OBSTRUENT VOICING AND LARYNGEAL FEATURES IN ARABIC

Abdulrahman Dallak[1,2], Ghada Khattab[2], Jalal Al-Tamimi[3]

[1]Jazan University, KSA, [2]Newcastle University, UK, [3]Université Paris Cité, LLF, CNRS, France
adallak@jazanu.edu.sa, ghada.khattab@ncl.ac.uk, jalal.al-tamimi@u-paris.fr

ABSTRACT

The paper investigates the laryngeal specification of obstruents in Jazani Arabic and whether fundamental frequency ($f0$) interacts with consonant voicing. It presents an acoustic analysis of voice onset time (VOT) and $f0$ perturbations. It is hypothesized that $f0$ perturbations should be predictable from the patterning of VOT in that voicing lead should induce a lowering effect on $f0$ and long lag should induce an increasing effect on $f0$. Twenty Jazani adult speakers produced target words with voiced and voiceless stops in controlled carrier sentences. VOT and $f0$ estimates were measured, the latter dynamically from the midpoint of the closure to the midpoint of the vowel at 10 ms intervals. The results show that VOT patterns in Arabic signal two distinct laryngeal specifications and that $f0$ perturbations appear to be predictable from the patterning of VOT. The results are then discussed in light of Laryngeal Realism and how they inform phonological representation.

Keywords: Jazani Arabic, voicing contrast, Laryngeal Realism, $f0$ perturbation

1. INTRODUCTION

It has long been observed that voiced stops /b,d,g/ in voicing languages such as Russian and Dutch are characterised by robust voicing lead whilst voiceless stops /p,t,k/ in aspirating languages such as English and German are characterised by robust long lag [1, 2]. The former implicates the specification of [voice] on the voiced series whilst the voiceless series in the same (voicing) language are unspecified. For the latter case, this implicates the specification of [spread glottis] on the voiceless series whilst the voiced series are unspecified. This observation is made formally explicitly in the so-called Laryngeal Realism (LR) [3, 4]; an experimental approach mapping VOT realisations into phonological features.

1.1. Laryngeal feature beyond VOT

While VOT patterning has taken centre stage when it comes to its role in laryngeal representation, $f0$ perturbation has been argued to also participate in the specification of laryngeal contrasts. In fact, recent studies [5] have shown that some languages appear to replace the VOT distinction with other emerging cues such as $f0$ and closure duration to maintain the laryngeal contrast. According to [6], the stops in Swiss German are realised as lenis stops [b̥,d̥], fortis unaspirated stops [p,t], and aspirated stops [pʰ,tʰ]. Following an investigation of the voicing contrast using VOT, stop closure and $f0$, the authors found that the (voicing) contrast in Swiss German stops is primarily signalled by closure duration for all the three categories, but not VOT. They reported that the unaspirated categories are accompanied by $f0$ raising, resembling that of prototypical voiceless stops. This shows that while VOT did not signal the voicing contrast in Swiss German stops, closure duration and $f0$ exhibited a robust pattern, reliably differentiating the three categories.

[5] examined the role of $f0$ in French and Italian (voicing languages) where the voiced stops are specified for [voice] and hence expected to induce a lowering effect on $f0$ while the voiceless stops are unspecified and expected to exhibit an inert pattern similar to the nasal baseline. The results depicted that $f0$ during the closure phase of the voiced stops was statistically more lowered than the nasal baseline in both languages, and following voiceless stops $f0$ was significantly raised in both languages. But what was unexpected is that $f0$ at the vowel onset and even midpoint was higher following the voiceless series given that French and Italian are both voicing languages, and based on LR, their voiceless series are laryngeally unspecified. Even more surprisingly, the authors reported that $f0$ at the vowel onset following voiced stops was statistically indistinguishable from the nasal baseline. Kirby et al. attribute the raising pattern of $f0$ after the unspecified voiceless stops to aerodynamic factors which inhibit phonation and which are independent from VOT and segment phonological voicing

In relation to Arabic, while the dialects surveyed in the literature such as Lebanese [7], Qatari [8], Najdi [9] etc. vary on what is encoded for the voiceless series, they are in agreement when it comes to the voiced series. That is, voiceless stops in Arabic have been observed to exhibit positive VOT with a varying degree of length within the aspiration range whilst voiced stops have been observed to exhibit voicing lead [7, 8, 9]. According to LR, Arabic stops appear to be over-specified. Both voicing categories of stops are specified simultaneously; [voice] for the voiced series and [spread glottis/tense] for the voiceless series [7, 8, 9, 10]. But whether both VOT and $f0$ exhibit a similar (or dichotomous) pattern and can be predictable from each other in Arabic is the focus of the present study.

