The bursts of stops can convey dialectal information

Charalambos Themistocleous[a)]
Centre for Linguistic Theory and Studies in Probability, University of Gothenburg,
Dicksonsgatan 4, 412 56 Gothenburg, Sweden
themistocleous@gmail.com

Abstract: This study investigates the effects of the dialect of the speaker on the spectral properties of stop bursts. Forty-five female speakers—20 Standard Modern Greek and 25 Cypriot Greek speakers—participated in this study. The spectral properties of stop bursts were calculated from the burst spectra and analyzed using spectral moments. The findings show that besides linguistic information, i.e., the place of articulation and the stress, the speech signals of bursts can encode social information, i.e., the dialects. A classification model using decision trees showed that skewness and standard deviation have a major contribution for the classification of bursts across dialects.

[AL]
Date Received: July 19, 2016 **Date Accepted:** September 27, 2016

1. Introduction

Stop consonants are produced with a complete closure followed by a release in the vocal tract, which allows the oral pressure to build up behind the occlusion point. The direct consequence of the increase of pressure is the abrupt release of the articulators that lets the air to exit from the oral cavity with a burst of noise (Löfqvist, 1992). Earlier studies demonstrated that the bursts, which are remarkably small segments of speech, can convey information about stops' place of articulation (POA) (Halle *et al.*, 1957; Löfqvist, 1992; Harrington and Cassidy, 1999). They can also convey information about the gender and age of the speakers (e.g., see Nissen and Fox, 2009).

The spectral moments can represent both the global and local properties of burst noise (e.g., see Forrest *et al.*, 1988; Kardach *et al.*, 2002). In an earlier study, Forrest *et al.* (1988) classified voiceless sibilants, using spectral moments calculated from Bark transformed spectra, with great accuracy (98% correctness). This study analyzes the bursts of voiceless stops produced by speakers of two dialects of Modern Greek, namely, Standard Modern Greek (SMG) and Cypriot Greek (CG). The two dialects share the same vowels but differ in their consonants. SMG contains voiceless [p t k c] and voiced stops [b d g ɟ] (Jong Kong *et al.*, 2012); by contrast, in CG voiced stops are always prenasalized. Also, in CG there are post-alveolar fricatives and affricates; these sounds do not exist in SMG. Notably, the CG phonemic inventory contains pairs of singleton and geminate consonants whereas the SMG phonemic inventory does not. CG voiceless geminate stops are longer and more aspirated than the corresponding singletons. The aim of this study is to investigate whether information about the dialect can be conveyed by the spectral properties of the burst noise of voiceless stop consonants. To this purpose, the study employs a spectral moments analysis of stop bursts along with temporal information of bursts. This study can contribute to the understanding of typical speech productions and how they can convey linguistic and sociolinguistic information with potential applications in the development of systems for dialect identification.

2. Methodology

2.1 Speakers

Twenty SMG and 25 CG female speakers between 19 and 29 yrs old born and raised in Athens and Nicosia, respectively, participated in the study. The speakers were selected using a demographic questionnaire based on sociolinguistic criteria, e.g., same age, gender, education—all were university students—and social background. The speakers were bilingual in Greek and English and reported no speech or hearing disorders.

[a)]Author to whom correspondence should be addressed.

2.2 Speech materials

The bursts of voiceless stop consonants with four places of articulation, i.e., bilabial [p], alveolar [t], palatal [c], and velar [k], were elicited from a series of CVCV words. The words contained the target consonant in word initial /CV̀*sa* (e.g., /‘pasa, ‘kasa, ‘tasa, etc./) and word medial position /sàCV/ (e.g., /‘sapa, ‘saka, ‘sata, etc./), in both stressed and unstressed syllables (e.g., /‘pasa vs pa‘sa, ‘sapa vs sa‘pa, etc./). In Greek, [c] and [k] are allophones of the phoneme /k/ ([c] appears before front vowels and [k] before back vowels), so to elicit these sounds, two vowel environments were included in the experimental design, i.e., /a/ and /i/. The keywords were uttered in a carrier phrase, which was slightly modified to fit the dialects of the speakers:

(1) SMG: “/ ‘ipes <keyword> ‘pali /” (You said <keyword> again).
(2) CG: “/ ‘ipes <keyword> ‘pale /” (You said <keyword> again).

