ARTICULATORY AND ACOUSTIC GRADIENCE IN SASAK WORD-FINAL STOPS /K, ʔ/

Jonathan Yip[1], Diana Archangeli[2]

[1]The University of Hong Kong, [2]University of Arizona
[1]yipjonat@hku.hk, [2]dba@email.arizona.edu

ABSTRACT

The sounds of a language are often assumed to exhibit some degree of physical categoricity. In this study, we examine the contrast between Sasak word-final /-k/ and /-ʔ/ by collecting audio and ultrasound for 17 /-k/ and 25 /-ʔ/ words from 9 talkers. A principle components analysis of tongue contours showed that talkers contrast velar and glottal articulations in the terms of dorsal height, while dip tests for bimodality indicate that only three talkers' articulations varied bimodally. F1 at vowel offset was the most distinctive acoustic cue distinguishing /k/ from /ʔ/, although this contrast occurred less consistently than the articulatory contrast. As with the PCA data, little bimodality was found in the acoustics. Further analysis showed categoricity across words depending on the talker. These results suggest that both acoustic properties and articulatory behaviors across and within talkers are rather gradient, challenging previous assumptions about phonetic uniformity.

Keywords: speech articulation, phonological contrast, phonetic variation, ultrasound, Sasak

1. INTRODUCTION

Much speech research has shown that sounds that behave phonologically as distinct sound units are subject to influences that can render their phonetic properties more continuous and gradient [2]. At the same time, talkers should produce phonologically contrastive sounds with a certain amount of physical categoricity (i.e. articulatory or acoustic invariants) in order for their percepts to be kept distinct [14] [4]. Phonetic variation that is attributable to differences between talkers ([9], [7]) and between phonological contexts ([5], [3]) has been shown to be relatively systematic.

A sometimes overlooked source of phonetic variation lies in the differences between the articulatory characteristics of contrastive sounds. Some sounds require articulatory movements that are more affected by coarticulatory influence than other sounds, as shown in Recasens & Rodríguez [13]. Productions of such sounds could thus obscure their degree of physical (articulatory and/or acoustic) differentiation from other sounds in the language. Sound segments that vary greatly may impose a level of phonetic gradience onto talker's productions, thereby reducing the overall degree of phonetic categoricity between phonological units.

In this study, we examine a phonological contrast in the Sasak language (Malayo-Polynesian; Lombok, Indonesia) involving morpheme-fina oral stops /k/ and /ʔ/. First, we investigate whether this sound contrast can be found in the articulations and acoustics of speakers' productions. We then explore whether these properties differ categorically and whether words containing these sounds exhibit clear phonetic differences according to the identity of the sound. For articulation, we collected ultasonic images of the tongue contour at the moment of stop constriction, and for acoustics, we examined phonetic cues that might distinguish /k/ from /ʔ/. If the articulatory behaviors or their corresponding acoustic cues vary categorically, then measures of their phonetic properties should fall into bimodal distributions, that is, an opposition between /k/-like versus /ʔ/-like articulatory and/or acoustic patterns. However, if such measures are not categorical, then it can be concluded that talkers fail to maintain strict, systematic distinctions between such sounds.

2. METHODS

2.1. Participants

Participants were nine native speakers of Sasak (six male, three female) who resided in Mataram, Lombok, Indonesia. All talkers also spoke Bahasa Indonesia and English, and none reported any speaking or hearing deficits. Participants' ages ranged from 19 to 33 (mean = 24.3±4.2 years).

2.2. Target items

Target items were 42 two- to four-syllable Sasak words ending in either /k/ or /ʔ/. In all cases, the vowel in the final syllable was /a/ and the preceding sound(s) was a bilabial stop or nasal+plosive sequence (/p, b, m, mp, mb/), e.g. *kepak* [kə.pak] and *begerambaq* [bə.gə.ram.baʔ]. This design ensured consistency in the physical position of the tongue (low dorsum and retracted root) among target items. 17 items ended in /-ak/, and 25 items ended in /-aʔ/. For prosodic consistency, each item was presented in the sentence frame: *Tulis kata ___ adeng-adeng.* [tu.lis ka.tə ___ a.deŋ a.deŋ] 'Write the word ___ slowly.' This frame was chosen so that the sound /a/ would follow each target word, thereby limiting the phrase's influence on each word-final stop articulation. Each item was presented in four iterations, resulting in a total of 168 productions per talker.

