INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Voicing contrasts in the singleton stops of Palestinian Arabic: Production and perception

Nour Tamim, Silke Hamann

University of Amsterdam, The Netherlands
n.j.s.tamim@uva.nl, silke.hamann@uva.nl

Abstract

This study investigates the stop voicing contrast in Palestinian Arabic (PA) by examining Voice Onset Time (VOT) in both production and perception. An acoustic analysis of the recordings of 8 speakers showed that word-initial voiced stops in sentence context have an average VOT of -93 msec, and word-initial voiceless stops one of 29 msec. PA thus belongs, like most dialects of Arabic, to true voicing languages, i.e., languages with a contrast between voicing lead and short lag VOT.

We furthermore tested whether the phoneme /b/, without voiceless counterpart /p/ in PA, has similar VOT values to /d, dˤ/, which have voiceless counterparts /t, tˤ/. Similarly, we compared /k/, without counterpart /g/ in the PA dialect we investigated, to /t, tˤ/. For /b/ we found very similar VOT values to /d, dˤ/, while for /k/ we found a difference to /t, tˤ/, attributable to a general tendency of velars to have longer VOT than denti-alveolars. We thus found no evidence for a less contrastive realization of unpaired plosives in PA.

In a categorization experiment of the denti-alveolar phoneme pairs with the same 8 speakers, VOT proved sufficient as a perceptual cue, though f0 of the following vowel also influenced the categorization.

Index Terms: plosives, Palestinian Arabic, Voice Onset Time, perception, dispersion.

1. Introduction

Palestinian Arabic is a dialect spoken by the people of Palestine, and is part of the South Levantine Arabic dialect group, which contains the dialects of Arabic spoken in Lebanon, Syria, Jordan, and Palestine. The Palestinian dialect (henceforth: PA) shares features of Modern Standard Arabic (henceforth: MSA). With respect to singleton stops, the focus of the present study, MSA has the partly asymmetric inventory given in Table 1.

Table 1: *Singleton stops in MSA (adapted from* [1]*), with those in urban PA framed.*

bi-labial	denti-alveolar	pharyngeal denti-alv.	velar	uvular	glottal
b	d	dˤ			
	t	tˤ	k	q	ʔ

PA diverges from the MSA inventory in that the glottal stop phoneme has disappeared in many words, e.g. رأس 'head' MSA /raʔs/ vs. PA /raːs/. Furthermore, there are large dialectal differences in the correspondent of MSA /q/: it is a glottal stop in the urban dialect of PA, /k/ in the rural dialect, /g/ in the Bedouin dialect, and /q/ in the Druze dialect [2]. There is no native /g/ in PA; however, it can occur in the realization of some loanwords like French *gateau*. Similarly, no native /p/ exists; nevertheless, it can occur in some of the adaptations of /p/ in loanwords such as English *chips*. The latter was tested in a small-scale study with 10 PA speakers who each read 14 loanwords containing /p/ and where the results show that only 2% of these /p/ were realized as voiceless [p] [3]. The present study therefore considers /b/ and /k/ in PA as unpaired phonemes with respect to voicing.

In most Arabic dialects, voiced stops are produced with voicing lead and voiceless stops with short lag (but see e.g. Najdi Arabic [4] and Qatari [5], which contrast prevoiced with aspirated stops). However, there are big differences in the duration of the voicing lag of voiceless stops between experimental studies, as discussed by Alghamdi [6], who attributes this not only to dialectal variation, but also to the choice of participants and material employed. As for the last factor, studies differ (amongst other things) in whether they recorded isolated words (e.g. [6, 7, 8]), words in sentences (e.g. [5, 6, 10]), or spontaneous speech (e.g. [8, 11]), which makes a cross-comparison difficult, because the temporal characteristic of Voice Onset Time (VOT) is shorter in sentences than in isolated words [12].

Only one previous experimental study looked at voicing in PA: Adam [11] compared the VOT of initial alveolar stops in spontaneous speech by five agrammatic Palestinians with Broca's Aphasia to that of five control speakers. He found no overlap between VOT values for /t/ and /d/ in the control group, and concluded that VOT is a reliable acoustic cue for the distinction between these two stops in PA.

