Behavior Research Methods (2020) 52:1893–1905
https://doi.org/10.3758/s13428-020-01353-z

Matrices of the frequency and similarity of Arabic letters and allographs

Sami Boudelaa[1,2] · Manuel Perea[3,4,5] · Manuel Carreiras[4,6]

Published online: 19 February 2020

Abstract

Indicators of letter frequency and similarity have long been available for Indo-European languages. They have not only been pivotal in controlling the design of experimental psycholinguistic studies seeking to determine the factors that underlie reading ability and literacy acquisition, but have also been useful for studies examining the more general aspects of human cognition. Despite their importance, however, such indicators are still not available for Modern Standard Arabic (MSA), a language that, by virtue of its orthographic system, presents an invaluable environment for the experimental investigation of visual word processing. This paper presents for the first time the frequencies of Arabic letters and their allographs based on a 40-million-word corpus, along with their similarity/confusability indicators in three domains: (1) the visual domain, based on human ratings; (2) the auditory domain, based on an analysis of the phonetic features of letter sounds; and (3) the motoric domain, based on an analysis of the stroke features used to write letters and their allographs. Taken together, the frequency and similarity of Arabic letters and their allographs in the visual and motoric domains, as well as the similarities among the letter sounds, will be useful for researchers interested in the processes underpinning orthographic processing, visual word recognition, reading, and literacy acquisition.

Keywords Arabic letters · Allographs · Sounds · Frequency · Visual similarity · Phonetic similarity · Motoric similarity

The study of letter similarity (or confusability) and letter frequency has a long history over several decades within the fields of psychology and psychophysics (see Mueller & Weidemann, 2012, for a review). Continued interest in the study of this topic is predicated on the widely held belief that a good understanding of what drives perceived similarity among letters and the availability of reliable statistics regarding their distributional properties are crucial for a number of reasons. First, the study of letter properties lays the groundwork for the study of how letters are represented in the cognitive system, since letters of individual words are thought to represent the first "language-specific" stage of the reading process, following the work done by oculomotor control mechanisms enabling fixation on the word and the early visual processing that allows visual feature extraction (Carreiras, Armstrong, Perea, & Frost, 2014; Dehaene, Cohen, Sigman, & Vinckier, 2005; Grainger, 2008). Second, since mastery of alphabetic reading is generally thought to require, as a first step, the ability to map letters and letter strings onto the sounds of the language (Bowey, 2005; Snowling & Hulme, 2011), the study of letter properties can provide valuable information to educators regarding the complexity of letter forms and guide the choice of the order in which the learner is exposed to these letters. Finally, the investigation of letter properties promotes empirical investigations with a view toward gaining a better understanding of how the visual system functions.

For many years, researchers have sought to establish letter frequency databases for different languages such as Russian (Gusein-Zade, 1988), English (Mayzner & Tresselt, 1965), and Spanish (Li & Miramontes, 2011) in order to provide normative frequency data for researchers interested in verbal

✉ Sami Boudelaa
s.boudelaa@uaeu.ac.ae

[1] Department of Cognitive Sciences, United Arab Emirates University, Al Ain 15551, UAE

[2] Department of Psychology, University of Cambridge, Cambridge, UK

[3] Universitat de València, Valencia, Spain

[4] Basque Center for Cognition, Brain, and Language, Donostia-San Sebastian, Spain

[5] Nebrija University, Madrid, Spain

[6] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

learning and retention, anagram problem solving, word recognition thresholds, and linguistic analyses. Similar interest in developing letter similarity/confusability matrices is evident in a long research tradition spanning several decades, with the early work, mainly on English, seeking to identify typefaces, fonts, and letters that were more or less legible, with the aim of improving printing and typesetting (Roethlein, 1912; Tinker, 1928). More recently, research has come to focus on understanding the visual system and how it represents and processes letters as visual objects, without losing interest, however, in attempting to make written text more comprehensible or helping learners to acquire reading skills more easily (Boles & Clifford, 1989; Fiset et al., 2009; Liu & Arditi, 2001; Mueller & Weidemann, 2012). Collectively, these studies have played a fundamental role in enabling the design and implementation of many well-controlled empirical studies seeking to pin down the dynamics of letter processing (e.g., Evans, Lambon Ralph, & Woollams, 2017; Grainger, Dufau, Montant, Ziegler, & Fagot, 2012; Kinoshita & Kaplan, 2008; Schelonka, Graulty, Canseco-Gonzalez, & Pitts, 2017).

Despite the importance of having reliable letter similarity matrices and letter frequency counts, this type of information is available only for a handful of Indo-European languages. Other languages, such as Modern Standard Arabic (henceforth MSA), suffer from a lack of lexical resources in general and computerized databases about letter similarity and letter frequency in particular. MSA is the language taught at most schools, colleges, and universities in the Arab world and is the one used in the media, literature, and formal settings such as political meetings (e.g., Kamusella, 2017; Versteegh, 2014). This language, despite its importance for the study of letter processing and letter representation by virtue of its very special writing system, as we will detail below, has very few published lexical resources. Notable exceptions are Aralex (Boudelaa & Marslen-Wilson, 2010) and Arabicorpus (Parkinson, 2000). Therefore, researchers interested in the study of Arabic letter processing, Arabic reading, and developing better Arabic reading tools, and psycholinguists interested in cross-linguistics investigations of letter and word processing are in dire need of reliable information about the distributional characteristics of letters and their similarities.

The aim of this study is to provide, for the first time, (a) comprehensive statistical information about Arabic letters and their allographs and (b) a similarity/confusability matrix of Arabic letters and allographs in the visual, auditory, and motoric domains. We begin by providing some relevant background about the orthographic system and its importance for the study of letter processing. Second, we provide a detailed statistical count of the frequencies of Arabic letters and their allographs based on a 40-million-word corpus. Third, we present a visual similarity matrix of Arabic letters and their allographs based on ratings by 125 participants, followed by a phonetic similarity matrix based on theory-driven phonetic features and a motoric similarity matrix based on the strokes required to write each letter and its allographic variants. We conclude by highlighting the importance of this new set of information on the distributional and structural properties of Arabic for future investigation of this language in different research fields.

The Arabic writing system

MSA is a Semitic language written from right to left in a cursive manner. The MSA alphabet consists of 28 letters, 22 of which always connect to the following letter using a ligature, while the remaining 6 connect to the preceding but not the following letter. MSA is the fifth most common language in the world, with over 300 million speakers. One of the most important features of the Arabic writing system is "allography," whereby the shapes of 15 of the 28 letters differ considerably depending on their location within the letter sequence (initial, middle, final, and isolated). For instance, the letter ع, which stands for a voiced pharyngeal fricative represented by /ʕ/ in IPA notation, takes the shape عـ word-initially, ـعـ word-medially, ـع word-finally when preceded by a ligating letter, and ع word-finally when preceded by a non-ligating letter. The remaining 13 letters (e.g., ب, ث, د, ر) preserve their shapes regardless of their position within the word, but have ligature marks on either side (e.g., ـبـ, ـثـ) or only on their right-hand side (e.g., ـد, ـر). Another important feature of the MSA orthographic system is the use of a cursive writing system even in typing, a rare feature among the world's writing systems, including typologically related languages such as Hebrew. A final unique aspect of MSA is that a given letter can have up to three diacritic symbols superposed on it, thus creating a highly complex visual percept. This is illustrated by the second letter خ of the word مـخٌّ "brain," which shows a single dot diacritic underneath a gemination sign indicating that the consonant خ is doubled, and the nunation sign, which denotes the indefinite article *-un*.

The complexity of this orthographic system has given rise to many studies across several research areas. For instance, in the field of reading, Asadi, Khateb, and Shany (2017) showed that unlike Indo-European languages, where reading processes are seen as the product of decoding abilities and listening comprehension, MSA requires an extended model that includes the orthographic and the morphological domains in order to capture the intricacies of reading in Arabic. Similarly, some researchers have suggested that the complexity of the Arabic orthographic system leads to slower processing than in related languages such as Hebrew (Ibrahim, Eviatar, & Aharon-Peretz, 2002), while others (Taha & Saiegh Haddad, 2017) have argued that this feature leads Arabic orthography learners to rely on morphological structure much earlier in the course of learning to read and spell than their Indo-European counterparts.