2.3. Procedure

Recordings were collected by the authors in a quiet classroom at the Mataram Lingua Franca Institute in Mataram, Lombok. During audio and ultrasound recording sessions, Sasak stimulus prompts were shown one at a time on a presentation laptop in front of the participant. One author monitored the quality of the live ultrasonic scan, while the other gave instructions to the participant and manually advanced through each stimulus.

Ultrasound video recordings were collected with the audio signal embedded directly into the file. Audio was captured with an omnidirectional, condenser earset microphone positioned at one side of the talker's mouth and then digitized with an analog-to-digital interface connected to the collection laptop. For video recording, ultrasonic scan data were captured using a ClarUs-EXT portable scan unit (Telemed) and a 2- to 4-Hz convex transducer probe. Ultrasonic data were fed via USB into a high-performance collection laptop running EchoWave II software [16], which constructed real-time ultrasonic image frames at a frequency of nearly 60 frames per second. The video and embedded audio stream were simultaneously written into a single file using screen capture software, with a 640×480 pixel image resolution and unlimited recording duration.

In order to physically stabilize the transducer into a midsagittal position underneath the lower jaw, a freely-poseable camera lighting arm was clamped onto the desk where the participant was seated, and two additional arms positioned to the sides and above the participant's head were used to provide padded forehead rests onto which the participant was instructed to lean against during data collection.

Audio and ultrasonic video were synchronized by aligning the acoustic release bursts of 9 to 12 productions of the post-alveolar click [k͡!] with their associated ultrasonic video frame, as produced by one of the authors at the end of each recording. The latency between the onset of the acoustic burst and the time of the image frame at which a significant downward motion of tongue-blade position was observed was computed for each click production, and the arithmetic mean of these latencies was used to adjust the timing of video image frames relative to the acoustic signal in each recording.

Frames for analyzing sound segments were identifed based on the acoustic recording in Praat [1]. The frame representing the articulation of each word-final stop was the frame occurring at or immediately before the acoustic onset of the [k]- or [ʔ]-constriction. This moment was the time at which the resonant formant structure for the preceding vowel disappeared, i.e. the offset of the preceding vowel [a]. The segment frames were then programmatically extracted for tongue contour tracing. Additionally, measures of the spectral properties (F1 and F2 in bark) and durations (in ms) of the preceding /a/ were extracted from the acoustic signal. Formant values were taken at 10-ms intervals from the first to final glottal pulse associated with [a]. Because the word-final Sasak stops of interest are typically released without audible cues, stop release cues were not analyzed. In each segment frame, the visible tongue contour was traced and exported in Cartesian coordinates using EdgeTrak software [10] and then later converted into polar coordinates relative to the scan origin, which was the physical location of the transducer.

2.4. Statistical analysis

Tongue contour data were analyzed using a principal components analysis (PCA) similar to that used by other authors [15] [8] [17]. Loadings for the first and second principal components were computed for each talker using the *princomp* function in R [12]. For each talker, articulatory (PC1, PC2 coefficients) and acoustic (F1, F2, duration) data were analyzed using linear mixed-effects regression models with the fixed effect of *Sound* (/k, ʔ/) and random intercepts for *Item*. For analyses of data pooled across talkers, random intercepts for *Talker* were also included. To test for bimodal distributions, PCA and acoustic data were submitted to Hartigan's dip test [6] using the R package *diptest* [11].

