this conclusion.

3. Duration metric and difference metric tested: An experiment

The current experiment utilizes the proposed duration and difference metrics for statistical analysis and

compares them with the traditional duration difference and the duration ratio measures in Arabic and Japanese.

3.1. Methodology

The stimuli for this study consist of 18 MSA CVCVC vs. CV:CVC words and 18 Japanese CVCV vs. CV:CV words. The Arabic items selected for this experiment were inspired by (but not taken from) HASSAN (2002), while the Japanese items were selected from TSUKADA (2012b). The target vocalic in the stimuli from both languages was the first rather than the second/final vocalic because the final vocalic is subject to certain phonological processes such as shortening and lengthening (see ALDHOLMI, 2022). Following the same protocol by ALDHOLMI (2022), 22 male and 18 female native speakers of Arabic ($n = 40$) produced the items using an Arabic carrier sentence (/ʔanaa ʔaktubu ___ ʔaidʕan/ “I write ___ as well”). The Arabic participants spoke different Arabic dialects, including Najdi (Qassimi, $n = 8$), Hijazi (Jeddah and Madinah, $n = 11$), Southern (Faifa and Abha, $n = 12$), and Northern (Northern Borders, $n = 9$) dialects. The Arabic participants met face-to-face with the experimenter or other linguists who had volunteered to help the researcher collect the data at different Saudi institutions.

Twenty-four male and 16² female native speakers of Japanese ($n = 40$; the initial sample comprised 41 participants, but one was excluded for unclear speech) produced the Japanese items using a Japanese carrier sentence adopted from TSUKADA (2011, p. 991) (/tsugiwa ___ to iimasu/ “Next I say the word ___”) and performed the task entirely online (using Phonic.ai, 2023). Approximately half of the Japanese participants ($n = 19$) came from Osaka, while the rest did not specify their origin. The target vocalics were isolated from the adjacent consonants by the experimenter, using both visual and auditory judgements for all items. Praat (BOERSMA, WEENINK, 2021) was used for segmenting and marking the boundaries of segments for all items.

3.2. Results

As shown in Figs. 1 and 2 and Table 3, Arabic has larger means and greater variability for both short and

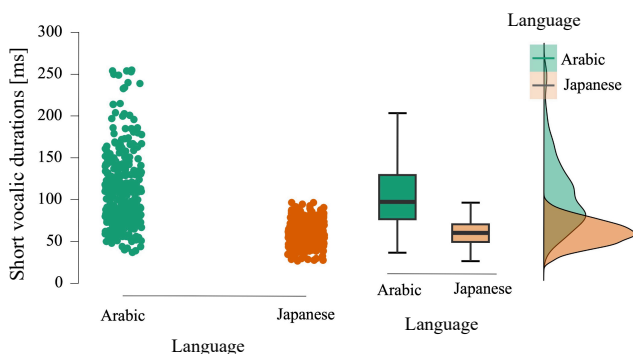


Fig. 1. Short vocalic durations in Arabic and Japanese.

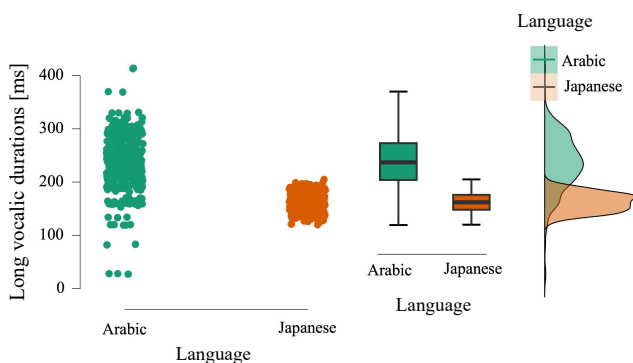


Fig. 2. Long vocalic durations in Arabic and Japanese.

²It would have been desirable to maintain gender balance for both Arabic and Japanese speakers, had the Japanese speakers been as accessible to the researcher as the Arabic speakers were. Nonetheless, an attempt was made to maintain a similar female-to-male ratio in both groups, although previous studies did not always have gender-balanced groups. For instance, Tsukada had 7 Arabic speakers (4 male and 3 female) in some studies (e.g., TSUKADA, 2011) and 9 Arabic speakers (6 male and 3 female) in some other studies (e.g., TSUKADA, 2012a). It should also be recalled that speakers maintain duration distinction in languages that exhibit durationally contrastive vocalics such as Hebrew regardless of gender (e.g., AMIR *et al.*, 2012).