1.2. [tense] vs. [spread glottis] characteristics

The feature [tense] was proposed in [11] to account for the aspiration contrast with an eye on the participation of the other cues in signaling the voicing/aspiration contrast [12]. [tense] is not necessarily accompanied by long VOT but explicitly presupposes that the voicing contrast is still maintained by other acoustic cues such as burst duration, intensity, voicing%, etc. In contrast, [spread glottis] was first proposed to signal the aspiration contrast per se [13]. The scope of each feature is still debatable among phoneticians and phonologists [2, 14, 15, 16], but it is crucial to highlight that [tense] does not necessarily implicate (robust) aspiration.

2. THE PRESENT STUDY

Given (i) the dichotomous pattern of VOT and $f0$ in Swiss German, French, and Italian, (ii) the unexpected raising pattern of $f0$ following lenis and unaspirated stops in Swiss German, and (iii) the unexpected raising pattern of $f0$ following the unspecified voiceless stops in voicing languages (i.e., French and Italian), the present study aims to investigate the mapping between these two acoustic correlates in the Jazani Arabic dialect to assess whether both VOT and $f0$ exhibit a similar pattern and can be predictable from each other. In order to examine the nuances of $f0$ pattern, it is crucial to utilise a nasal baseline as was done in [6, 5, 17].

2.1. Methods

2.1.1. Participants

Twenty male subjects from Jazan (Saudi Arabia), aged 22 to 30 and with no reported speech or hearing disorders, were recruited for the study. The subjects' speech was recorded using Tascam Dr-05 Linear PCM in a sound-proof booth. Subjects were asked to read the sentences out loud twice while being recorded. The recordings were digitised at 44.1 kHz sampling rate, with 16-bit quantization and in a mono channel.

2.1.2. Materials

Disyllabic words containing the target sounds /b,d,g,t,k/ and /m/ were inserted in varying natural sentences. /m/ was used as nasal baseline for examining consonant voicing effect on $f0$ [6]. The study only looked at VOT and $f0$ patterns of post-vocalic word-initial stops in three vowel contexts /iː,aː,uː/. This experiment generated 600 tokens for the analysis (20 subj$\times$ 3 vowel $\times$ 2 rep $\times$ 5 cons).

2.1.3. Acoustic analysis

Data were manually transcribed using a newly developed transliteration system for Arabic (ATR), which was then forced aligned using the WebMAUS Arabic via the Munich Automatic Segmentation (MAUS) [18] as in Fig 1. Two acoustic correlates were measured in this experiment: (i) VOT, and (ii) $f0$. VOT was manually measured for post-vocalic word-initial stops from the onset of the release burst to onset of the vowel using visual inspection of oscillogram and broadband spectrograms [19]. The onset of the release burst was identified by an abrupt vertical "spike" in the spectrogram following stop closure. The onset of the vowel was identified by the first vertical striations on the spectrogram. For voiced stops, VOT represents closure plus release duration. $f0$ estimates were measured dynamically from the midpoint of the closure (even for the nasal) to the midpoint of the vowel at 10 ms intervals using `Praat`'s version of `VoiceSauce` [20] For voiceless stops, $f0$ estimates were measured from vowel onset to midpoint.

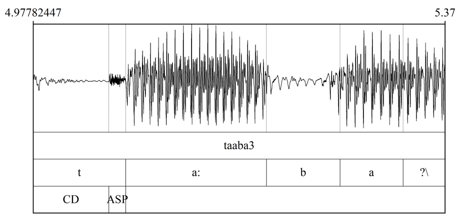

Figure 1: Schematic TextGrid showing how VOT was measured for /taːbaʕ/ 'followed'

2.1.4. Statistical analysis

Statistical analysis was carried out in `R` [21]. We modelled the difference between voiced vs. voiceless stops using estimation statistics [22] based on median difference, 95% CI, and effect size.

The CIs coupled with effect size are derived from 5000 bootstrap resampling, bias-corrected and accelerated. The effect size used Cohen's *d*. For modelling $f0$ contour, we used a generalised additive mixed model (GAMM) using `mgcv` [23]. It was fitted to the (Hz) value of $f0$ with voicing by phoneme interaction as parametric term, and following the recommendation of [24] for random effect. The results reported here are for the context /aː/. The model used 'scat' family [25], and accounted for autocorrelated residuals. GAMM graphs were produced using `tidyverse` [26], and `tidymv` [27].