2.3 Procedure

The recordings for the CG materials were conducted in a quiet room at the University of Cyprus in Nicosia and for the SMG materials in a recording studio in Athens. A male CG speaker and a female SMG speaker provided standard instructions before the recording to the SMG and CG speakers, respectively, e.g., to speak in their standard pace, seat appropriately in front of the microphone, and keep a designated distance. The target words were presented in standard Greek orthography. (Stress marks are conventionally represented in Greek orthography). The stimuli were randomized for every repetition and speaker. Between the repetitions there was a 2-mine break. No instructions about the prosodic pattern or any explanation about the purposes of the experiment were provided. The speakers read sentences out loud from a computer screen, at a comfortable, self-selected rate. Recordings were made on a Zoom H4n audio recorder where voice was sampled at 44.1 kHz. Segmentation of the onset and offset of the stop burst was conducted through simultaneous inspections of the waveform and the spectrogram using Praat (Boersma and Weenink, 2016). To demarcate the onset of the stop burst, the sharp increase in the diffuse noise energy and the rapid increase in zero crossings have been employed. The sharp decrease in the diffuse noise energy was used to designate the burst offset (Nissen and Fox, 2009). The segmentation decisions were evaluated twice by the author and another trained phonetician. To calculate the acoustic parameters the Discrete Fourier Transformations (DFTs) were averaged using the time-averaging procedure of Shadle (2012). Within time-averaging, a number of DFTs were taken from across the duration of the burst and averaged for each token. Then the first four spectral moments: centroid or centre of gravity (COG), standard deviation (SD), skewness, and kurtosis were calculated from the burst spectra. The stimulus material consisted of 2160 stimuli. Specifically, a total sum of 1440 productions was produced for [p] and [t] (i.e., 45 Speakers × 2 POA × 2 Stress Positions × 2 vowel environments × 4 Repetitions) and 720 productions for the [c] and [k] that appear only before /i/ and /a/, respectively, 2 × (i.e., 45 Speakers × 1 POA × 2 Stress Positions × 1 Vowel Environment × 4 Repetitions). To minimize speaker’s attention on the keywords, filler words were added in the same carrier sentences.

2.4 Statistics

Linear mixed effects models of the effects of the POA, Dialect, and Stress as fixed effects on duration, COG, SD, skewness, and kurtosis as response variables (RVs) were conducted using R (R Core Team, 2016) and lme4 (Bates *et al.*, 2015). As random effects, the models included intercepts for Speakers and Keywords. To analyze the data, several models were compared starting with the three fixed factors and their interactions. The Speaker and the Keyword were employed as random effects. The final model was selected using model comparison. Changes in the Akaike information criterion and the Bayesian information criterion or Schwarz criterion for each model have been employed as Diagnostics for detecting overparameterization and model simplification (Bates *et al.*, 2015). The RVs were transformed in a logarithmic scale following the requirements of the linear mixed models,

$$\mathrm{RV} \sim \mathrm{POA} * \mathrm{Dialect} * \mathrm{Stress} * \mathrm{Consonant\ Position} + (1|\mathrm{Speaker}) + (1|\mathrm{Keyword}). \quad (1)$$

2.5 Classification and decision trees

Moreover, to estimate the contributions of spectral moments and duration for the classification of Dialect, three classification models were fitted: a linear discriminant analysis (lda) model, a flexible discriminant analysis (fda) model, and a C5.0 machine learning and classification model. The spectral moments and the duration were used as predictors and Dialect as the RVs. Lda, fda, and C5.0 were evaluated with respect to Receiver Operating Characteristic Curves and the Sensitivity of the training model, the resulting true-positive rate (i.e., the sensitivity), and the specificity of the model. As C5.0 outperformed both the lda and the fda, we report these classification results only. Moreover, to estimate model predictions, the decision tree was trained on 90% of the data and used to evaluate the 10%, treated as unknown data. A classification model using a repeated ten-fold cross-validation with three repetitions is also provided. The statistical analysis and the classification were carried out in R (R Core Team, 2016). The following R packages have been used for the analysis: the lme4, which provided functions for fitting generalized linear mixed models (Bates *et al.*, 2014; Kuznetsova *et al.*, 2016), the caret (Kuhn, 2016), and the C5.0 package (Kuhn *et al.*, 2015).

3. Results

The spectral properties of burst noise were investigated from frequency spectra; examples of stop bursts' waveforms for [p t c k] are shown in Fig. 1. In the following, we report the results from the statistical analysis.