Table 1: LMER estimates of articulatory and acoustic measures for /k/ and /ʔ/ productions. Talker-specific results are reported by column, and results for data pooled across talkers are presented in the rightmost column. Grey cells indicate significant comparisons, and asterisks indicate the level of significance for the *p*-value indicated as follows: *** < 0.001 < ** < 0.01 < * < 0.05 < . < 0.1

measure	sound	S1	S2	S3	S4	S5	S6	S7	S8	S11	pooled
PC1	/k/	***9.26	***9.16	***9.47	***6.83	***7.54	*2.52	***6.76	***4.50	1.05	***6.34
	/ʔ/	-6.30	-6.23	-6.44	1.89	-5.18	-1.26	-4.71	-3.06	-0.71	-4.29
PC2	/k/	0.37	1.13	1.65	**3.01	0.36	0.03	0.27	***3.80	**2.88	***1.52
	/ʔ/	-0.25	-0.74	-1.12	-2.05	-0.25	-0.57	-0.15	-2.58	-1.96	-1.08
offset F1	/k/	6.44	7.97	5.84	5.92	6.86	6.40	7.57	6.38	7.29	6.74
	/ʔ/	***7.00	***8.57	***7.04	***6.81	***7.42	***7.03	**7.91	6.33	*7.85	***7.33
offset F2	/k/	***10.90	11.30	*10.23	10.33	10.47	10.47	10.53	10.05	11.33	10.62
	/ʔ/	10.45	**11.52	10.07	**10.52	*10.63	10.55	***11.07	***10.54	11.68	**10.78
duration	/k/	***142.0	**86.4	***85.7	91.7	109.2	**95.1	101.7	113.2	*84.2	***101.0
	/ʔ/	71.2	73.4	68.4	85.4	105.2	82.6	97.8	116.8	75.4	86.3

3. RESULTS

Results from the linear mixed-effects models for PC1 and PC2 coefficients, offset-F1 and -F2, and vowel-duration measures are reported in Table 1.

PC1 coefficients corresponded to dorsal height, and PC2 coefficients corresponded to tongue-root position. In general, /k/ had higher PC1 and PC2 values (higher dorsum, more advanced root) than /ʔ/ (Figure 1). LMER results for PC1 coefficients indicate a significant difference between /k/ and /ʔ/ articulations for each talker as well as overall. For PC2, results were less consistent among talkers, but the pooled analysis revealed a significant overall difference. A dip test for PC1 coefficients pooled across talkers indicated that a bimodal distribution was unlikely, as shown in Table 2, while within individuals, PC1 coefficients were bimodally distributed for only three talkers (S1, S3, S5). For PC2, coefficients were not bimodally distributed in the pooled and individual-talker analyses.

In the acoustic data, /k/ was generally produced with lower offset-F1, lower offset-F2, and longer preceding vowel duration than /ʔ/. Patterns for F2 values were less consistent, with five talkers producing lower offset F2 in /k/ than /ʔ/. This result suggests some variation in the degree of fronting for /k/-constrictions among talkers. None of the dip tests for F1, F2, and durational measures pooled across talkers indicated bimodal distribution, and for individual-talker analyses, only talker S4's F1 values and talker S1's durational values were distributed bimodially. None of the talkers' F2 measures had bimodal distribution.

In order to explore whether the articulations among word items fell into the expected categories,

Figure 1: PCA loadings for lingual contours at constriction onset by one talker (S4). The black line with dots shows the average contour, and the red contours show the range of PC1. The dashed blue contours indicate the range of PC2.

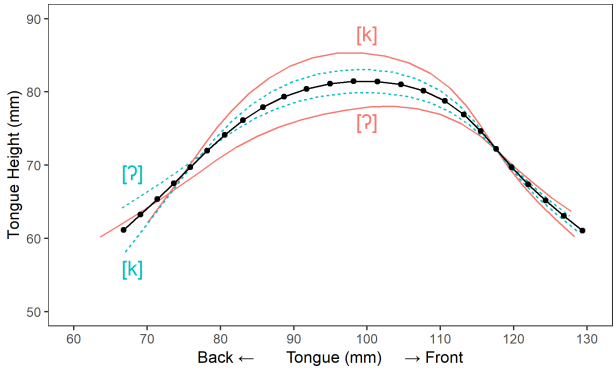


Table 2: *p*-values from dip tests for the articulatory and acoustic data, by talker and pooled together. Shaded cells indicate values below an α-level of 0.05.