The present study investigates the voicing contrast of stops in the PA dialect, with the aim to fill a gap in the phonetic work on Arabic in general and the Palestinian dialect in particular. We investigate the urban dialect spoken in Nablus city, where MSA /q/ is a glottal stop. The glottal stop was excluded from our acoustic measurements because of its very short VOT (preliminary measures showed values between 10 and 15 msec, cf. also [7] for the Iraqi dialect).

Besides establishing VOT values of voiced and voiceless stops in PA, the present study is interested whether the unpaired stop /b/ in PA shows VOT values that differ from the voiced phonemes in the pairs /d/ – /t/ and /dˤ/ – /tˤ/. The same question holds for unpaired /k/: does its VOT realization differ from that of the paired voiceless phonemes /t/ and tˤ/. These questions are based on studies on dispersion (e.g. [13, 14]), showing that e.g. in languages with only one sibilant, this sound is usually realized with values in the center of the sibilant noise continuum, while languages with two sibilants avoid the middle of the continuum; they have one sibilant with higher and one with lower frequency noise, i.e. a dispersed inventory. Recent findings by Olson and Hayes-Harb [15] have not supported a

http://dx.doi.org/10.21437/Interspeech.2021-1079

realization in the center of the VOT continuum for unpaired plosives in Arabic: their six Saudi Arabian Arabic speakers had VOT values for /b/ that were similar to those of /d/, showing no shift towards the middle of the VOT continuum. Their study reports a similar non-centering effect for /k/, but they add that in the participants' native dialect, MSA /q/ is pronounced as [g], hence /k/ is not unpaired, which might have influenced their production of /k/ in MSA (the variety they were asked to produce in the experiment).

A third objective of the present study is the question whether the realizations of the pharyngealized pair /dˤ/ – /tˤ/ differ from that of the non-pharyngealized pair /t/ – /d/. This question is based on the study by Bellem [16] who found that the pharyngealized voiceless /tˤ/ in Saudi and Muslim Baghdadi dialects of Arabic had short lag VOT values that differed from the long lag of the voiceless /t/ and /k/.

A last objective is to test whether the acoustic findings are reflected in the perceptual behavior of the speakers, i.e., whether VOT is a strong cue in distinguishing voiced from voiceless phonemes in PA.

2. Acoustic study

2.1. Participants and material

Eight native Palestinian Arabic speakers participated in the study (mean age 27; four males). All subjects speak English as a second language and reported having normal speech and hearing. To ensure dialect homogeneity, all informants were from Nablus city.

The 24 stimuli were real words with one of the six stops of urban PA in word-initial or intervocalic position, see Table 2.

Table 2: *List of all stimuli; with voiced (top) and voiceless segments (bottom; all boldfaced).*

ba:b	'door'	**d**a:x	'daze'	**dˤ**ɑ:ʕ	'lost'
bu:r	'wild'	**d**u:r	'turn around'	**dˤ**u:ʔ	'light'
bi:r	'well'	**d**i:n	'religion'	**dˤ**i:f	'add'
mu**b**a:ħ	'permissible'	hu**d**a:	female name	sˤu**dˤ**ɑ:ʕ	'headache'
ta:ʒ	'crown'	**tˤ**ɑ:r	'flew'	**k**a:z	'kerosene'
tu:t	'blueberry'	**tˤ**u:b	'bricks'	**k**u:b	'cup'
ti:n	'figs'	**tˤ**i:n	'mud'	**k**i:s	'bag'
fu**t**a:t	'crumbs'	ħu**tˤ**ɑ:m	'wreckage'	zu**k**a:m	'cold'

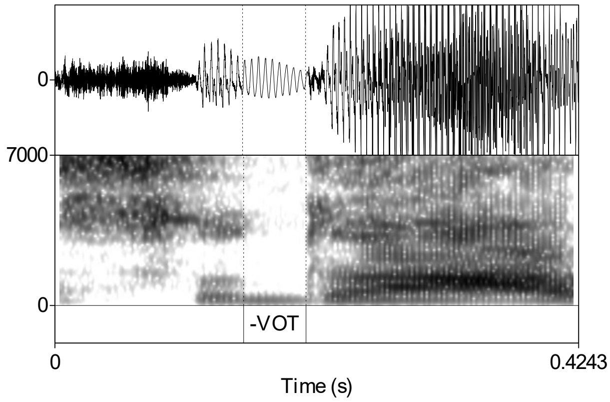

Figure 1: *Illustration of how the negative VOT was labelled, with an example token of* sˤu**dˤ**ɑ:ʕ.