3.1 Temporal properties

POA, Dialect, Stress, and Consonant Position had significant effects on the log transformed Duration, and resulted in significantly different slopes from the intercept [$\beta = 2.33$, standard error (SE) $= 0.06$, $t(71.60) = 36.03$, $p < 0.001$]. Specifically, POA resulted in significant effects for stops articulated at the Alveolar POA [$\beta = 0.38$, SE $= 0.08$, $t(44.30) = 5.01$, $p < 0.001$], the Palatal [$\beta = 1.15$, SE $= 0.09$, $t(44.40) = 12.32$, $p < 0.001$], and the Velar [$\beta = 0.80$, SE $= 0.09$, $t(44.40) = 8.64$, $p < 0.001$]. Also, the interaction of POA and Dialect had significant effects on Duration, which were manifested as different models for the CG Palatals [$\beta = -0.25$, SE $= 0.10$, $t(1737) = -2.61, p < 0.01$] and CG Velars [$\beta = -0.39$, SE $= 0.10$, $t(1737) = -4.02$, $p < 0.001$]. The interaction of Stress × Dialect had a significant effect, resulting in a significantly different slope for the CG Stressed bursts [$\beta = 0.17$, SE $= 0.08$, $t(1738) = 2.14$, $p < 0.05$]. The position of the burst in the utterance affects the temporal properties of burst and that was evident in the significant interaction between POA × Dialect × Stress × Consonant Position, which resulted in significantly different slopes for the CG Stressed Palatals in Word Medial position [$\beta = 0.60$, SE $= 0.19$,

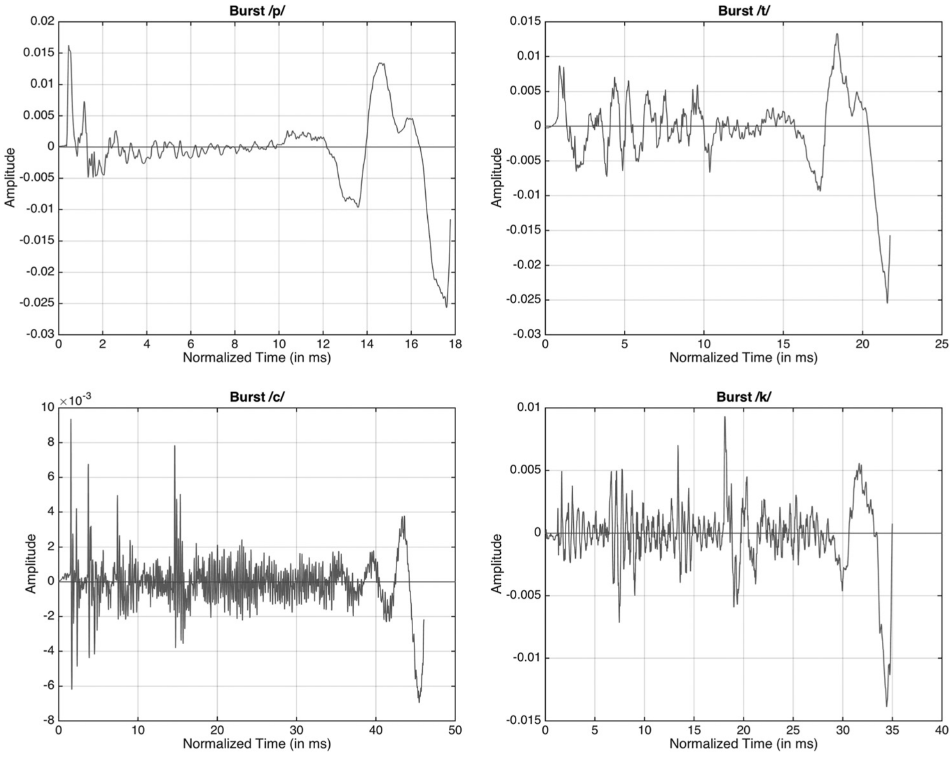

Fig. 1. Waveform of the burst of SMG [p t c k].

$t(1737) = 3.10,\ p < 0.001$] and CG Stressed Velars in Word Medial position [$\beta = 0.39,\ \mathrm{SE} = 0.20,\ t(1738) = 1.99,\ p < 0.05$].

3.2 COG

There is an overall higher COG for the Palatal stops whereas the Bilabial and Velar stops associate with the lowest COG. The Dialect did not have significant effects on the log transformed COG. POA had a significant effect, which was manifested by the effects of the Alveolar [$\beta = 0.59,\ \mathrm{SE} = 0.11,\ t(31.8) = 5.37,\ p < 0.001$], the Palatal [$\beta = 1.29,\ \mathrm{SE} = 0.13,\ t(31.8) = 9.58,\ p < 0.001$], and the Velar Bursts [$\beta = 0.29,\ \mathrm{SE} = 0.13,\ t(31.8) = 2.16,\ p < 0.05$] on the intercept [$\beta = 14.22,\ \mathrm{SE} = 0.09,\ t(43.9) = 166.48,\ p < 0.001$].