Talker	PC1	PC2	F1	F2	V dur.
S1	0.017	0.885	0.898	0.773	0.034
S2	0.461	0.944	0.773	0.895	0.996
S3	0.020	0.459	0.470	0.534	0.850
S4	0.992	0.914	0.009	0.325	0.602
S5	0.03	0.711	0.781	0.976	0.828
S6	0.700	0.990	0.958	0.990	0.728
S7	0.134	0.876	0.992	0.509	0.932
S8	0.996	0.993	0.939	0.514	0.907
S11	0.607	0.903	0.924	0.870	0.963
pooled	0.925	0.903	0.993	0.982	0.985

Figure 2: Bar graphs for PC1 estimates of word item produced by talkers S5 (top) and S8 (bottom). White and grey bars indicate (orthographically-presented) /k/ and /ʔ/ items, respectively.

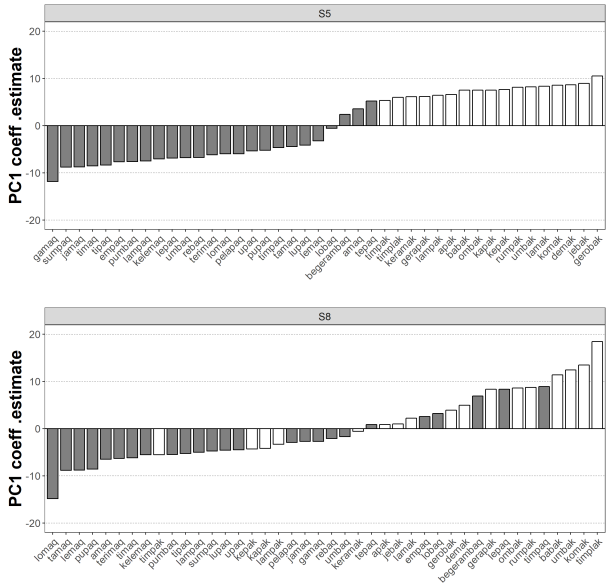
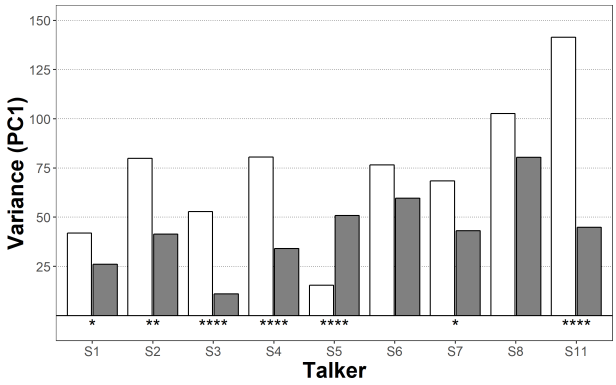


Figure 3: Comparisons of variance in PC1 coefficients for each talker. White bars indicate variance in /k/ tokens, and grey bars indicate variance among /ʔ/ tokens. Asterisks indicate the level of significance of F-test comparisons.



estimates for each word item's PC1 coefficient from talkers with the largest and smallest dip statistics (S5 and S8, respectively) are shown in Figure 2. Talker S5 makes a strong articulatory distinction between word-final /k/ and /ʔ/ across word items, whereas talker S8's articulations are far less categorical, with less separation between PC1 coefficients for /k/- and /ʔ/-words.

4. DISCUSSION & CONCLUSION

Results in this study show that words ending in /k/ and /ʔ/ indeed contrast in articulatorily (PC1=dorsal height and PC2=tongue root advancement) and acoustically (F1, F2, and vowel duration). However, an examination of the distribution of the measured variables reveals a general lack of categoricity (that is, bimodal distribution). Clear categorical differentiation between /k/- and /ʔ/ was demonstrated only in articulatory measures (PC1 coefficients) for three talkers and in acoustic measures (F1, vowel duration) for two talkers, based on the dip-test results. When measures were pooled across talkers, none exhibited any bimodal variation between these two sounds.