In order to ensure that the participants produced PA rather than Standard Arabic, they were asked in PA to read each test word three times, inserted in the PA carrier sentence given in (1).

(1) /ʒi:t ʔaqraʔ __ w ʔru:ħ/ 'I came to read __ and go'

The test words were presented in written form in Arabic script. In total, this yielded 576 tokens (24 words x 8 speakers x 3 repetitions). Participants were recorded in a quiet room on a laptop with an in-built microphone.

VOT was measured as the time from the stop release to the beginning of voicing (as indicated by glottal pulsing). VOT boundaries were set and VOT durations were measured using Praat [17], for an illustration of a negative VOT in a token of the word sˤu**dˤ**ɑ:ʕ, see Figure 1.

2.2. Results

The VOT durations in msec for the PA stops split by word-initial and intervocalic position are given in Table 3 on the next page, the distribution of tokens, also split by position, are given as density plots in Figure 2 below.

Separate linear mixed-effects models for voiced and voiceless plosives were carried out in the program R [18, 19, 20]. In both, VOT was fixed factor, while random factors were segment (contrast coded for alveolar vs. labial or velar place, and for pharyngealized vs. plain alveolar) and position (initial

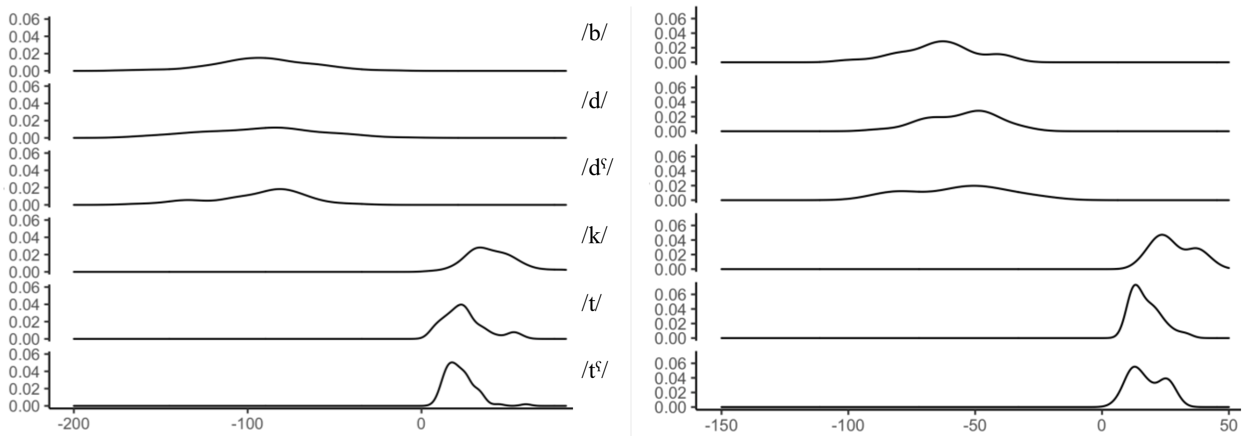

Figure 2: *Density plots showing the distribution of tokens, with VOT values (in msec) along the x-axis; left panel: word-initial position, right panel: word-medial position.*

Table 3: *Mean VOT durations (in msec) of PA stops split by position, standard deviation in brackets.*

voiced	b	d	$d^ʕ$
initial	-91 (28)	-93 (33)	-94 (27)
intervocalic	-64 (15)	-55 (14)	-57 (19)

voiceless	t	$t^ʕ$	k
initial	25 (12)	22 (9)	41 (15)
intervocalic	17 (6)	18 (7)	28 (8)

vs. medial). The models also accounted for segment and position as random slopes per speaker. For the voiced plosives, no significant difference could be found between the VOT values of the single segments, but a general difference between initial and intervocalic position ($p = 0.00091$). For the voiceless plosives, there was again an effect of position ($p < 0.0001$), but also one of segment: /k/ is realized with significantly different VOT values than /t, tˤ/ ($p < 0.0001$), but /tˤ/ does not differ significantly from /t/.