Table 3. Short and long vocalic means [ms], SDs, minimums and maximums of durations in Arabic and Japanese.

| Length | Language | Mean | SD | Minimum | Maximum |
|--------|----------|--------|-------|---------|---------|
| Short | Arabic | 108.95 | 42.43 | 50.00 | 254.63 |
| | Japanese | 60.51 | 14.56 | 27.00 | 97.00 |
| Long | Arabic | 237.11 | 52.55 | 27.00 | 413.94 |
| | Japanese | 161.88 | 17.37 | 120.00 | 205.00 |

long vocalics (short: $M = 108.95$, $SD = 42.43$; long: $M = 237.11$, $SD = 52.55$) than Japanese (short: $M = 60.51$, $SD = 14.56$; long: $M = 161.88$, $SD = 17.37$).

The repeated-measures ANOVA with the vocalic length (short vs. long) as a within-subject factor and language as a between-subject factor was performed to test mean differences. As detailed in Table 4, the test provided evidence for a statistically significant difference between short and long vocalics with a very large effect size, $F(1, 78) = 2047.16$, $p < 0.001$, $\omega^2 = 0.86$, and between Arabic and Japanese, also with a large effect size, $F(1, 78) = 182.51$, $p < 0.001$, $\omega^2 = 0.53$. There was also a statistically significant interaction between the two factors with an intermediate effect, $F(1, 78) = 27.89$, $p < 0.001$, $\omega^2 = 0.07$.

Table 4. Repeated-measures ANOVA results.

| Factors | df | F | p | ω^2 |
|---------------------------|----|---------|-------|------------|
| Vocalic length | 1 | 2047.16 | <.001 | 0.86 |
| Language | 1 | 182.51 | <.001 | 0.53 |
| Vocalic length * language | 1 | 27.89 | <.001 | 0.07 |

Thus, we have strong evidence that Arabic and Japanese differ significantly in terms of duration for both short and long vocalics and that, within each language, short vocalics are shorter than their counterparts. Table 5 presents the duration differences and duration ratios for both languages. The duration difference for Arabic (128.16 ms) and for Japanese (101.82 ms) and the duration ratio for Arabic (0.48) and Japanese (0.37) are similar to those calculated and obtained from the data provided in (TSUKADA, 2011). Hence, the duration difference may be misinterpreted as indicative of an overall similarity between the vocalic duration in Arabic and Japanese, which is not precisely the case.

Table 5. Duration differences and duration ratios in Arabic and Japanese.

| | Language | Mean | SD | Minimum | Maximum |
|--------------------------|----------|--------|-------|----------|---------|
| Duration difference [ms] | Arabic | 128.16 | 54.20 | -167.47* | 329.84 |
| | Japanese | 101.82 | 19.45 | 45.00 | 294.00 |
| Duration ratio | Arabic | 0.48 | 0.29 | 0.20 | 3.04 |
| | Japanese | 0.37 | 0.06 | 0.20 | 0.54 |

* Note that the negative value is one of the issues caused by using the duration difference, as explained in the introduction.

Now consider both the proposed duration metric and the difference metric in Table 6. The duration metric for Arabic (173.03 ms) was substantially larger than that for Japanese (111.20 ms). Likewise, the difference metric for Arabic (64.08 ms) was considerably greater than that for Japanese (51.30 ms).

Table 6. Duration metrics and difference metrics in Arabic and Japanese.

| | Language | Mean | SD | Minimum | Maximum |
|------------------------|----------|--------|-------|---------|---------|
| Duration metric [ms] | Arabic | 173.03 | 39.32 | 42.50 | 283.54 |
| | Japanese | 111.20 | 14.85 | 74.50 | 145.50 |
| Difference metric [ms] | Arabic | 64.08 | 42.43 | 50.00 | 254.63 |
| | Japanese | 51.30 | 28.31 | -52.00 | 259.00 |

Thus, based on the aforementioned data, we observe that the duration metric and the difference metric better represent the vocalic duration facts in both languages. The values are re-reported side-by-side in Table 7, which arguably illustrates how the substantial dissimilarity between Arabic and Japanese and between short and long vocalics is reflected more clearly in the duration metric and the difference metric than in the duration difference and the duration ratio. To support this claim, an inverse regression was performed to test which of the four variables (duration difference, duration ratio, duration metric, or difference metric) would most accurately predict the language. We first compare the duration difference and the duration metric, as these two are similar; both inform us about the actual duration of the short vs. long vowels. Next, we compare the duration ratio and the difference metric, as these two are also similar; both inform

us about the relationship between two values. Despite the similarity in purpose between the members of each group, the difference metric and duration metric both have the added benefit of being able to inform us about the mean vocalic duration measures as well.