3. RESULTS

3.1. VOT MODEL

Fig. 2 presents an estimation plot using two different y-axes. The y-axis on the left side represents the VOT value in ms whilst that on the right side represents effect size of the median difference Δ between the two voicing categories. The grey curve indicates the resampled distribution of Δ, median, and 95% CI, given the observed data. The results

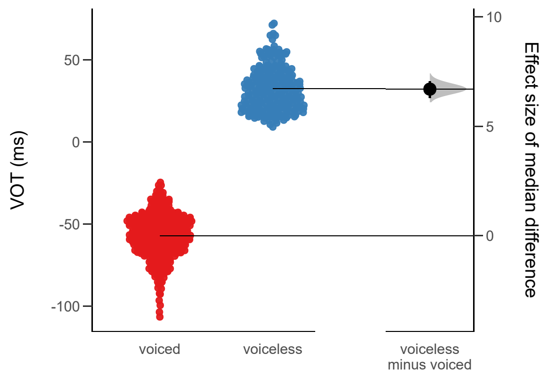

Figure 2: Estimation plot showing the distribution of each voicing category coupled with effect size.

show that voiced stops exhibited an averaged VOT value of -75 ms ([b]= -57.9, [d]= -56.9, [g]= -56.5). In contrast, the voiceless series exhibited an averaged VOT value of 32.5 ms ([t]=31.0, [k]=34.1). The estimated difference between both categories is 90.5 ms [95% CI 87.6, 93.6], exhibiting a robust effect size d = 6.7 [95% CI 6.3, 7.06]. More specifically, for the difference between [t]-[d], the results exhibited a difference value of 87.9 ms [95% CI 84.6, 91.6]. For [k]-[g], the estimated difference was 90.5 ms [95% CI 87.6, 93.6].

3.2. $f0$ CONTOUR MODEL

Table 1 presents the GAM model summary of $f0$. It shows that voiced stops induced a significantly reduced pattern of -19.45 Hz averaged across all the intervals measured relative to the intercept which represents the nasal baseline /m/. For the voiceless stops, the model shows that voiceless stops induced a statistically non-significant pattern of -8.29 Hz averaged across all the intervals measured relative to the nasal baseline.

Predictors	Estimates
Intercept (nasal)	152.57^{***}
	(8.65)
voiced.stop	-19.45^{**}
	(6.21)
voiceless.stop	-8.29
	(6.41)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1: Summary of the GAM model fitted to $f0$

Fig.3 presents the estimates of $f0$ contour for voiced stops and the nasal baseline along with the difference curve. It shows that voiced stops induced a robust reduced pattern throughout the measured intervals from closure midpoint to vowel midpoint.

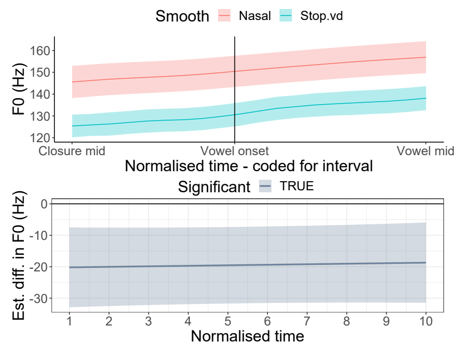

Figure 3: Estimated $f0$ splines for nasal baseline and vd stops coupled with a difference curve

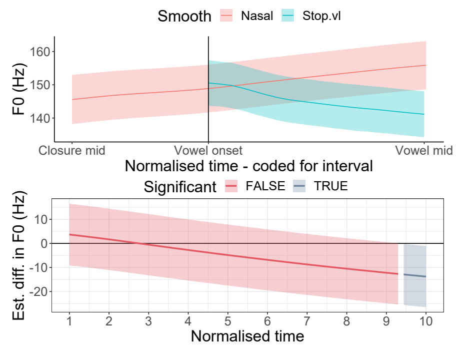

Figure 4: Estimated $f0$ splines for nasal baseline and vl stops coupled with a difference curve

Fig. 4 presents the estimates of $f0$ contour for voiceless stops and the nasal baseline along with the difference curve from vowel onset to midpoint. It shows that while the voiceless stops induced a raising pattern at vowel onset, this raising pattern is weak and negligible. Crucially, the difference curve presents no evidence for significant differences between the two trajectories except toward the vowel midpoint.