2.3. Discussion

The results of the acoustic study show that the voiced stops in PA have negative VOT (voicing lead) and the voiceless stops have short positive VOT (short lag), in line with results for most of the other Arabic dialects.

The VOT values of word-initial voiceless /t/ and /k/ in PA found in the present study are similar to the results of previous studies on other Arabic dialects (with the exception of the dialects of Iraq [7], Jordanian [9], Qatar [5], and Saudi (Najdi) [4], which show much longer VOT durations for voiceless stops). The prevoicing for the voiced PA stops /b/ and /d/, on the other hand, is longer than what has been reported for other Arabic dialects (e.g., for /b/ this ranged from -51 msec. in Qatari Arabic [5] to -85 msec. in Iraqi Arabic [7]).

With respect to the pharyngealized stops in PA, VOT values of /tˤ/ were not significantly different from /t/, and neither were the VOT values of /dˤ/ significantly different from those of /d/. We could thus not establish a different, third VOT region for the pharyngealized voiceless stops (contra [16]). The average VOT for word-initial PA /tˤ/ (22 msec) is similar to that in the dialects spoken in Iraq [7] and Lebanon [8, 21]. The average negative VOT for the word-initial /dˤ/ (-94 msec) resembles that of the Iraqi dialect [7] with -90 msec, and is longer than that of the Lebanese dialect [8, 21].

Looking at the difference between paired and unpaired phonemes, the unpaired /b/ in PA has slightly longer negative VOT values in intervocalic position than the two paired voiced stops, but this difference is not significant and furthermore in the opposite direction of what would be expected for an unpaired voiced segment, i.e., a more centralized realization with VOT values that are shifted in the direction of 0 msec. In initial position, the VOT realizations of /b/ were in-distinguishable from that of /d/ and /dˤ/. For /k/, the other unpaired phoneme in PA, we found significantly longer positive VOT realizations than for /t/ and /tˤ/, independent of position. This difference is again in the opposite direction of what would be expected for an unpaired phoneme, but in line with the cross-linguistic tendency that the VOT for voiceless stops increases the further back the place of articulation is in the vocal tract [22, 23].

3. Perceptual study

3.1. Participants and material

The 8 subjects of the acoustic study also participated in a perception experiment which was conducted three days after the production experiment. For the creation of the stimuli, a female native speaker was recorded in a sound-attenuated room, producing tokens of the four stops in word-initial position in the following words:

(2) تاب /ta:b/ ‘repented’ – داب /da:b/ ‘dissolved’
طاف /tˤa:f/ ‘circled’ – ضاف /dˤa:f/ ‘added’

Based on these recordings, ambiguous bursts with mean values for burst duration and burst amplitude were created for /d, t/ ($burst_1$) and for /dˤ, tˤ/ ($burst_2$).

Then, two stimuli continua, one for /d, t/ and one for /dˤ, tˤ/, were created. For the endpoint stimuli, one realization of each initial plosive in the words in (2) (excluding the burst) was used. To the endpoint plosives of /d, t/, $burst_1$ and the vowel and final plosive taken from a realization of /ta:b/ were added. To the endpoint plosives of /dˤ, tˤ/, $burst_2$ and a vowel and final fricative taken from a realization of / tˤa:f/ were added.