We fitted an inverse binary logistic regression model, first using the *duration difference* as a predictor variable and the *language* as a predicted variable. The results indicated a significant improvement in fit relative to an intercept-only model, $\chi^2(1) = 79.58$, $p < .001$, and that the duration difference was a statistically significant predictor of language, $\chi^2(1) = 60.60$, $p < .001$. Table 8 shows the -2 log-likelihood ($-2LL$) and the pseudo- R^2 values of the first model (model 1). As shown, in order from the largest pseudo- R^2 value to the smallest, the Nagelkerke R^2 , Tjur R^2 , Cox and Snell R^2 , and McFadden R^2 exhibited relatively similar, low values. These values become important later when we compare with another predictor variable.

Table 9 shows that the sensitivity of the model was 78.60%, the specificity of the model was 66.70%, and the overall accuracy was 72.60%.

The model was re-fitted using the *duration metric* value as a predictor variable. The results again showed a significant improvement in fit for the second model (model 2) relative to an intercept-only model, $\chi^2(1) = 572.06$, $p < .001$, and that duration metric was a statistically significant predictor of the language, $\chi^2(1) = 159.56$, $p < .001$. When the duration difference was used as a predictor for the language, the $-2LL$ value was lower while the pseudo- R^2 values (Table 10) were higher than those obtained in the previous model, demonstrating the development of better fit in model 2.

Table 7. Vocalic durations, duration differences, duration ratios, duration metrics, and difference metrics in Arabic and Japanese.

| | Short duration [ms] | | Long duration [ms] | | Duration difference [ms] | | Duration ratio | | Duration metric [ms] | | Difference metric [ms] | |
|------|---------------------|----------|--------------------|----------|--------------------------|----------|----------------|----------|----------------------|----------|------------------------|----------|
| | Arabic | Japanese | Arabic | Japanese | Arabic | Japanese | Arabic | Japanese | Arabic | Japanese | Arabic | Japanese |
| Mean | 108.95 | 60.51 | 237.11 | 161.88 | 128.16 | 101.82 | 0.48 | 0.37 | 173.03 | 111.20 | 64.08 | 51.30 |
| SD | 42.43 | 14.56 | 52.55 | 17.37 | 54.20 | 19.45 | 0.29 | 0.06 | 39.32 | 14.85 | 42.43 | 28.31 |

Table 8. The $-2LL$ and pseudo- R^2 values for model 1.

| | χ^2 | df | p | $-2LL$ | Nagelkerke R^2 | Tjur R^2 | Cox and Snell R^2 | McFadden R^2 |
|---------------------|----------|------|-------|--------|------------------|------------|---------------------|----------------|
| Model | 79.58 | 1 | <.001 | 918.55 | 0.14 | 0.12 | 0.10 | 0.08 |
| Duration difference | 60.60 | 1 | <.001 | | | | | |

Table 9. Confusion matrix (sensitivity and specificity rates) and accuracy rate of model 1.

| Observed | | Predicted | | |
|-------------------------------|--------|-----------|-------|--------------------|
| | | Language | | Percentage correct |
| Language | Arabic | Japanese | | |
| | Arabic | 240 | 120 | 66.70 |
| Japanese | 77 | 283 | 78.60 | |
| Overall percentage (accuracy) | | | 72.60 | |

Table 10. The -2LL and pseudo- R^2 values for model 2.

| | χ^2 | df | p | -2LL | Nagelkerke R^2 | Tjur R^2 | McFadden R^2 | Cox and Snell R^2 |
|-----------------|----------|----|-------|--------|------------------|------------|----------------|---------------------|
| Model | 572.06 | 1 | <.001 | 426.06 | 0.73 | 0.65 | 0.57 | 0.54 |
| Duration metric | 159.56 | 1 | <.001 | | | | | |

As shown in Table 11, the sensitivity (85.30%), specificity (92.80%), and overall accuracy (89.00%) all improved in model 2.