In the visual word recognition domain, researchers have been interested in establishing the role of allography and whether Arabic cognitive representations contain a level that corresponds to abstract letter identities (Boudelaa, Norris, Mahfoudhi, & Kinoshita, 2019; Carreiras, Perea, & Abu Mallouh, 2012; Friedmann & Haddad-Hanna, 2012; Perea, Abu Mallouh, & Carreiras, 2010). This line of research relates to a much broader set of issues in cognitive science regarding the types of representations used in reading and whether letter recognition is subserved by a hierarchical processing system that involves both case-specific and case-independent representations of alphabetic stimuli (Petit, Midgley, Holcomb, & Grainger, 2006; Rothlein & Rapp, 2014, 2017). In this respect, Boudelaa et al. (2019) reported a series of priming experiments looking at whether a target word (e.g., يـعـود "be happy") is facilitated more by a nonword transposed letter (TL) prime that does not cause allographic changes (e.g., يـ ع دود) than a TL prime that causes such changes (e.g., يـدعـود). The results showed that the non-allographic TL primes produced significantly greater facilitation than allographic TL primes, indicating that Arabic readers use allographic variation to resolve the uncertainty in letter order during the early stages of orthographic processing. Similar results were reported by Yakup, Abliz, Sereno, & Perea (2014, 2015) for Uyghur, a Turkic language spoken in Western China that uses the Arabic orthographic system, suggesting that visual form changes that Arabic letters undergo as a function of their position in the word play a critical role in guiding the reading process.

Finally, in the field of automatic language processing, there has been a recent surge in the study of the characteristics of typed and handwritten Arabic letters to develop algorithms that can automatically process Arabic written scripts (Abandah, Younis, & Khedher, 2014; Cowell & Hussain, 2002; Khorsheed, 2002). The development of new lexical resources related to letter frequency and letter similarity can only help to further spur interest in MSA and provide the tools necessary to conduct well-controlled and replicable research.

Letter and allograph frequencies

Here we provide the frequency of Arabic letters and their allographs based on the 40-million-word corpus previously used by Boudelaa and Marslen-Wilson (2010) to develop the Aralex database. These frequency figures were calculated as percentages over the non-diacritized version of Aralex. In Table 1, we provide the frequencies of the 28 letters of the alphabet along with the letter frequencies published online by Mohsen Madi (2010) for comparison.

There are numerous similarities between the frequency statistics of the current study and Madi's (2010), as demonstrated by a Pearson correlation analysis ($r = 0.9$), suggesting a close match between the two sets of frequencies. The small discrepancies in the frequency counts between the two studies are probably due to the use of different kinds of corpora. The current study's 40-million-word corpus comes from contemporary written sources, namely newspaper articles, as detailed in Boudelaa and Marslen-Wilson (2010). In contrast, Madi (2010) relied on a small corpus of a little more than one million words derived mainly from old Arabic books such as ا بداية وا نهاية *The Beginning and The End* of Ibn Katheer (1300–1373) and ا حـ يق ا خـ تـ و *The Sealed Nectar* by Al Mubarkafoori, which is a compilation of the sayings of the Prophet of Islam produced in classical Arabic 14 centuries ago, or on books that deal with Islamic jurisprudence and hence use mostly older Arabic, such as تـ حـ فـ ة ا ع و ين *The Masterpiece of the Brides* by Al-Shuri.

It is important to further note that the current letter frequency values make intuitive sense, because the four letters with the highest frequencies are on the one hand the letters و and لـ, which respectively correspond to the function words "and" and "in order to," and the letters تـ ,يـ on the other, which are in fact inflectional affixes. At the same time, the letters with the lowest frequencies correspond either to marked sounds that are very rare across the world languages, such as the pharyngealized alveolar and the pharyngealized interdental ظـ, or indeed to letters that do not correspond to function words or affixes, such as ذ ,ثـ and خـ.

In Table 2 below we present for the first time the frequencies of Arabic letters broken down by allograph.

For each letter of the alphabet, we determined the frequency of its allographic form in isolation and at the onset, middle, and offset of the word. Thus, for the majority of letters, such as ع *ain*, and غ *ghayn*, we report the frequencies of four allographs, whereas for others, such as د *daal*, ذ *thaal*, *raa*, and *zein*, we report only two values because they have only two allographic forms. For the letter أ *alif,* we report values for seven allographic forms because this letter has different interchangeable variants such as أ, أ , and ا. Finally, for the letter تـ, *taa*, we report values for six allographs, four of which are for the *taa maftuuha*, "open taa," and two for *taa marbuuta* "closed taa" As can be clearly seen from Table 2, allographs of the same letter do not occur with the same frequency across the board. For instance, the allograph بـ, baa, with a frequency of 2%, is much more common than the allograph ب with only 0.22%. The frequencies of other letter allographs (e.g., ,0.26 0.76) are much more evenly distributed.

An interesting theoretical question that allograph frequencies can help address is whether the effects of allographic changes in visual word recognition experiments, such as those reported by Friedmann and Haddad-Hanna (2012) and Boudelaa et al. (2019), can be modulated by

Table 1. Percentage frequencies of the 28 Arabic letters in the current study and in Madi 2010

Arabic Letter	Current Study	Madi 2010	Arabic Letter	Current Study	Madi 2010
أ	1.43	2.76		0.70	0.51
ﺑ	4.17	3.47		1.10	0.38
ﺗ	6.87	3.18	ﻇ	0.23	0.26
ﺛ	0.43	0.43	ع	2.50	2.84
ﺟ	1.51	1.00	ﻏ	0.57	0.37
ﺣ	1.84	1.25	ﻓ	2.82	2.64
ﺧ	0.88	0.76	ﻗ	2.14	2.13
د	2.57	1.81	ﻛ	1.99	3.17
ذ	0.37	1.49	ﻟ	8.40	11.55
	4.61	3.75	ﻣ	5.77	8.08
	0.85	0.48	ﻧ	5.44	8.25
	2.74	1.82	ه	4.91	4.49
	1.11	0.64	و	8.36	8.36
	1.06	0.63	ﻳ	7.12	6.64

allographic frequency. From a practical point of view, these data can help educators not only in making informed choices about the development of teaching materials that reflect the frequency of different letters and their allographs, but also in modulating their instructional focus. For instance, when teaching the letter ع the instructor can, based on allograph frequency data, dedicate more time to teaching the allograph ع than the allograph ع , given that the latter is much more frequent than the former and may not need as much time to be learned.

Subjective Letter Similarity Experiment

The technique that we employed to construct the similarity matrix is based on data obtained under normal (untimed)

Table 2 Percentage frequency (% Frq) of 116 Arabic letter allographs (Allog)

Allog	% Frq	Allog	% Frq	Allog	% Frq	Allog	% Frq	Allog	% Frq	Allog	% Frq
أ	0.51	ﺟ	0.05		0.05	ﻇ	0.00	ﻛ	0.09	ؤ	0.05
أ	0.92	ﺠ	0.48		1.04	ﻈ	0.03	ﻜ	0.72	ؤ	0.1
آ	0.04	ﺠ	0.05		0.12		0.02	ﻜ	0.26	و	5.32
ا	3.39	ﺞ	0.93		1.53	ﻇ	0.18	ﻛ	0.92	و	3.04
ا	11.02	ﺣ	0.05		0.01	ع	0.08	ﻟ	0.11	ى	0.04
إ	0.25	ﺤ	0.51		0.32	ع	0.71	ﻟ	5.02	ى	0.2
ﺑ	0.11	ﺤ	0.1		0.04	ع	0.17		0.37	ﺋ	0.01
ﺑ	2	ح	1.18		0.74	ع	1.54		2.9	ﺋ	0.47
ب	0.22	ﺧ	0.01		0.01	ﻏ	0.01	ﻣ	0.11	ﻳ	0.19
ب	1.84	ﺨ	0.31		0.26	ﻏ	0.17	ﻣ	1.45	ﻳ	1.79
ﺗ	0.71	ﺨ	0.02		0.04	ﻐ	0.02		1.02	ي	0.91
ﺗ	1.73	خ	0.54		0.75	غ	0.37		3.19	ي	4.23
ﺗ	0.32	د	0.69		0.03	ﻓ	0.07	ﻧ	0.44		0.14
ت	4.11	د	1.88		0.19	ﻓ	1.4	ﻧ	1.15	ﺋ	0.08
ة	0.31	ذ	0.11		0.05	ﻓ	0.15	ن	1.3	ﺋ	0.05
ة	1.35	ذ	0.26		0.43	ف	1.2	ن	2.55		
ﺛ	0.02		1.14		0.02	ﻗ	0.08	ه	0.3		
ﺛ	0.14		3.47		0.26	ﻗ	0.55	ﻫ	0.74		
ﺛ	0.03		0.24		0.06	ﻗ	0.14	ﻫ	1.32		
ث	0.24		0.61		0.76	ق	1.37	ﻫ	2.55		

reading conditions and is comparable to the approach used in previous studies examining letter knowledge in children (Treiman, Kessler, & Polo, 2006; Treiman, Levin, & Kessler 2007, 2012) and letter similarity in adults (Simpson, Mousikou, Montoya, & Defior, 2013). Participants were speakers of MSA who were required to rate letter pairs on a scale from 1 (*not similar at all*) to 7 (*very similar*). We anticipate that the matrix presented here will also prove useful to researchers in any field of investigation in which Arabic letters are used as stimuli and where a measure of visual similarity between stimuli is required.