For the intermediate stimuli, a different strategy was used in the creation of stimuli with negative VOT values than for stimuli with positive VOT values. For stimuli with negative VOT values, parts of approx. 10 msec (two glottal pulses) were removed stepwise from the prevoicing of the voiced endpoint stimulus of each pair. For the intermediate stimuli with positive VOT values, parts of approximately 4 msec were removed stepwise from the voice lag of the voiceless endpoint stimulus of each pair. Larger steps were chosen for the part of the continuum with negative VOT values than for the part with positive VOT values to yield some balance between the two VOT types in the final stimuli set. The resulting set consisted of 10 stimuli on the /da:b/ – /ta:b/ continuum (seven having negative VOT), and 12 stimuli on the /dˤa:f/ – /tˤa:f/ continuum (8 having negative VOT). All stimuli on all continua were thus manipulated.

3.2. Procedure

The experiment was designed with an Experiment MFC script in Praat. Each stimulus was played three times in randomized order, resulting in a total of 66 stimuli. The stimuli were blocked for the two continua. Participants had to listen to the stimuli via headphones in a quiet room and had to categorize what they thought they heard by clicking on one of two answer categories. Answer categories were the two words of each continuum given in (2), presented in Arabic script. Instructions were given in PA and included the information that the two possible answers might not be used equally often by the participant. Participants could listen to a stimulus again if they were unsure, and had a self-timed break in the middle of the experiment.

3.3. Results

The category boundary between the non-pharyngealized denti-alveolar stops /d/ and /t/ in our experiments was at a VOT value of -18 msec, see Figure 3, and the boundary between /dˤ/ and /tˤ/ at -34 msec, see Figure 4.

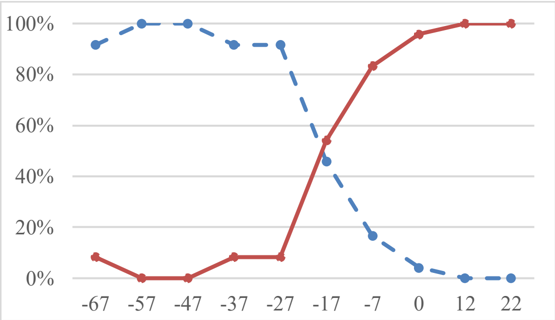

Figure 3: *Percentage categorization as* /d/ *(blue line) or* /t/ *(red line).*

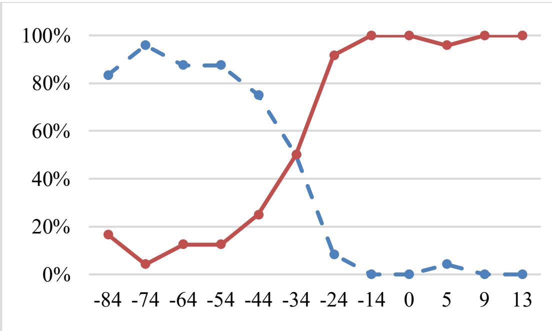

Figure 4: *Percentage categorisation as* /dˤ/ *(blue line) or* /tˤ/ *(red line).*

3.4. Discussion

The results of the perception experiment show that participants responded systematically to the VOT difference in the stimuli, and that VOT is an important acoustic cue in the identification of homorganic stops in PA. Stimuli with positive VOT values were identified almost 100% as voiceless on both the pharyngealized and the plain denti-alveolar continua. The category boundary between /d/ and /t/ at -18 msec is close to the boundary between /dˤ/ and /tˤ/ at -34 msec. Surprising is that these boundaries are not closer to 0 msec, i.e. that stimuli with short negative VOT values triggered categorization as voiceless /t/ and / tˤ/.

A first explanation for the unexpected location of the category boundaries could be the fact that the stimuli set contained more stimuli with negative VOT values (70% for the plain denti-alveolars, and 67% for the pharyngealized ones). It is therefore possible that participants, despite explicit instruction, tried to use both answer categories equally often. Such behavior could explain the boundary location, but cannot account for another observation in our data, namely that stimuli with a VOT of -67 msec on the /t/ – /d/ continuum were categorized as /t/ in 10% of the cases. Similarly, the identification of the stimuli on the left part of the /dˤ/ – /tˤ/ continuum (with negative VOT values) was not 100% voiced: /dˤ/ with long prevoicing did not always evoke a voiced category by listeners. These findings suggest that there is another acoustic cue that played a role in the identification experiment besides VOT.