Thus, all indicators demonstrated that the duration metric proposed in the current study is a better alternative to the duration difference used in previous studies. We compare the two other indicators (the duration ratio vs. the difference metric), following the same steps used in comparing the duration difference and the duration metric.

An inverse binary logistic regression model (model 3) was performed with the *duration ratio* as a predictor and the *language* as a predicted variable. The output showed that, compared to an intercept-only model, model 3 demonstrated a significant improvement in fit, $\chi^2(1) = 101.92$, $p < .001$, and that duration ratio was a statistically significant predictor of the language, $\chi^2(1) = 68.31$, $p < .001$. The -2LL value (896.20) and the pseudo- R^2 values (Cox and Snell $R^2 = 0.17$, Nagelkerke $R^2 = 0.13$, Tjur $R^2 = 0.12$, and McFadden $R^2 = 0.10$) were very similar (Table 12)

to those obtained when the duration difference was used as an indicator.

The sensitivity (59.7%), specificity (70.8%), and overall accuracy (65.3%) of the model, as shown in Table 13, indicated that this model exhibited poor sensitivity and slightly poor overall accuracy.

Running the model again with the *difference metric* as a predictor, model 4 showed a significant improvement in fit relative to the intercept-only model, $\chi^2(1) = 537.29$, $p < .001$. It also indicated the difference metric as a statistically significant predictor of the language, $\chi^2(1) = 167.56$, $p < .001$. The -2LL value (460.84) and the pseudo- R^2 values (Nagelkerke $R^2 = 0.71$, Tjur $R^2 = 0.60$, Cox and Snell $R^2 = 0.53$, and McFadden $R^2 = 0.52$) were highly similar (Table 14) to those obtained when using the duration metric as an indicator.

The sensitivity (87.2%), specificity (84.2%), and overall accuracy (85.7%) of the model, as shown in Table 15, were notably higher than those in the previous model and indicated good fit. Thus, running the model

Table 11. Confusion matrix (sensitivity and specificity rates) and accuracy rate of model 2.

| Observed | | Predicted | | |
|-------------------------------|----------|-----------|----------|--------------------|
| | | Language | | Percentage correct |
| | | Arabic | Japanese | |
| Language | Arabic | 307 | 53 | 85.30 |
| | Japanese | 26 | 334 | 92.80 |
| Overall percentage (accuracy) | | | | 89.00 |

Table 12. The -2LL and pseudo- R^2 values for model 3.

| | χ^2 | df | p | -2LL | Cox and Snell R^2 | Nagelkerke R^2 | Tjur R^2 | McFadden R^2 |
|----------------|----------|----|-------|--------|---------------------|------------------|------------|----------------|
| Model | 101.92 | 1 | <.001 | 896.20 | 0.17 | 0.13 | 0.12 | 0.10 |
| Duration ratio | 68.31 | 1 | <.001 | | | | | |

Table 13. Confusion matrix (sensitivity and specificity rates) and accuracy rate of model 3.

| Observed | | Predicted | | |
|-------------------------------|----------|-----------|----------|--------------------|
| | | Language | | Percentage correct |
| | | Arabic | Japanese | |
| Language | Arabic | 215 | 145 | 59.70 |
| | Japanese | 105 | 225 | 70.80 |
| Overall percentage (accuracy) | | | | 65.30 |

Table 14. The -2LL and pseudo- R^2 values for model 4.

| | χ^2 | df | p | -2LL | Nagelkerke R^2 | Tjur R^2 | Cox and Snell R^2 | McFadden R^2 |
|-----------------|----------|----|-------|--------|------------------|------------|---------------------|----------------|
| Model | 537.29 | 1 | <.001 | 460.84 | 0.71 | 0.60 | 0.53 | 0.52 |
| Duration metric | 167.56 | 1 | <.001 | | | | | |

Table 15. Confusion matrix (sensitivity and specificity rates) and accuracy rate of model 4.

| Observed | | Predicted | | |
|-------------------------------|----------|-----------|----------|--------------------|
| | | Language | | Percentage correct |
| | | Arabic | Japanese | |
| Language | Arabic | 314 | 46 | 87.20 |
| | Japanese | 57 | 303 | 84.20 |
| Overall percentage (accuracy) | | | | 85.70 |

again with the *difference metric* as a predictor significantly improved the model's goodness of fit compared to using the *duration ratio* as a predictor.