4. DISCUSSION

This study investigated two acoustic measures, VOT and $f0$ contour, to assess the laryngeal specification of stops in the Jazani Arabic dialect along with exploring whether VOT and $f0$ exhibit a predictable mapping pattern, given the cases addressed briefly in Introduction.

4.1. VOT and LR

According to the predictions of LR, robust voicing lead implicates the specification of [voice] on the voiced stops whilst long lag implicates the specification of [spread glottis] on the voiceless stops. The results show that voiced stops exhibited an average VOT value of -75 ms, congruent with the available literature on Arabic [7, 8, 9]. For the voiceless stops, the present results show that voiceless stops exhibited an average VOT value of 32.5 ms. Compared to the pattern of voiceless stops in other Arabic dialects such as Najdi and Qatari [9, 8], respectively, this observed value constitutes a short VOT value for the voiceless series. More specifically, in Najdi, voiceless stops showed a mean VOT value of 47.6 ms. Likewise, voiceless stops in Qatari Arabic showed a value of 55 ms. This long positive VOT values for Najdi and Qatari Arabic are similar to that of the aspirating languages such as German 67.33 ms [2] and English 56.05 ms [28].

The observed value of the voiceless series in the present dialect is instead relatively similar to the voiceless series in the Lebanese dialect, which is 28.62 ms [7], and that of the unspecified (lenis) stops in aspirating languages such as English (25 ms) [28].

Taken together (i) the absence of robust long lag on the voiceless series and (ii) the null pattern exhibited by the voiceless series on $f0$ contour, the present results demonstrate that voiceless stops in the current dialect are not specified with [spread glottis]. While this suggests that voiceless stops are unspecified, this paper is an excerpt from a larger project [29] where robust evidence is presented showing active participation of other cues in signalling the voicing contrast in the present dialect such as voicing%, harmonics, intensity, burst duration, etc. Thus, we propose that voiceless stops in the present dialect are specified with [tense], which in the current context indicates that the voicing contrast is upheld by other acoustic cues beyond VOT (and $f0$). The fact that researchers working within the LR tradition use [spread glottis] to signal robust aspiration while others [2] use [tense] to signal the same manifestation reveals that the scope of each feature is still controversial. In this study, we deviate from [spread glottis] due to the absence of robust aspiration and adopt [tense] due to the active participation of the other cues in signalling the voicing contrast (cf.[14, 12]).

4.2. $f0$ contour and LR

Given the specification of [voice] on the voiced series, it follows naturally that voiced stops should induce a robust lowering pattern compared to the nasal baseline [6, 5]. This robust pattern is observed in $f0$ contour results where the voiced stops exhibit a significant lowering pattern throughout all the measured intervals. For the voiceless stops, the pattern observed rules out the specification of [spread glottis] on the voiceless series; otherwise [spread glottis] should induce significantly higher $f0$ contour in relation to the baseline [1, 9]. Similar to the phonological mapping of VOT, the pattern exhibited by the voiceless stops on the $f0$ contour largely indicates a null effect and specification. This indicates that VOT and $f0$ (contour) exhibit a matching pattern and appear to be predictable from each other at least for the present dialect. This suggests that they provide evidence for a laryngeal matching specification. This is in line with the predictions of LR, which appear to be language-specific.

5. CONCLUSION

This study examined the acoustic realisation of voicing distinction in stops in the Jazani Arabic dialect. It showed that voiced stops are specified as [voiced], and voiceless stops are unspecified or at best specified with a lower numerical value for [tense] [1] building on an ongoing project [29]. It also showed that VOT and $f0$ contour exhibit a matching pattern given the appropriate feature specified for the voiceless series.

6. ACKNOWLEDGEMENTS

For AD, we acknowledge Jazan University for PhD scholarship. For JAT, this work is partially supported by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083), contributing to the IdEx Université Paris Cité - ANR-18-IDEX-0001.

7. REFERENCES

[1] J. Beckman, M. Jessen, and C. Ringen, "Empirical evidence for laryngeal features: Aspirating vs. true

voice languages," *Journal of linguistics*, vol. 49, no. 2, pp. 259–284, 2013.

[2] M. Jessen, "Phonetics and phonology of tense and lax obstruents in german," *Phonetics and Phonology of Tense and Lax Obstruents in German*, pp. 1–414, 1998.

[3] G. K. Iverson and J. C. Salmons, "Aspiration and laryngeal representation in germanic," *Phonology*, vol. 12, no. 3, pp. 369–396, 1995.

[4] P. Honeybone, "Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications," *The internal organization of phonological segments*, vol. 319, p. 54, 2005.