The only cue in the stimuli that could explain a systematic preference for voiceless plosives even in the very voiced regions of the two continua is the following vowel. For both continua, the vowels (and rest of the words) were taken from recordings where the vowel was preceded by a voiceless stop (/ta:b/ and / tˤa:f/). Cross-linguistic studies have shown that voicing of preceding consonants determines the f0 of the following vowel [24, 25, 26, 27]: f0 falls after voiceless consonants and shifts upward after voiced consonants. An additional measurement we performed at the onset of the vowel in the stimuli showed indeed that f0 is higher (on average 29 Hz) after the voiceless denti-alveolars than after the voiced ones: /tˤa:f/ with 270 Hz vs. /dˤa:f/ with 241 Hz, and /ta:b/ with 265 Hz vs. /da:b/ with 236 Hz (see also similar findings of higher f0 onset frequency in post-voiceless vowels for Qatari Arabic [5] and Najdi Arabic [4]).

We therefore assume that high f0 at vowel onset was used as additional cue by the participants in our perception experiment, partly overriding the information provided by the VOT values.

4. Conclusions

This study is the first acoustic study to explore VOT in the production and perception of the PA dialect of Arabic. The acoustic results show that PA, like most other dialects of Arabic, is a prevoicing language: the voiced stops are produced with voicing lead while the voiceless stops are produced with short-lag, i.e. short positive VOT values.

With respect to the gaps in the plosive inventory of PA and their possible impact on the acoustic realization, we found no evidence for more centralized values on the VOT continuum for the unpaired phonemes /b/ and /k/. It seems that the existence of at least two phoneme pairs that are contrasting in a phonological feature such as [±voice], and the mapping of the two feature specifications onto two distinct regions on the VOT continuum is transferred to the unpaired phonemes /b/ and /k/ that do not contrast on this continuum (cf. [15] for a similar argument). Future studies need to shed light on the question whether there is also phonological evidence for a feature like [+voice] being shared by the PA phoneme class /b, d, dˤ/ and/or a feature like [–voice] being shared by /t, tˤ, k/.

Another interesting factor for future acoustic and phonological studies is the possible PA adaptation of the plosives /p/ and /g/ in recent loanwords from e.g. English, and whether the gaps in the PA inventory are filled by loan phonemes. Preliminary findings [3] did not confirm these expectations; however, more detailed studies on this point are necessary.

The current study also shed light on the perception of plosives in PA. The results of the categorization task with pharyngealized and non-pharyngealized denti-alveolars showed that VOT is the major perceptual cue in the identification of voiced and voiceless PA stops, but that the f0 of the following vowel seems to play an additional role in the perception of the voicing contrast. Future perception studies need to investigate more systematically the weighting of these two and possible further perceptual cues for voicing in PA but also in other Arabic dialects.

5. Acknowledgements

We would like to thank all our PA participants for taking part in our production and perception experiment.

6. References

[1] M. Embarki, “Phonetics”. In: Owens (ed.) *The Oxford handbook of Arabic linguistics*. pp. 23–44. Oxford: Oxford University Press, 2013.

[2] M. Jarrar, N. Habash, F. Alrimawi, D. Akra, and N. Zalmout, “Curras: An annotated corpus for the Palestinian Arabic dialect.” *Language Resources and Evaluation*, 2016, pp. 1–31.

[3] N. Tamim, “English /p/ in Palestinian Arabic loanwords.” Unpublished term paper in the Master’s programme General Linguistics of the University of Amsterdam, 2017.

[4] N. Al-Gamdi, J. Al-Tamimi, and G. Khattab, “The Acoustic Properties of Laryngeal Contrast in Najdi Arabic Initial Stops.” In: Calhoun, Escudero, Tabain and Warren (eds.) *Proceedings of the 19th ICPhS*. Canberra: Australasian Speech Science and Technology Association; 2019, pp. 2051–2055.

[5] V. Kulikov, “Laryngeal contrast in Qatari Arabic: Effect of speaking rate on voice onset time.” *Phonetica*, 2020, vol. 77, pp. 163–185.

[6] M. Alghamdi, *Analysis, synthesis and perception of voicing in Arabic*. PhD thesis, University of Reading, 1990.