Method

Participants

A total of 125 participants, aged 20 to 24, were recruited to take part in this experiment. All participants were literate MSA speakers who were undergraduate students in the female campus of the faculty of Humanities and Social Sciences at United Arab Emirates University. All participants spoke English as a second language but declared Arabic (i.e., MSA and the Emirate Dialect) their dominant language. This experiment was approved by the ethics committee of United Arab Emirates University, and all participants gave their written consent to take part in it in return for 50 AED.

Stimuli

As in the previous study, we selected four allographs for each letter of the alphabet except for the letters (a) , , , and ظ, for which only two allographs were included; (b) the letter ه, for which only three allographs were used; and (c) the letter أ *alif*, for which eight different allographs were included. This choice, which was based on pilot testing, resulted in a total of 110 allographs. Each allograph was paired with every other allograph, including itself, resulting in 6105 pairs. These were used to build 15 experimental lists consisting of 407 experimental pairs each. Each participant was randomly assigned to one list. To ensure that subjects were assessing the visual, and not phonetic, similarity between the different allograph pairs, a further 32 foil pairs were built consisting of the 28 Arabic letters paired with Latin letters to create four conditions. The first consisted of cross-alphabet letter pairs that were both phonetically and visually similar. These were pairs like ـأ L, which share the straight downward-directed stroke. The second condition consisted of Arabic-Latin pairs which were phonetically similar but visually dissimilar, such as ـنـ-N, which share phonetic features [+coronal, +nasal, +continuant, +sonorant] but look very different visually. The third condition consisted of cross-alphabet pairs that were phonetically dissimilar but visually similar, like ـخـ-G, which share the downward-directed semicircular stroke. The final condition comprised pairs that were neither phonetically nor visually similar, such as ـذ I. The ordering of the letters within each pair was counterbalanced across lists, such that each letter appeared almost half of the time in the first position and half in the second.

Design and procedure

The presentation of the stimuli and recording of responses were controlled by desktop computers running SuperLab 5. On each trial, two stimulus allographs appeared at the center of the screen in Traditional Arabic 72-point font size in black against a white background. Participants were instructed to ignore the sounds of the letters and to rate the letter pairs on the computer keyboard based purely on visual similarity on a scale from 1 (*not at all similar*) to 7 (*very similar*). No time limits were imposed, and participants responded at their own pace. Participants could advance to the following trial only after providing a response to the current trial. To emphasize the importance of paying attention to the shape of the allograph, participants were also asked to rate a number of geometrical shapes (e.g., squares, rectangles, circles) on their similarity in shape. The experiment lasted about 15 minutes.

Results and discussion

An initial screening was performed on the data in order to detect cases in which the participants may have misunderstood or not correctly followed the instructions. This resulted in the exclusion of no data points at all. A second screening process tested whether participants' knowledge of the letter sounds exerted a strong influence on their responses, by examining the ratings assigned to the Arabic-Latin letter pairs. We have linearly rescaled the similarity ratings on the 1–7-point scale into distances on a 0–1 scale. In order to take into account the fact that human-generated similarity judgments are likely to be logarithmic on actual distance, we used the following formula: Distance = [exp(7) − exp (Distance$_1$)]/[exp(7) − exp(1)], where Distance$_1$ is the distance between a given pair of letter allographs. This formula simply rescales the similarity score provided by the participants into a distance metric that can be fed to the hierarchical clustering technique.

Table 3 Mean (and standard deviation) of the visual distance between cross-alphabet Roman–Arabic letter pairs

+P+V	+P−V	−P +V	−P−V
0.77 (0.07)	0.87 (0.03)	0.81 (0.07)	0.91 (0.05)

Note: +P+V = phonetically and visually similar; +P−V = phonetically similar but visually dissimilar; −P+V = phonetically dissimilar but visually similar; −P−V = phonetically and visually dissimilar

Table 3 suggests that although the overall perceived visual distance among cross-alphabet letters is large, the +P+V pairs (e.g., ﻟـ-L) and –P+V (e.g., غـ-G) pairs were perceived as significantly closer in visual space than the +P−V (e.g., ﺑـ-B) and the –P−V pairs (e.g., -E). Thus, phonetic similarity did not modulate the perceived distance among the cross-alphabet pairs, with the visually similar pairs perceived to be the same distance from each other regardless of phonetic similarity, and the visually dissimilar pairs being rated as maximally distant from each other regardless of whether they were phonetically similar. A series of paired two-tailed t tests revealed +P+V to be significantly different from +P−V ($p < 0.00$) and –P−V ($p < 0.00$), but not from –P+V ($p = 0.48$). More interestingly, –P+V was also reliably different from +P−V ($p < 0.01$) and –P−V ($p < 0.02$). This pattern of results clearly demonstrates that participants carried out the task solely based on the visual similarities of the letter pairs and completely ignored the phonetic dimension as instructed.

Where the within-alphabet letter and allograph pairs are concerned, the full visual similarity matrix for 110 allographs can be accessed at https://osf.io/yqns4/, with the distance measures rescaled using the distance formula mentioned above. The dendrogram in Fig. 1 displays the hierarchical relationships of the 110 Arabic allographs used in this experiment.

The general technique we use here is hierarchical clustering, which aims to group similar objects into groups called clusters (Kassambara, 2017; Jajuga, Sokolowski, & Bock, 2002; Stahl, Leese, Landau, & Everitt, 2011). The end point of such an approach is to create a set of clusters that are distinct from each other, while the objects within each cluster are broadly similar to each other. Hierarchical clustering typically operates on a distance matrix. It starts by treating each observation as a separate cluster, then it iteratively identifies the two clusters closest to each other and merges them until no clusters are left unmerged. The main output of hierarchical clustering is a *dendrogram*, which is simply a diagram that shows the hierarchical relationships between objects. The main use of a dendrogram is to work out the best way to allocate objects to clusters, and this usually requires (1) the computation of the distance (similarity) between two given clusters using a distance metric (e.g., Euclidean distance, city block) and (2) selecting a linkage criterion to determine whether the distance is computed between the two most similar parts of a cluster (single-linkage), the two least similar bits of a cluster (complete-linkage), the center of the clusters (mean or average-linkage), or some other criterion.

In this study, all dendrograms are based on the standard Euclidean distance metric and use “ward. D2” as a linkage criterion to determine the distance between sets of observations as a function of the pairwise comparisons (Murtagh & Legendre, 2014). However, since hierarchical cluster analysis can typically yield as many cluster solutions as there are cases to be clustered (Clatworthy, Buick, Hankins, Weinman, & Horne, 2005), one needs to determine the appropriate cluster solution using objective formal rules and equations to identify the optimal number of clusters in a sample. Here we have opted for the “gap statistic,” which operates by taking the input of the hierarchical clustering analysis and compares the change in within-cluster dispersion with that expected under a reference null distribution. The gap statistic has been reported to outperform other methods (Tibshirani & Walther, 2005) and to provide quite stable solutions (Yan and Ye, 2007). Upon applying this method to our data, the results suggest that the value that maximized the gap statistic was 0.94, with an optimal number of 19 clusters (Table 4).