3.3 SD

The SD was smaller for [p] and [k] and greater for [t] and [c]. POA and the Dialect had a significant effect on the log transformed SD. Specifically, POA resulted in significantly different slopes from the intercept [$\beta = 7.50,\ \mathrm{SE} = 0.057278,\ t(65) = 131.02,\ p < 0.001$] for bursts articulated at the Alveolar [$\beta = 0.395,\ \mathrm{SE} = 0.069537,\ t(44) = 5.69,\ p < 0.001$] and Palatal POA [$\beta = 0.447212,\ \mathrm{SE} = 0.085,\ t(43) = 5.25,\ p < 0.001$]. The Dialect had significant effects as well, manifested by a significantly different slope for CG bursts [$\beta = 0.31,\ \mathrm{SE} = 0.064,\ t(147) = 4.78,\ p < 0.001$].

3.4 Skewness

[p] and [k] had higher skewness than [t] and [c]. POA, Dialect, and Stress resulted in significantly different slopes from the intercept of the log transformed Skewness [$\beta = 1.47,\ \mathrm{SE} = 0.12,\ t(51) = 12.74,\ p < 0.001$]. Specifically, POA resulted in significant effects for bursts articulated at the Alveolar POA [$\beta = -0.78,\ \mathrm{SE} = 0.14,\ t(35) = -5.43,\ p < 0.001$] and the Palatal POA [$\beta = -1.69,\ \mathrm{SE} = 0.18,\ t(35) = -9.59,\ p < 0.01$]. Moreover, the interaction of POA × Dialect and POA × Dialect × Stress were significant, as was evident by the significantly different slopes of CG Palatals [$\beta = 0.45,\ \mathrm{SE} = 0.16,\ t(1628) = 2.866,\ p < 0.01$] and the CG Stressed Palatals [$\beta = -0.58,\ \mathrm{SE} = 0.23,\ t(1627) = -2.59,\ p < 0.01$].

3.5 Kurtosis

Overall, kurtosis was high for [p] and [k] and low for [t] and [c]. This resulted in statistically significant effects for [t] and [c]. POA and the interaction of POA × Dialect resulted in significantly different slopes from the intercept of the log transformed kurtosis [$\beta = 3.28,\ \mathrm{SE} = 0.18,\ t(50) = 18.47,\ p < 0.001$]. Specifically, POA resulted in significant effects for burst articulated at the Alveolar [$\beta = -1.55,\ \mathrm{SE} = 0.19,\ t(34) = -8.19,\ p < 0.001$] and the Palatal places of articulation [$\beta = -2.64,\ \mathrm{SE} = 0.23,\ t(36) = -11.65,\ p < 0.001$]. The interaction of POA × Dialect was manifested by the significantly different productions of CG Velars [$\beta = -0.396,\ \mathrm{SE} = 0.197,\ t(1462) = -2.01,\ p < 0.05$].

3.6 Classification model

Sections 3.1–3.5 showed that information about the POA, the Stress, and the Dialect is represented by effects on the spectral moments and the duration of the bursts. The classification model showed that all the spectral moments and the duration contribute to the classification of the Dialect, with skewness and SD as the most important cues. Specifically the contribution of the spectral moments and duration was the following: 100% skewness, 96% SD, 35% COG, 33% kurtosis, and 16% Duration. The resulting tree is shown in Fig. 2. The classification model was evaluated over a test data set (classification accuracy = 82.45%, kappa = 0.64; the kappa is a metric of the overall accuracy to the accuracy expected by chance). A classification model using a repeated ten-fold cross-validation with 3 repetitions, yielded similar results: classification accuracy = 80%, kappa = 0.58). At each node of the tree (see Fig. 2), C5.0 selects the attribute that most effectively splits the data into subsets. The splitting criterion is the difference in entropy (i.e., the normalized information gain). The attribute that carries the highest normalized information gain is chosen to make the decision. Each decision is a binary yes or no classification. The C5.0 algorithm then recurs on the smaller subsets (see Quinlan, 1993).