[7] S. H. Al-Ani, *Arabic phonology: An acoustical and physiological investigation*. The Hague: Mouton, 1970.

[8] J. Al-Tamimi and G. Khattab, “Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops.” *Journal of Phonetics*, 2018, vol. 71, pp. 306–325.

[9] F. Mitleb, “Voice onset time of Jordanian Arabic stops.” *The Journal of the Acoustical Society of America*, 2001, vol. 109, no. 5, p. 2474.

[10] J. E. Flege and R. Port, “Cross-language phonetic interference: Arabic to English.” *Language and Speech*, 1981, vol. 24, pp. 125–146.

[11] H. Adam, “VOT-analysis: The production of stops by agrammatic Palestinians.” *International Journal of Linguistics*, 2012, vol. 4, pp. 300–308.

[12] L. Lisker and A. Abramson, “Some effects of context on voice onset time in English stops.” *Language and Speech*, 1967, vol. 10, no. 1, pp. 1–28.

[13] J. Liljencrants and B. Lindblom, “Numerical simulation of vowel quality systems: the role of perceptual contrast.” *Language*, 1972, vol. 48, pp. 839–862.

[14] P. Boersma and S. Hamann, “The evolution of auditory dispersion in bidirectional constraint grammars.” *Phonology*, 2008, vol. 25, no. 2, pp. 217–270.

[15] D. E. Olson and R. Hayes-Harb, “Voice onset time in Arabic and English stop consonants.” *Al-Arabiyya*, 2019, vol. 52, no. 1, pp. 29–48.

[16] A. Bellem, “Towards a Comparative Typology of Emphatics: Across Semitic and into Arabic Dialect Phonology.” PhD diss., University of London, 2007.

[17] P. Boersma and D. Weenink. *Praat: doing phonetics by computer* [Computer program]. 2018, version 6.0.43, http://www.praat.org/

[18] R Development Core Team. *R: a language and environment for statistical computing*. 2016, version 3.4.3. www.r-project.org

[19] D. Bates, M. Mächler, B. Bolker and S. Walker. “Fitting linear mixed-effects models using lme4.” *Journal of Statistical Software*, 2015. http://arxiv. org/abs/1406.5823

[20] A. Kuznetsova, P. B. Brockhoff & R. H. B. Christensen. “lmerTest Package: Tests in Linear Mixed Effects Models”. *Journal of Statistical Software*, 2017, vol. 82, pp. 1–26, https://doi.org/10.18637/jss.v082.i13

[21] G. H. Yeni-Komshian, A. Caramaza, and M. S. Preston, “A study of voicing in Lebanese Arabic.” *Journal of Phonetics* 1977, vol. 5, no. 1, pp. 35–48.

[22] L. Lisker and A. Abramson, “A cross-language study of voicing in initial stops: Acoustical measurements.” *WORD*, 1964, vol. 20, no. 3, pp. 384–422.

[23] T. Cho and P. Ladefoged, “Variation and universals in VOT: Evidence from 18 languages.” *Journal of Phonetics,* 1999, vol. 27, pp. 207–229.

[24] A. S. Abramson and L. Lisker, “Relative power of cues: F0 shift versus voice timing.” V. Fromkin (ed.) *Phonetic linguistics: Essays in honor of Peter Ladefoged*, pp. 25–33. Orlando: Academic Press, 1985.

[25] I. Lehiste and G. E. Peterson, “Some basic considerations in the analysis of intonation.” *The Journal of the Acoustical Society of America*, 1961, vol. 33, no. 4, pp. 419–425.

[26] D. Whalen, A. Abramson, L. Lisker, and M. Mody, “F0 gives voicing information even with unambiguous voice onset times.” *The Journal of the Acoustical Society of America*, 1993, vol. 93, no. 4, pp. 2152–2159.

[27] J. P. Kirby, and D. R. Ladd, “Effects of obstruent voicing on vowel F 0: Evidence from ‘true voicing’ languages.” *The Journal of the Acoustical Society of America*, 2016, vol. 140, no. 4, pp. 2400–2411.