Table 4 shows that the largest of the 19 clusters consisted of nine allographs, and the smallest consisted of two. The within-

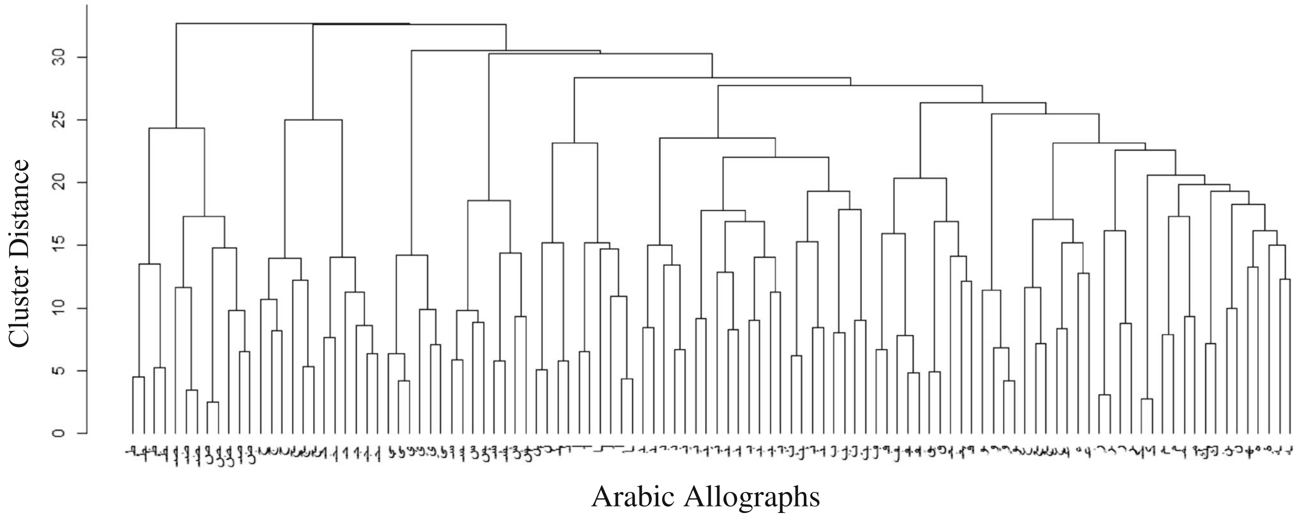

Fig. 1 Hierarchical clustering (dendrogram) using the nearest neighbor method. The vertical axis of the dendrogram represents the distance or dissimilarity between clusters. The horizontal axis represents the 110 Arabic allographs

Table 4 Optimal number of clusters based on visual similarity as suggested by the gap method, the members of each class, and its within-cluster sum of squares

Cluster number	Cluster members	Within-cluster SS
1	عع غغ ع غ	1.083648
2	آأإا أ ا	1.134387
3	ؤ ؤ و و	0.72502
4	ﺋ ﺌ ى ي ﻰ ﻴ	0.056119
5	ﺌ ﺒ ﺘ ﺋ ﺑ ﺗ ن ي ﻳ	0.909182
6	ﺑﺘﺜﺜ ب ت ث ث	1.736229
7	ﺔ ة ﻚ ن ﻬ ﻫ ﻜ ﻨ ﻪ	7.776124
8	ﺟﺤﺨ ج ح خ	0.540865
9	ﺟ ﺣ ﺧ ﺠ ﺤ ﺨ	0.621634
10	ﺪ ﺬ د ذ	0.177571
11		1.583232
12		0.000000
13		0.307081
14	ظ	1.515929
15	ع ﻐ ق ﻗﻘ	0.383974
16	ف ﻔ ﻗ ﻓﻔ	0.025204
17	ﻛ ﻜ	0.059794
18	ﻟﻠ	0.000000
19	ﻣﻤ	0.012434

cluster sum of squares (SS), which measures the amount of variance in the data, is < 2 for all clusters except Cluster 7. Although the within-cluster SS is influenced by the number of observations and is therefore often not directly comparable across clusters with different numbers of observations, the preponderance of low SS for all clusters save one suggests that the clusters are highly consistent, with very little variability. In addition, the total SS is 40.62 and the between-cluster SS is 21.97, suggesting that data points cluster neatly in a 19-dimensional space of visual attributes.

The component members of each cluster share a number of characteristics that the participants relied on to assign their similarity ratings. For example, Cluster 14 in Table 3 features the allographs ﻈ ﻂ ﻇﻃ, which share the egg-shaped loop with a vertical stroke, and the only difference between them is the dot above the first and third members of this set. Similarly, the eighth cluster in the same table features the six allographs ﺟﺤﺨ ج ح خ, with the first three ligating to the right (i.e., to the preceding letter), while the second three do not. Two main features cut across the members of this cluster: the downward-directed semicircle and the acute angle it makes at its upper end. Even Cluster 7, which consists of nine seemingly heterogeneous allographs overall, reveals a clear structure at a lower level of granularity, with the allographs ﻨ and ن sharing the downward-directed semicircle, while the ﻬ , ﺔ , ﻫ , ه, ﺔ, share the closed loop written on or above the line. The final two members of this cluster are the isolated ﻚ and the right-ligating ﻚ. One reason these two allographs are grouped with Cluster 7 is arguably the small dot-like shape in the middle of the two allographs, which allies them with the four dot-bearing allographs in this cluster.

Table 4 further suggests that phonetic similarity among allographs played little or no role in the similarity judgment process. This is clearly illustrated by Cluster 1, for example, where the allograph corresponds to a voiceless glottal stop sound, whereas the allographs ع ع ع and the allographs ﻐ ﻐ غ correspond to a voiced pharyngeal fricative and a voiced velar fricative, respectively. More importantly, perhaps, the cluster membership as illustrated in Table 4 is in keeping with recent psycholinguistic and neurolinguistics research on Arabic letter allography (Boudelaa et al., 2019; Friedmann & Haddad-Hanna, 2012; Yakup, Abliz, Sereno, & Perea, 2014, 2015). For instance, the allographs ج and ﺟ are two different instantiations of the abstract letter ﺟ, but they belong to Clusters 8 and 9, respectively. This strongly suggests that different allographic shapes of the same abstract letter were treated as two different perceptual objects in our similarity judgment task. Further credence for this idea comes from the recent demonstration by Boudelaa et al. (2019) that transposed-letter priming (TL-priming) is modulated by allographic changes, such that a target word like ي عدون "be happy" is easier to recognize when preceded by the non-allographic TL-prime يع دون than when preceded by the allographic TL-prime ي دعون. Similar results were reported by Yakup et al. (2014, 2015) for Uyghur, a non-Semitic language that uses the Arabic writing system, and by Friedmann and Haddad-Hanna (2012), who showed that Arabic dyslexic patients' letter migration errors when reading aloud were reduced for words in which letter transposition or letter substitution caused allographic changes.

The current experiment refines and extends the recent findings of Wiley, Wilson, and Rapp (2016) in a number of ways. For example, those authors studied the similarity structure of 45 Arabic letter shapes in a timed same–different judgment task with experienced and novice speakers. Our study included 110 allographs, allowing us to provide the principled similarity structure displayed in Fig. 1 above for allograph groups absent from Wiley et al.'s study. Consider, for instance, the letter ﺋ. In our study, this letter meaningfully clusters with its allographic variant in a right-ligating context (i.e., ﻲ), with the allograph called *alif maqsuura* in isolation with or without a *hamza* ى ئ, and with the *alif maqsuura* ligating to the right with and without the glottal stop, *hamza* ﺊ ﻰ . The same letter ﺋ in Wiley et al. (2016) clusters with ﻣ and ﻫ in the latency and accuracy data of the expert subjects, respectively, making it more difficult to isolate the basis of the visual similarity underlying such clusters. Further, Wiley et al. (2016) did not include the glottal stop, *hamza*, either by itself ( ) or in the context of the different letters that can support it, such as *alif* أ,

alif maqsuura ئـ, *waaw* ؤ, or *nabrah* ـئـ. Presumably, Wiley et al.'s choice is reasonably predicated on the standard view that the *hamza* is not a letter of the alphabet. We have opted for completeness and included the glottal stop in our analysis. In doing so, we have gained the novel insight that this letter is typically treated like a dot when it occurs in the context of a supporting letter. Thus, ؤ clusters with و ـو, while ـئـ clusters with ي ن ـتـ ـب. In contrast, isolated is treated like a full-fledged letter allograph and clusters with غ ع ـغـ غـ ـع ع, arguably because it is perceived as a miniature ع.