4. Discussion and conclusion

The findings above agree with a large body of literature that has shown that Arabic and Japanese contrast short and long vowels (e.g., TSUKADA, 2013), as well as with previous observations that Arabic short vowels weigh approximately 50% of their long counterparts while Japanese short vowels weigh less than 50% of their long counterparts (e.g., TSUKADA, 2011). The duration difference and the duration ratio were, respectively, 128.16 and 0.48 for Arabic vocalics and were, respectively, 101.82 and 0.37 for Japanese vocalics. The duration differences (128.16 and 101.82 ms) do not reflect the short and long durations in Arabic or Japanese; Arabic short vowels are approximately 55% longer than Japanese short vowels, Arabic long vowels are approximately 65% longer than Japanese long vowels, and, overall, Arabic vocalics are approximately 60% longer than Japanese vocalics. Likewise, the duration ratio does not convey much information about vocalic duration within-language (e.g., Arabic or Japanese) or between the two languages (Arabic and Japanese) nor in comparison with other languages. Based on the data we obtained in this experiment, the duration ratios in Arabic and Japanese are relatively similar: 0.48 in Arabic and 0.37 in Japanese. That is, the durations of short and long vowels in Arabic are nearly double those in Japanese, but we cannot deduce this from the duration ratio.

In comparison, the duration metric (173.03 ms) and difference metric in Arabic (± 64.08 ms) diverged from the duration metric (111.20 ms) and the difference metric in Japanese (± 51.30 ms). The duration metric shows the average length of both short and long vowels; we can see clearly that Arabic vocalics are considerably longer than Japanese vocalics. The duration metric shows the extent to which short vocalics and long vocalics are similar or different within and between Arabic and Japanese, and we can ascertain that the difference between short and long vocalics in Arabic is greater than that in Japanese and that, moreover, that duration is more variable in Arabic than in Japanese. The two metrics together show that the duration metric of short and long vocalics in Arabic (173.3 ms) is

very close to the duration of long vocalics in Japanese ($111.20 \pm 51.30 = \mathbf{162}$ ms) and that the duration metric of short and long vocalics in Japanese (111.20 ms) is also similar to the duration of short vowels in Arabic ($173.03 - 64.08 = \mathbf{108.95}$ ms).

Neither the duration difference nor the duration ratio is a factual duration unit. Unlike the duration metric, the duration difference does not provide the actual duration of vocalics in Arabic vs. Japanese. Likewise, the duration ratio is a completely different measurement unit that no longer expresses the duration in time units and cannot indicate the duration of short vowels relative to long ones in Arabic vs. Japanese. The duration ratio cannot be used to compare vocalic durations between dialects or languages, because two different languages that have two distinct duration measurements for short and long vocalics may still have similar or even identical duration ratios. For instance, the duration ratio for Palestinian vocalics is approximately 0.39 (SAADAH, 2011) and for Japanese vocalics in the current experiment was 0.37. These two values are extremely similar, but overall, Palestinian short and long vocalics are both longer than their Japanese counterparts. Using the actual vocalic duration measurements in a statistical test to compare vocalics in Palestinian and Japanese should reveal a significant difference, while using the duration ratio is unlikely to reveal any differences. This is probably the reason why the duration and difference metrics were better predictors of the language.

To summarize, this paper shows how the duration difference and duration ratio measures used in previous studies are not optimal metrics for comparing vocalic duration within and across languages. We propose two alternative metrics: the duration metric and the difference metric. Using data from a previous study (TSUKADA, 2011), we illustrate the difference between the duration ratio and the duration difference, on the one hand, and between the duration metric and difference metric, on the other hand. We then conduct an experiment to examine the new metrics. The findings show that short and long vocalic durations differ in both Arabic and Japanese and that Arabic and Japanese also differ in terms of short and long vowel durations. More importantly, the key finding is that the proposed metrics were better predictors of the language than the traditional measures. This finding invites researchers on the vocalic duration, whether pho-

neticians, language acquisitionists, or speech pathologists, to consider using (and testing) of the proposed metrics. We also call for a revisiting of the findings established in previous literature, especially those studies that compared several languages or dialects (e.g., ALGHAMDI, 1998). Future research can survey languages and dialects that have shown similar or dissimilar duration ratios and examine whether the proposed metrics will reveal patterns that differ from those revealed by the traditional duration ratio and duration difference measures.