[5] J. P. Kirby and D. R. Ladd, "Effects of obstruent voicing on vowel f0: Evidence from 'true voicing' languages," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 2400–2411, 2016.

[6] D. R. Ladd and S. Schmid, "Obstruent voicing effects on f0, but without voicing: Phonetic correlates of swiss german lenis, fortis, and aspirated stops," *Journal of Phonetics*, vol. 71, pp. 229–248, 2018.

[7] J. Al-Tamimi and G. Khattab, "Acoustic correlates of the voicing contrast in lebanese arabic singleton and geminate stops," *Journal of Phonetics*, vol. 71, pp. 306–325, 2018.

[8] V. Kulikov, "Laryngeal contrast in qatari arabic: Effect of speaking rate on voice onset time," *Phonetica*, vol. 77, no. 3, pp. 163–185, 2020.

[9] N. Al-Gamdi, "Voicing contrast in najdi arabic stops: Implications for laryngeal realism," Ph.D. dissertation, Newcastle University, 2022.

[10] N. Al-Ghamdi, J. Al-Tamimi, and G. Khattab, "The acoustic properties of laryngeal contrast in najdi arabic initial stops," in *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS)*, S. Calhoun, P. Escudero, M. Tabain, and . P. Warren, Eds., Melbourne, Australia; Canberra, Australia, 2019, pp. 2051–2055.

[11] R. Jakobson, C. G. Fant, and M. Halle, *Preliminaries to speech analysis: The distinctive features and their correlates*. MIT press, 1952.

[12] H. Kim, G. N. Clements, A. Rialland, R. Ridouane, and H. van der Hulst, "The feature [tense]," *Features in phonetics and phonology: Unpublished work from George N. Clements and his colleagues*, pp. 159–178, 2015.

[13] C.-W. Kim, "A theory of aspiration," *Phonetica*, vol. 21, no. 2, pp. 107–116, 1970.

[14] R. Ridouane, G. N. Clements, and R. Khatiwada, "Language-independent bases of distinctive features," in *Tones and Features*, J. A. Goldsmith, E. Hume, and L. Wetzels, Eds. Berlin, Boston: DE GRUYTER, Oct. 2011, pp. 264–291.

[15] M. Halle and K. N. Stevens, "A note on laryngeal features," *Quarterly Progress Report, Research Laboratory of Electronics*, pp. 198–213, 1971.

[16] N. Chomsky and M. Halle, *The sound pattern of English*. ERIC, 1968.

[17] J. J. Ohala *et al.*, "Phonetic explanations for nasal sound patterns," in *Nasálfest: Papers from a symposium on nasals and nasalization*. Stanford University Language Universals Project Palo Alto, CA, 1975, pp. 289–316.

[18] J. Al-Tamimi, F. Schiel, G. Khattab, N. Sokhey, D. Amazouz, A. Dallak, and H. Moussa, "A romanization system and webmaus aligner for arabic varieties," in *13th Conference on Language Resources and Evaluation (LREC 2022)*, 2022.

[19] P. Boersma, "Praat: doing phonetics by computer [computer program]," *http://www. praat. org/*, 2020.

[20] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu, "Voicesauce: A program for voice analysis," *Energy*, vol. 1, no. H2, pp. H1–A1, 2010.

[21] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/

[22] J. Ho, T. Tumkaya, S. Aryal, H. Choi, and A. Claridge-Chang, "Moving beyond p values: data analysis with estimation graphics," *Nature methods*, vol. 16, no. 7, pp. 565–566, 2019.

[23] S. N. Wood, "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 1, pp. 3–36, 2011.

[24] M. Soskuthy, "Evaluating generalised additive mixed modelling strategies for dynamic speech analysis," *Journal of Phonetics*, vol. 84, p. 101017, 2021.

[25] M. Wieling, "Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between l1 and l2 speakers of english," *Journal of Phonetics*, vol. 70, pp. 86–116, 2018.

[26] H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. FranÃ§ois, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. MÃ¼ller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani, "Welcome to the tidyverse," *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, 2019.

[27] S. Coretta, *tidymv: Tidy Model Visualisation for Generalised Additive Models*, 2022, r package version 3.3.2. [Online]. Available: https://CRAN.R-project.org/package=tidymv

[28] G. J. Docherty, "The timing of voicing in british english obstruents," in *The Timing of Voicing in British English Obstruents*. De Gruyter Mouton, 2011.

[29] A. Dallak, "Arabic Obstruents: Laryngeal Contrast and Representation," Ph.D. dissertation, Newcastle University, forthcoming.