Finally, our study provides strong empirical support for Wiley et al.'s observation that allographs of letters in the middle position (e.g., ـجـ ـخـ ـحـ ـثـ ـتـ ـب) are identical to the corresponding allographs in the initial position when the ligature to the right is ignored (i.e., جـ خـ حـ ثـ تـ بـ). Based on the structure of Clusters 5 and 9 in our data, it is clear that participants ignored the right ligation of the middle allographs and grouped them with their counterparts in the initial position. This is a seemingly surprising outcome, since ligation is not only taught as part of the letter form to Arabic learners, but it also provides crucial information about word length and lexical stress position (Boudelaa et al., 2019). It is however consistent with recent research that reports comparable masked repetition priming effects for isolated letter pairs with similar (e.g., ـفـ فـ) and with dissimilar (ـع ع) visual features across letter positions (Carreiras, Perea, Gil-López, & Abu Mallouh, 2013. Furthermore, event-related potential (ERP) data recorded continuously while subjects performed a masked same–different matching task with visually similar (e.g.,   ) and visually dissimilar (e.g., ـعـ ع) allographs clearly show an early ERP (P/N150) associated with visual form similarity, and a later ERP component (P300) related to abstract letter representations. Specifically, allographs like ـع-ع showed clear electrophysiological response differences early on in processing, while brain responses later in processing were modulated by abstract letter representations such that ـع-ع were perceived as equally similar as - (Carreiras, Perea, Gil-López, Abu Mallouh, & Salillas, 2013).

Phonetic letter similarity

The ability to quantify the phonetic similarity between words is important in many fields, including computational linguistics, dialectometry, applied linguistics, psycholinguistics, and cognitive neuroscience. The literature provides a number of methods for measuring the degree of phonetic similarity between segments. Some of these are based on experimental studies showing, for instance, the degree of confusability of different segments (Klatt, 1968; Greenberg & Jenkins, 1964; Mohr & Wang, 1968). Others are based on more theoretical arguments (Austin, 1957). Others still have opted for quantifying the degree of similarity between segments by counting the number of differences in their specifications in terms of phonetic/phonological features (Ladefoged, 1970). Here we opted for the use of phonetic features to quantify the amount of similarity/difference among the various Arabic letter sounds. Our choice is predicated on recent reports in the literature suggesting that similarity between component speech sounds is much better captured by theoretically driven measures based on phonetic/phonological features than empirically derived measures based on confusability (Bailey & Hahn, 2005; Hahn & Bailey, 2005). Accordingly, we focused on providing a similarity metric that simultaneously compares consonants and vowels using 16 features from phonological theory. Specifically, these consist of a first set of three *Major Class* features that define the major classes of sounds in the language into consonantal, sonorant, and approximant. A second set consists of seven *Place of Articulation* features, namely, labial, coronal, dorsal, pharyngeal, anterior, distributed, and high, serving to define the specific articulator involved in producing the sound. A third set of four features, continuous, lateral, nasal, and strident, pertains to the *manner* in which the letter sound is produced. Finally, a fourth set consists of one *Laryngeal* feature, voicing, that distinguishes voiced from voiceless segments, and a fifth set comprises a *Quantity* feature, categorizing segments as long and short. The full matrix of features for the 28 consonants and 6 vowels of the language is accessible here: https://osf.io/mx5t7/.

Using these features, each letter was then converted into a vector consisting of 16 elements of 0s and 1s (0 if the feature did not apply to the letter and 1 if it did). We then performed the same hierarchical clustering procedure on these vectors as before in order to determine the similarity structure underlying them (see Fig. 2).

Visual inspection of the dendrogram in Fig. 2 suggests that there are seven distinct phonetic sound clusters, with an average number of letter sounds per cluster ranging from two to eight. However, to more objectively determine the optimal number of groups that the 36 letter sounds cluster into, we used the gap statistic as before. The results of this analysis suggest that the optimal number of clusters is five, with a maximal value of 0.23. The sizes of these clusters, displayed in Table 5, range from 5 to 10 members.

Interestingly, the different clusters make intuitive sense. For instance, the members of Cluster 1 are all back fricative consonants except for the voiceless glottal stop أ / /, which is part of this cluster because it shares many features with the voiceless

glottal fricative ـهـ /h/, which in turn naturally clusters with the back fricatives ع ـخـ ـحـ غـ / x ħ γ/. Similarly, the members of Cluster 2 are all bilabial consonants except for the palatal approximant يـ /y/ arguably added to this cluster due to its similarity to the bilabial approximant و /w/, which shares the place feature of bilabial with all the other members of the cluster. The largest cluster, Cluster 3 with 10 members, consists of consonants that are all non-back consonants with

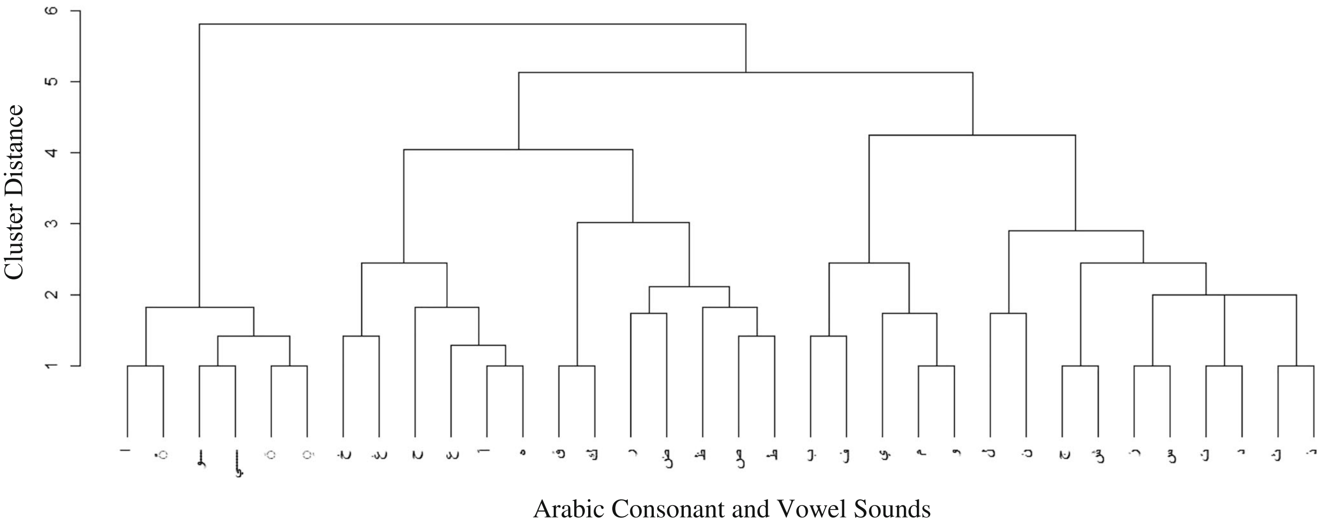

Fig. 2 Hierarchical clustering (dendrogram) using the nearest neighbor method. The vertical axis of the dendrogram represents the distance or dissimilarity between clusters. The horizontal axis represents the 34 Arabic sounds

places of articulation starting with the ﺟ /j/ at the palate and progressing anteriorly to the dental area with the ذ /ð/ and ث /θ/ sounds. Cluster 4 includes seven sounds, all emphatic. In the environment of such sounds, the low front vowel phoneme /æ/ of the language is standardly pronounced as a low back vowel /a/, which is the typical manifestation of phonetic emphasis in Arabic. The only non-emphatic sound in this cluster is the velar ك /k/, arguably added to this cluster by virtue of sharing the features back, voiceless, and plosive with the sound ق /q/. Finally, Cluster 5 includes the six vowels of the language.

It is interesting to note that the within-cluster SS is 8.59 on average, while the total SS and between-cluster SS stand at 90.6 and 47.6, respectively, suggesting a high degree of consistency within the component members of each cluster. Furthermore, our theoretically driven measure of similarity based on phonetic features is in agreement with empirically derived measures based on confusability as shown by hidden Markov recognition systems. For instance, Maaly, Elobeid, and Ahmed (2002) reported that the sounds /ħ/ and / / are highly confusable and that their automatic Arabic phoneme recognizer failed to distinguish between them. It is also with consistent with the phonological neutralization processes at play in many Arabic dialects. For instance, in the Egyptian dialect spoken in Cairo, the interdental voiceless fricative ث /θ/ is typically realized as ت /t/ (e.g., ث ن /θæmæn/ "price" pronounced ت ن /tæmæn/) or /s/ (e.g., ثانية /θaanyæ/ "second" pronounced انية /saanyæ/). These phonemes /θ, t, s/ are members of Cluster 3. Analogously, phonological speech errors made by children learning Arabic (e.g., ق بي /qalbi/ "my heart" pronounced as ك بي /kalbi/ "my dog") also seem to target phonemes that are members of the same clusters (Dyson & Amayreh, 2000).