Acknowledgments

The author would like to thank King Saud University, Riyadh, Saudi Arabia, for funding and supporting this research through Researchers Supporting Project number (RSPD2023R580).

References

1. ABD ALMISREB A., ABIDIN A.F., TAHIR N.M. (2016), An acoustic investigation of Arabic vowels pronounced by Malay speakers, *Journal of King Saud University – Computer and Information Sciences*, **28**(2): 148–156, doi: [10.1016/j.jksuci.2015.08.003](https://doi.org/10.1016/j.jksuci.2015.08.003).
2. AHMED A.A.M. (2008), *Production and perception of Libyan Arabic vowels*, Ph.D. Thesis, Department of Communication and Language Sciences, Newcastle University.
3. AL-ANI S.H. (1970), *Arabic Phonology: An Acoustical and Physiological Investigation*, De Gruyter Mouton.
4. ALDHOLMI Y. (2022), Medial vowel temporal acoustics in Arabic and Japanese polysyllabic words, *Archives of Acoustics*, **47**(3): 319–329, doi: [10.24425/aoa.2022.142006](https://doi.org/10.24425/aoa.2022.142006).
5. ALDHOLMI Y., ALOTAIBI S., ALROUQI M. (2021), Rating nonnativeness in L1-Japanese L2-Arabic speakers' vowels, [in:] *ExLing 2021: Proceedings of the 12th International Conference of Experimental Linguistics*, pp. 9–12.
6. ALGHAMDI M.M. (1998), A spectrographic analysis of Arabic vowels: A cross-dialect study, *Journal of King Saud University – Journal of Arts*, **10**(1): 3–24.
7. ALMBARK R., HELLMUTH S. (2015), Acoustic analysis of the Syrian vowel system, [in:] *Proceedings of the 18th International Congresses of Phonetic Sciences*.
8. AL-TAMIMI J., BARKAT-DEFRADAS M. (2003), Inter-dialectal and inter-individual variability in production and perception: A preliminary study in Jordanian and Moroccan Arabic, [in:] *AIDA 5th Conference Proceedings (Association Internationale de Dialectologie Arabe)*, pp. 171–186.
9. AMIR N., AMIR O., ROSENHOUSE J. (2014), Colloquial Arabic vowels in Israel: A comparative acoustic study of two dialects, *The Journal of the Acoustical Society of America*, **136**(4): 1895–1907, doi: [10.1121/1.4894725](https://doi.org/10.1121/1.4894725).
10. AMIR N., TZENKER O., AMIR O., ROSENHOUSE J. (2012), Quantifying vowel characteristics in Hebrew and Arabic, [in:] *Proceedings of the Afeka-AVIOS Speech Processing Conference*, pp. 37–43.
11. BLOMGREN M., ROBB M., CHEN Y. (1998), A note on vowel centralization in stuttering and nonstuttering individuals, *Journal of Speech, Language, and Hearing Research*, **41**(5): 1042–1051, doi: [10.1044/jslhr.4105.1042](https://doi.org/10.1044/jslhr.4105.1042).
12. BOERSMA P., WEENINK D. (2021), *Praat: Doing phonetics by computer*, Version 6.2.03, <https://www.fon.hum.uva.nl/praat/> (access: 12.03.2022).
13. CANTINEAU J. (1956), The phonemic system of Damascus Arabic, *Word*, **12**(1): 116–124, doi: [10.1080/00437956.1956.11659595](https://doi.org/10.1080/00437956.1956.11659595).
14. COWAN W. (1970), The vowels of Egyptian Arabic, *Word*, **26**(1): 94–100, doi: [10.1080/00437956.1970.11435584](https://doi.org/10.1080/00437956.1970.11435584).
15. FERGUSON C.A. (1957), Two problems in Arabic phonology, *Word*, **13**(3): 460–478, doi: [10.1080/00437956.1957.11659647](https://doi.org/10.1080/00437956.1957.11659647).
16. FLEGE J.E., PORT R. (1981), Cross-language phonetic interference: Arabic to English, *Language and Speech*, **24**(2): 125–146, doi: [10.1177/002383098102400202](https://doi.org/10.1177/002383098102400202).
17. HARRIS Z.S. (1942), The phonemes of Moroccan Arabic, *Journal of the American Oriental Society*, **62**(4): 309–318, doi: [10.2307/594035](https://doi.org/10.2307/594035).
18. HASSAN Z.M. (1981), *An experimental study of vowel duration in Iraqi spoken Arabic*, Ph.D. Thesis, Department of Linguistics & Phonetics, University of Leeds.
19. HASSAN Z.M. (2002), Gemination in Swedish and Arabic with a particular reference to the preceding vowel duration: An instrumental and comparative approach, [in:] *Proceedings of Fonetik TMH-QPSR*, **44**(1): 81–84.
20. HONG S.A., SARMAH P. (2009), Korean perception of Arabic phonemes, *Current Issues in Linguistic Interfaces*, **2**: 375–386.
21. JAKOBSON R., FANT C., HALLE M. (1963), *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*, MIT Press, Cambridge.
22. KALALDEH R. (2018), Acoustic analysis of Modern Standard Arabic vowels by Jordanian speakers, *International Journal of Arabic-English Studies*, **18**(1): 23–48.
23. KENT R.D., VORPERIAN H.K. (2018), Static measurements of vowel formant frequencies and bandwidths: A review, *Journal of Communication Disorders*, **74**: 74–97, doi: [10.1016/j.jcomdis.2018.05.004](https://doi.org/10.1016/j.jcomdis.2018.05.004).
24. KOTBY M.N. *et al.* (2011), The Arabic vowels: Features and possible clinical application in communication disorders, *Folia Phoniatrica et Logopaedica*, **63**(4): 171–177, doi: [10.1159/000316323](https://doi.org/10.1159/000316323).
25. LABABIDI Z., PARK H. (2014), L1-English tense-lax vowel system influence on L2-Arabic short and long vowel learning, [in:] *Perspectives on Arabic Linguistics XXVIII: Papers from the Annual Symposium on Arabic Linguistics*, pp. 63–88, doi: [10.1075/sal.4.03lab](https://doi.org/10.1075/sal.4.03lab).