As shown in Figure 3, F-tests comparing variance in PC1 coefficients indicate that 6 of 9 talkers produced significantly greater variation in tongue height for /k/ than /ʔ/, with two additional talkers having the same trend but lacking a significant difference. One talker (S5) exhibited the opposite pattern, i.e. greater height variation in /ʔ/ than /k/. This outcome suggests that talkers did not achieve full (velar) constriction for many of their lingual articulations of word-final /k/, whereas for /ʔ/ no such lingual movement was ever needed. Higher rates of lingual-gesture reduction in /k/ may thus account for the overall pattern of phonetic gradience with this sound because productions of /k/ were more variable and less categorically distinguished than expected. Reductions of /k/ gestures may indicate a process of debuccalization affecting this sound in word-final contexts, thereby causing its merger with word-final /ʔ/. If Sasak is undergoing such a sound change, then the data shown here indicate that this process is both lexically-specific (word items exhibit different degrees of change or reduction) ands still incomplete (overall phonetic differences between the sounds are maintained, but some talkers produce greater categoricity than others).

The outcome of this study lends support to the idea that while phonological contrasts in a language may necessarily be categorical, such distinctions can be obscured by phonetically-relevant processes that render articulatory and acoustic patterns more gradient. These patterns may be induced by phonetic variation in speech (e.g. inter-talker variation, articulatory reduction, and contextual variation) and/or cause phonological change (e.g. systematic debuccalization of word-final /k/). Hence, the phonetic variation observed here involves an interaction between language- and talker-specific factors. Talkers produce a general contrast between /k/ and /ʔ/ but do not maintain clear-cut distinctions among their individual productions.

5. REFERENCES

[1] Boersma, P., Weenink, D. 2012. Praat: doing phonetics by computer. Version 5.3.14. http://www.praat.org.

[2] Chitoran, I., Cohn, A. 2009. Complexity in phonetics and phonology: Gradience, categoriality, and naturalness. In: Pellegrino, F., Marisco, E., Chitoran, I., Coupé, C., (eds), *Approaches to Phonological Complexity*. Berlin: Mouton de Gruyter 21–46.

[3] Cole, J., Lindebaugh, G., Munson, C. M., McMurray, B. 2010. Unmasking the acoustic effects of vowel-to-vowel coarticulation: a statistical modeling approach. *Journal of Phonetics* 38, 167–184.

[4] Diehl, R. L., Lotto, A. J., Holt, L. L. 2004. Speech perception. *Annual Review of Psychology* 55, 149–179.

[5] Fowler, C. A. 1994. Invariants, specifiers, cues: an investigation of locus equations as information for place of articulation. *Perception & Psychophysics* 55, 597–610.

[6] Hartigan, J., Hartigan, P. 1985. The dip test of unimodality. *Annals of Statistics* 13, 70–84.

[7] Johnson, K. 1997. Speech perception without speaker normalization: an exemplar model. In: Johnson, K., Mullennix, J. W., (eds), *Talker Variability in Speech Processing*. San Diego: Academic Press 145–165.

[8] Johnson, K. 2011. *Quantitative methods in linguistics*. Wiley-Blackwell.

[9] Ladefoged, P., Broadbent, D. E. 1957. Information conveyed by vowels. *Journal of the Acoustical Society of America* 29(1), 98–104.

[10] Li, M., Kambhamettu, C., Stone, M. 2005. Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics* 19(6–7), 545–554.

[11] Maechler, M. 2016. diptest. [R Package]. Version 0.75-7. http://cran.r-project.org. (last accessed 2-Dec-18).

[12] R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

[13] Recasens, D., Rodríguez, C. 2016. A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound. *Journal of Phonetics* 59, 58–75.

[14] Stevens, K. 1989. On the quantal nature of speech. *Journal of Phonetics* 17, 3–46.

[15] Stone, M. 2005. A guide to analyzing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics* 19, 455–501.

[16] Telemed, 2014. Echo Wave II [computer program]. Version 3.3.2. http://http://www.telemedultrasound.com.

[17] Turton, D. 2017. Categorical or gradient? an ultrasound investigation of /l/-darkening and vocalisation in varieties of english. *Laboratory Phonology* 8(1), 13.