Table 5 Optimal number of clusters based on phonetic letter similarities as suggested by the gap method, the members of each class, and its within-cluster sum of squares

Cluster number	Cluster members	Within-cluster SS
1	أ ح خ ع غ ه	428571.9
2	ب ف م و ي	000000.6
3	ت ث ج د ذ ل ن	700000.14
4	ظ ق ك	666667.8
5	و ا ي	.4166667

Finally, it is important to note that as far as we know, there are no phonetic confusion tables for Arabic like those available for English (e.g., Luce, 1986; Shattuck-Hufnagel & Klatt, 1979; Wickelgren, 1966). Interestingly, however, Bailey and Hahn (2005) have forcefully argued that measures of similarity based on theoretically motivated phonetic features, as we have applied here, are superior to similarity measures based on confusability from speech perception, speech production, and short-term memory. Therefore, we feel confident that the current phonetic similarity matrix can serve as the basis for further explorations either within a language (Kishon-Rabin & Rosenhouse, 2000) or across languages (Boudelaa, 2018; Khattab, 2002).

Motoric Letter Similarity

Our ability to generate similar shapes with different limbs or execution modes suggests the existence of a relatively abstract, effector-independent level of representation that specifies the forms of letters (Keele, 1981; Rapp & Caramazza, 1997). If this is so, then language users must somehow develop a motoric scheme that represents information about the characteristics of the strokes required to write down a given allograph. Research into the written spelling performance of patients with dysgraphia strongly supports the involvement of multiple representational types, including a relatively abstract, effector-independent representational level that specifies the features of the component strokes of letters (Rapp & Caramazza, 1997). Specifically, individuals with dysgraphia seem to make well-formed letter substitution errors in written spelling, such as writing "F-A-P-L-E" for TABLE, while correctly spelling the target word as [ti, ei, bi, el, i]). Similarly, neuroimaging research suggests that the motoric features of letters activate significant portions of the brain in the left intraparietal sulcus and in areas previously associated with spelling processes (Rothlein & Rapp, 2014).

Given the importance of understanding the content of motor plans used to execute letter writing, we sought to develop a motoric letter similarity matrix for Arabic letters and their allographs based on 26 stroke features we established to be necessary to uniquely identify each of 100 letter allographs of Arabic.[1] We used 10 generic features to capture the visuospatial characteristics of each allograph in terms of a set of strokes. Accordingly, for each letter allograph, we specified the number of strokes (1 to 5) required to create it and the shape of those strokes (i.e., line, curve). When the stroke was a line, we specified its shape as downward- or upward-directed and its orientation, horizontal or vertical. When the stroke was a curve, we defined its shape (clockwise or anti-clockwise). We also included the number and position of the dots as well as the overall shape of the allograph and the number of angles it contained. Finally, we determined whether the allograph's main part was above or below the line and whether its overall shape was a half or full loop with no dots. The combination of these features allowed us to quantize each of the 100 letter allographs into a 26-element vector that captured the motor scheme necessary to create it. These vectors, accessible at https://osf.io/v2gb7/, were then submitted to a hierarchical clustering analysis with a view to determining the similarity structure underlying the motor plans of the different allographs. The dendrogram in Fig. 3 displays the clusters yielded by the nearest-neighbor method.

Using the gap statistic suggests that the data optimally cluster into 12 groups with a maximal value of 0.40. The average within-cluster SS is 16.46, while the total SS is 418.62 and the between-cluster SS is 221.07, thus suggesting a high degree of consistency within clusters. Table 6 displays the members of each cluster along with the associated within-cluster SS.

According to Table 6, a number of motoric features seem to underlie the way in which the 100 Arabic allographs used here cluster. Specifically, these are the presence and to some extent the number and position of the dots, as well as the presence and shape of a loop. Thus, for instance, the six members of Cluster 12, ي يـ ق ـقـ ـقـقـ share two dots, and four of them exhibit a clockwise downward-directed loop. Similarly, the seven members of Cluster 10, ن ن feature a single dot above the allograph, while those of Cluster 5 ث ڎ share the three dots above the allograph itself. The importance of the presence and number of dots in this context is that they define whether the abstract motoric program required to write down a letter allograph can be completed with or without lifting the pen: When a dot is present, the letter allograph cannot be written without lifting the pen. Another dimension of similarity arising from Cluster 1, ذ ـل کـ دذكـ ـخـ ـخـ ـحـ ـحـ جـ ابـ يـ ي, is the presence of an angle, which can be either a right angle, as in ي يـ ذ ابـ ا كـل, or an acute angle, as in جـ ـحـ ـحـ ـخـ ذ د ـخـ کـ. A final example is Cluster 9, ـهـ , where the presence of a closed loop in all allographs save appears to underlie the motoric similarity of this group of allographs. One obvious reason the allograph clusters with this group is the presence of the line segment that it shares in shape and orientation with and in shape only with .

Overall, then, there is a clear sense in which one might claim that similarity in terms of the characteristics of the strokes—number, orientation, and direction—that are required to produce the different allographs has a significant weight in the structure of each cluster. The viability of the present matrix as a measure of similarity between the motoric plans required to write each letter allograph is consistent with the performance of patients with dysgraphia as described by Nashaat, Kilany, Hasan, Helal, Gebril, and Abdelraouf (2016). Some of these patients made letter substitution errors in writing (e.g., دأيت for أيت), where the downward-directed stroke that starts above the "discontinuous" line and ends with a straight stroke on the line –ـد– substitutes for a downward-directed stroke that begins on the line and ends underneath it, - –. Further research is needed to examine the extent to which the motoric plan of allograph writing maps onto the neurocognitive domains of Arabic processing.

Conclusion

We present new data on the frequencies and similarities of Arabic letters and their allographs in the visual, phonetic, and motoric domains. These sets of frequencies of Arabic letters and their allographs, which are based on a 40-million-word corpus, comprise the only frequencies of letter allographs available for MSA. The visual similarity matrix is based on ratings collected from untimed responses of 125 participants to clearly presented allographic variants of the same letter. This methodology preempts serious issues likely to be inherent in matrices formed from data generated in atypical reading conditions, using, for example, speeded naming or degraded presentation conditions. Our visual similarity

[1] The reason we did not use the 110 allographs used in Experiment 1 is that it was not always easy to translate the letter shapes as defined by visuospatial features into an appropriate stroke set. This difficulty stems from the fact that the letter allographs we left out, آ أ إ ا أ ؤ ؤ ـئـ ـئـ ـئـ, were all carrier letters for *hamza* and had identical shapes to allographs we have included in this study.

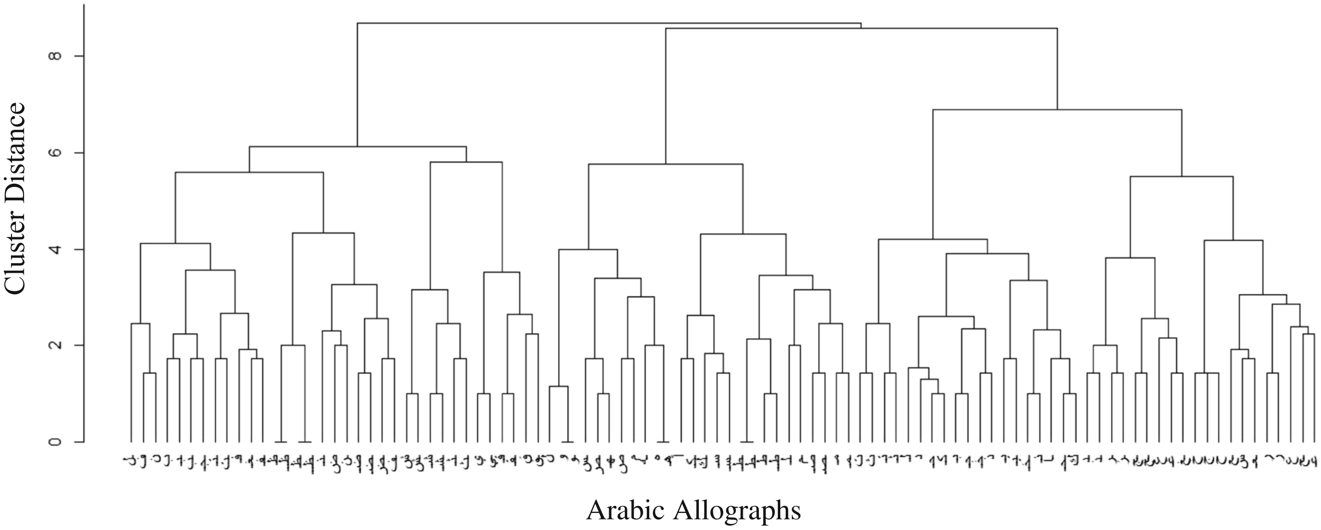

Fig. 3 Dendrogram of 100 Arabic letter allographs based on the motor scheme needed to produce them in writing