26. MITLEB F. (1984), Vowel length contrast in Arabic and English: A spectrographic test, *Journal of Phonetics*, **12**(3): 229–235, doi: [10.1016/S0095-4470\(19\)30879-4](https://doi.org/10.1016/S0095-4470(19)30879-4).
27. Phonic.ai (2023), *Phonic: Voice and video surveys*, <https://www.phonic.ai/> (access: 22.02.2022).
28. SAADAH E. (2011), *The production of Arabic vowels by English L2 learners and heritage speakers of Arabic*, Ph.D. Thesis, Department of Linguistics, University of Illinois at Urbana-Champaign.
29. TSUKADA K. (2009), An acoustic comparison of vowel length contrasts in Arabic, Japanese and Thai: Durational and spectral data, *International Journal of Asian Language Processing*, **19**(4): 127–138.
30. TSUKADA K. (2011), The perception of Arabic and Japanese short and long vowels by native speakers of Arabic, Japanese, and Persian, *The Journal of the Acoustical Society of America*, **129**(2): 989–998, doi: [10.1121/1.3531801](https://doi.org/10.1121/1.3531801).
31. TSUKADA K. (2012a), Comparison of native versus nonnative perception of vowel length contrasts in Arabic and Japanese, *Applied Psycholinguistics*, **33**(3): 501–516, doi: [10.1017/S0142716411000452](https://doi.org/10.1017/S0142716411000452).
32. TSUKADA K. (2012b), Non-native Japanese listeners' perception of vowel length contrasts in Japanese and Modern Standard Arabic (MSA), *Second Language Research*, **28**(2): 151–168, doi: [10.1177/0267658311435870](https://doi.org/10.1177/0267658311435870).
33. TSUKADA K. (2013), Vowel length categorization in Arabic and Japanese: Comparison of native and non-native Japanese perception, *Speech, Language and Hearing*, **16**(4): 187–196, doi: [10.1179/2050572813Y.0000000012](https://doi.org/10.1179/2050572813Y.0000000012).
34. ZALTZ Y., SEGAL O. (2021), The perception of Arabic vowel duration by L1 Hebrew speakers: Can a short training remold the effect of the native phonological system?, *Studies in Second Language Acquisition*, **44**(1): 143–163, doi: [10.1017/S0272263120000728](https://doi.org/10.1017/S0272263120000728).