builds on and significantly extends previous findings in the literature (e.g., Wiley et al., 2016). The phonetic similarity matrix is based on theoretically motivated major phonetic/phonological class features, an approach that has recently been demonstrated to be efficient in identifying cognitively relevant similarities while at the same time significantly avoiding spurious task-specific similarities that characterize similarity metrics based on the perception of speech in noise (Bailey & Hahn, 2005). Finally, the motoric similarity matrix is based on a set of stroke features necessary to implement each letter and its allographs. This sort of similarity matrix is not very common across languages, and the only one we know of is the motoric similarity matrix developed for English (Rapp & Caramazza, 1997). Collectively, these new data will be a valuable tool for psycholinguistic research directed toward the study of letter stimuli and the effects and time courses of their visual similarity (Boudelaa et al., 2019; Carreiras et al., 2012; Gutiérrez-Sigut, Marcet, & Perea, 2019; Perea et al., 2010). They will be equally useful in informing cognitive neuropsychological reading research (Friedmann & Haddad-Hanna, 2012; Khwaileh, Body, & Herbert, 2014; Prunet, Béland, and Idrissi, 1998). Finally, since alphabet knowledge is consistently recognized as the strongest and most durable predictor of later literacy achievement (Jones, Clark, & Reutzel, 2012), the current results have clear practical implications for developing strategies to increase the effectiveness of teaching alphabet knowledge to young MSA learners by capitalizing on the similarity structure underlying the different letter and allograph groups (Mahfoudhi, Everatt, & Elbeheri, 2011; Perea, Abu Mallouh, & Carreiras, 2013; Taha, 2013).

Table 6 Optimal number of clusters based on motoric letter similarity as suggested by the gap method, the members of each class, and within-cluster sum of squares

Cluster number	Cluster members	Within-cluster SS
1	ابـ جـ حـ حـ خـ خـ د ذ ك ک ل ن يـ ي	22.36364
2	أ ک ك	10.00000
3	بـ ب تـ ت جـ غـ ففـ فـ نـ	15.00000
4	تـتـ ثـثـ	12.85714
5	ثـ ث	27.81818
6	جـحـحخـ ععـ ع ع	15.14286
7	ج خ د ذ غـغـ غ	6.40000
8	لمـهـ هـ هو و	6.80000
9	مـ	15.42857
10	فـ ن ن	16.00000
11	ظ ظ ظ	20.40000
12	ققـ ق قي ي	29.33333

Authors' note This research was funded by two United Arab Emirates University College of Humanities and Social Sciences grants to Sami Boudelaa (G00002367 and G00003158).

References

Abandah, G. A., Younis, K. S., & Khedher, M. Z. (2014). Handwritten Arabic character recognition using multiple classifiers based on letter form. In *Proceedings of the 5th IASTED International Conference on Signal Processing, Pattern Recognition, & Applications* (SPPRA 2008), Feb. 13–15, Innsbruck, Austria.

Asadi, I. A., Khateb, A., & Shany, M. (2017). How simple is reading in Arabic? A cross-sectional investigation of reading comprehension from first to sixth grade. *Journal of Research in Reading, 40 (S1)*, S1–S22. doi:https://doi.org/10.1111/1467-9817.12093.

Austin, W. M. (1957). Criteria for phonetic similarity. *Language, 33*, 538–543.

Bailey, T. M., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language, 52*, 339–362.

Boles, D. B., & Clifford, J. E. (1989). An upper- and lowercase alphabetic similarity matrix, with derived generation similarity values. *Behavior Research Methods, 21*, 579–586.

Boudelaa, S. (2018). Non-selective lexical access in late Arabic-English bilinguals: Evidence from gating. *Journal of Psycholinguistic Research, 47*, 913–930.

Boudelaa, S., & Marslen-Wilson, W. D. (2010). ARALEX: A lexical database for Modern Standard Arabic. *Behavior Research Methods, 42*, 481–487.

Boudelaa, S., Norris, D., Mahfoudhi, A., & Kinoshita, S. (2019). Transposed letter priming effects and allographic variation in Arabic: Insights from lexical decision and the same-different task. *Journal of Experimental Psychology: Human Perception and Performance, 49*, 729–757.

Bowey, J. A. (2005). Predicting individual differences in learning to read. In M. J. Snowling & C. Hulme (Eds.), The science of reading: A handbook (pp. 155–172). Oxford: Blackwell. doi:https://doi.org/10.1002/9780470757642.ch9.

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences, 18*, 90–98. doi: https://doi.org/10.1016/j.tics.2013.11.005.

Carreiras, M., Perea, M., & Abu Mallouh, R. (2012). Priming of abstract letter representations may be universal: The case of Arabic. *Psychonomic Bulletin and Review, 19*, 685–690. doi:https://doi.org/10.3758/s13423-012-0260-8.

Carreiras, M., Perea, M., Gil-López, C., Abu Mallouh, R., & Salillas, E. , (2013) Neural correlates of visual versus abstract letter processing in Roman and Arabic scripts. *Journal of Cognitive Neuroscience, 25*, 1975–1985. doi:https://doi.org/10.1162/jocn_a_00438.

Clatworthy, J., Buick, D., Hankins, M., Weinman, J., & Horne, R. (2005). The use and reporting of cluster analysis in health psychology: A review. *British Journal of Health Psychology, 10*, 329–358.

Cowell, J., & Hussain, F. (2002). A fast recognition system for isolated Arabic character recognition. Paper presented at the *IEEE Information Visualization Conference*, London. UK.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences, 9*, 335–341.

Dyson A. T., & Amayreh, M. M. (2000). Phonological errors and sound changes in Arabic-speaking children. *Clinical Linguistics & Phonetics, 14*, 79–109.

Evans, G. A. L., Lambon Ralph, M. A., & Woollams, A. M. (2017). Seeing the meaning: Top-down effects on letter identification. *Frontiers in Psychology*, *8*, 322. doi: https://doi.org/10.3389/fpsyg.2017.00322.

Fiset, D., Blais, C., Arguin, M., Tadros, K., Éthier-Majcher, C., Bub, D., & Gosselin, F. (2009). The spatio-temporal dynamics of visual letter recognition. *Cognitive Neuropsychology, 26*, 23–35. doi:https://doi.org/10.1080/02643290802421160.

Friedmann, N., & Haddad-Hanna, M. (2012). Letter position dyslexia in Arabic: From form to position. *Behavioural Neurology, 25*, 193–203. doi:https://doi.org/10.3233/BEN-2012-119004.

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes, 23*, 1-35.

Grainger, J., Dufau, S., Montant, M., Ziegler, J. C., & Fagot, J. (2012). Orthographic processing in baboons (*Papio papio*). *Science, 336*, 245–248.

Greenberg, J. H., & Jenkins, J. T. (1964). Studies in the psychological correlates to the sound system of American English. *Word 20*, 157–177.

Gusein-Zade, S. M. (1988). Frequency distribution of letters in the Russian language. *Problemy Peredachi Informatsii: Archive, 24*, 102–107.

Gutiérrez-Sigut, E., Marcet, A., & Perea, M. (2019). Tracking the time course of letter visual-similarity effects during word recognition: A masked priming ERP investigation. *Cognitive, Affective, and Behavioral Neuroscience*, *19*(4), 966–984. doi:https://doi.org/10.3758/s13415-019-00696-1.

Hahn, U., & Bailey, T. M. (2005). What makes words sound similar? *Cognition, 97*, 227–267.

Ibrahim, R., Eviatar, Z., & Aharon Peretz, J. (2002). The characteristics of the Arabic orthography slow its cognitive processing. *Neuropsycholgy, 16*, 322–326.

Jajuga, K., Sokolowski, A., & Bock, H.-H. (2002). Classification, clustering and data analysis: Recent advances and applications. Springer-Verlag, Berlin.

Jones, C. D., Clark, S. K., & Reutzel, D. R. (2012). Enhancing alphabet knowledge instruction: Research implications and practical strategies for early childhood educators. *Early Childhood Education Journal, 41*, 81–89.

Kamusella, T. (2017). The Arabic language: A Latin of modernity? *Journal of Nationalism, Memory & Language Politics, 11*, 117–145.

Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning*. Sthda.com. Ketchen.

Keele, S. W. (1981). Behavioral analysis of movement. In V. B. Brooks (Ed.), *Handbook of physiology: Vol. II. Motor control* (pp. 1391-1414). Baltimore: American Physiological Society.

Khattab, G. (2002). /l/ production in English-Arabic bilingual speakers.

Khorsheed, M. (2002). Off-line Arabic character recognition: A review. *Pattern Analysis & Applications, 5,* 31–45.

Khwaileh, T., Body, R., & Herbert, R. (2014). A normative database and determinants of lexical retrieval for 186 Arabic nouns: Effects of psycholinguistic and morpho-syntactic variables on naming latency. *Journal of Psycholinguistic Research, 43*, 749–769.

Kinoshita, S., & Kaplan, L. (2008). Priming of abstract letter identities in the letter match task. *Quarterly Journal of Experimental Psychology, 61*, 1873–1885. doi:https://doi.org/10.1080/17470210701781114.

Kishon-Rabin, L., & Rosenhouse, J. (2000). Development of speech assessment tests for Arabic-speaking children. *Audiology, 39*, 269–277.

Klatt, D. H. (1968). Structure of confusions in short-term memory between English consonants. *Journal of the Acoustical Society of America, 44*, 401–407.

Ladefoged, P. (1970). The measurement of phonetic similarity. *Statistical Methods in Linguistics, 6,* 23–32.

Li, W., & Miramontes, P. (2011). Fitting ranked English and Spanish letter frequency distribution in US and Mexican presidential speeches. *Journal of Quantitative Linguistics, 18*, 359. doi:https://doi.org/10.1080/09296174.2011.608606.

Liu, L., & Arditi, A. (2001). How crowding affects letter confusion. *Optometry and Vision Science, 78*, 50–55.

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon* (Unpublished doctoral dissertation, Dept. of Psychology), Indiana University, Bloomington, Indiana.

Maaly, I. A., Elobeid, A. R. & Ahmed, K. M. A. (2002). New Parameters for Resolving Acoustic Confusability Between Arabic Phonemes in A Phonetic HMM Recognition System. Ashurst Lodge : WIT Press, Vol. 1. 1- 85312-925-9.

Madi, M. (2010). A study of Arabic letter frequency analysis. http://www.intellaren.com/articles.

Mahfoudhi, A., Everatt, J., & Elbeheri, G. (2011). Introduction to the special issue on literacy in Arabic. *Reading and Writing, 24*, 1011–1018.

Mayzner, M. S., & Tresselt, M. E. (1965). Tables of single-letter and digram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements, 1*, 13–32.

Mohr, B., & Wang, W. (1968). Perceptual distance and the specification of phonological features. *Phonetica 18*, 31–45.

Mueller, S. T., & Weidemann, C. T. (2012). Alphabetic letter identification: Effects of perceivability, similarity, and bias. *Acta Psychologica, 139*, 19–37. doi:https://doi.org/10.1016/j.actpsy.2011.09.014.

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification, 31*, 274–295.

Nashaat, N. H., Kilany, A., Hasan, H. M., Helal, S. I., Gebril, O. H., & Abdelraouf, E. R. (2016). Dysgraphia in Egyptian dyslexic children: Related abilities to writing performance in Arabic. *Journal of Innovations in Pharmaceutical and Biological Sciences, 3*, 110–115

Parkinson, D. (2000). ArabiCorpus. http://arabicorpus.byu.edu/search.php.

Perea, M., Abu Mallouh, R., & Carreiras, M. (2010). The search of an input coding scheme: Transposed-letter priming in Arabic. *Psychonomic Bulletin and Review, 17*, 375–380.

Perea, M., Abu Mallouh, R., & Carreiras, M. (2013). Early access to abstract representations in developing readers: Evidence from masked priming. *Developmental Science, 16*, 564-573. DOI: https://doi.org/10.1111/desc.12052.

Petit, J.-P., Midgley, K., Holcomb, P. J., & Grainger, J. (2006). On the time course of letter perception: A masked priming ERP investigation. *Psychonomic Bulletin & Review 13*, 674-81.

Prunet, J. F., Béland, R., & Idrissi, A. (1998). Arabic consonantal root extraction in a deep dyslexic patient. *Brain and Language, 65*, 241–243.

Rapp, B., & Caramazza, A. (1997). From graphemes to abstract letter shapes: Levels of representation in written spelling. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1130-1152.

Roethlein, B. E. (1912). The relative legibility of different faces of printing types. *American Journal of Psychology, 23*, 1–36.

Rothlein, D. & Rapp, B. (2014). The similarity structure of distributed neural responses reveals the multiple representations of letters. *Neuroimage, 89*, 331–344.

Rothlein, D., & Rapp, B. (2017). The similarity structure of distributed neural responses reveals abstract and modality-specific representations of letters. *Journal of Vision, 13*, 786-786.

Schelonka, K., Graulty, C., Canseco-Gonzalez, E., & Pitts, M. A. (2017). ERP signatures of conscious and unconscious word and letter perception in an inattentional blindness paradigm. *Consciousness & Cognition*, *71*, 54–56. https://doi.org/10.1016/j.concog.2017.04.009.

Shattuck-Hufnagel, S., & Klatt, D. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior, 18*, 41–55.

Simpson, I. C., Mousikou, P., Montoya, J. M., & Defior, S. (2013). A letter visual-similarity matrix for Latin-based alphabets. *Behavior Research Methods, 45*, 431–439, https://doi.org/10.3758/s13428-012-0271-4.

Snowling, M. J., Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology, 81*, 1–23.

Stahl, D., Leese, M., Landau, S., & Everitt, B. S. (2011). Cluster analysis. Hoboken NJ: Wiley.

Taha, H. (2013). Reading and spelling in Arabic: Linguistic and orthographic complexity. *Theory and Practice in Language Studies, 3*, 721–727.

Taha, H., & Saiegh-Haddad, E. (2017). Morphology and Spelling in Arabic: Development and Interface. *Journal of Psycholinguistic Research, 46*, 27–38. doi:https://doi.org/10.1007/s10936-016-9425-3

Tibshirani, R. & Walther, G. (2005). Cluster validation by prediction strength. *Journal of Computational and Graphical Statistics, 14*, 511–528.

Tinker, M. A. (1928). The relative legibility of the letters, the digits, and of certain mathematical signs. *Journal of General Psychology, 1*, 472–496.

Treiman, R., Kessler, B., & Pollo, T. C. (2006). Learning about the letter name subset of the vocabulary: Evidence from US and Brazilian preschoolers. *Applied Psycholinguistics, 27*, 211–227.

Treiman, R., Levin, I., & Kessler, B. (2007). Learning of letter names follows similar principles across languages: Evidence from Hebrew. *Journal of Experimental Child Psychology, 96*, 87–106.

Treiman, R., Levin, I., & Kessler, B. (2012). Linking the shapes of alphabet letters to their sounds: The case of Hebrew. *Reading and Writing, 25*, 569–585.

Versteegh, K. (2014). The Arabic language. Edinburgh: Edinburgh University Press.

Wickelgren, W. A. (1966). Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America, 39*, 388–398.

Wiley, R. W., Wilson, C., & Rapp, B. C. (2016). The effects of alphabet and expertise on letter perception. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 1186–1203. https://doi.org/10.1037/xhp0000213.

Yakup, M., Abliz, W., Sereno, J., & Perea, M. (2014). How is letter position coding attained in scripts with position-dependent allography? *Psychonomic Bulletin & Review, 21*, 1600–1606. https://doi.org/10.3758/s13423-014-0621-6.

Yakup, M., Abliz, W., Sereno, J., & Perea, M. (2015). Extending models of visual-word recognition to semicursive scripts: Evidence from masked priming in Uyghur. *Journal of Experimental Psychology: Human Perception and Performance, 41*, 1553–1562. https://doi.org/10.1037/xhp0000143.

Yan, M. & Ye, K. (2007). Determining the Number of Clusters Using the Weighted Gap Statistic. *Biometrics, 63*, 1031-1037.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.