4. Discussion

Speech signals convey linguistic and sociolinguistic information, namely, they provide information about the segments, e.g., the consonants and vowels—the prosody, e.g., the melodies of the utterances—and the social identity of the speakers, e.g., the dialect,

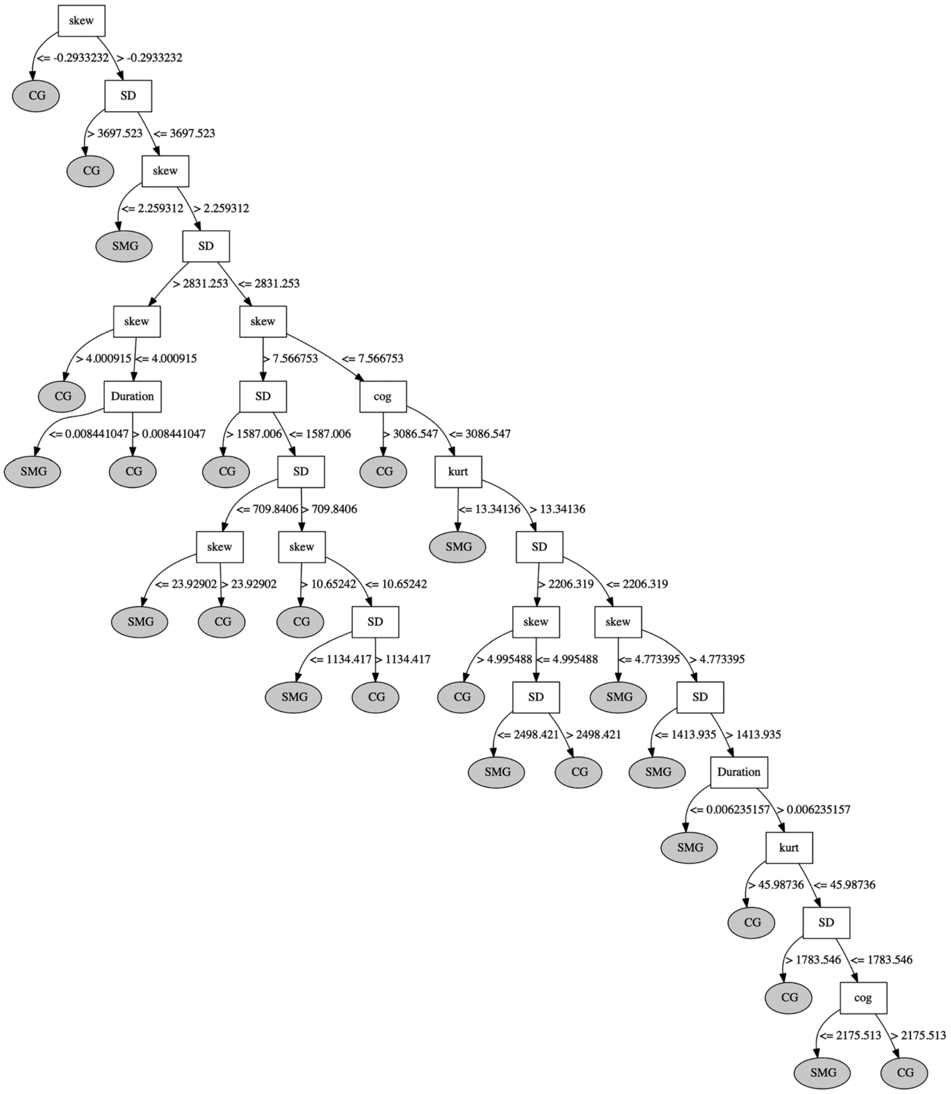

Fig. 2. Decision tree produced by the classification algorithm C5.0.

the age, and the gender of the speakers. This study showed that this information is encoded in a complex manner affecting not only the periodic but also the noisy parts of the utterances, such as the bursts of stop consonants [see also Themistocleous *et al.* (2016) for a similar study of Greek fricatives] and that differences in the spectral properties of the burst noise can be explained by the effects of the POA and stress. More specifically, the POA had significant effects on both the spectral moments and the duration. The interaction of stress and POA had a significant effect on the duration; also, the interaction of stress, POA, and dialect had a significant effect on skewness. Also, the spectral properties of the burst noise encode information about the dialect of the speakers. The decision tree produced by the machine learning and classification algorithm C5.0 showed that the skewness and the SD have the greatest contribution for the classification of the dialect of the speakers. This finding emphasizes that even the smallest segments of speech, such as the stop bursts, can convey both linguistic and more importantly dialectal information. Moreover, it suggests that the bursts can be employed as an acoustic cue for the identification of the dialect of speakers.

References and links

Bates, D., Mäechler, M., Bolker, B., and Walker, S. (**2014**). "lme4: Linear mixed-effects models using Eigen and S4," R package version 1.1-6.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Software **67**(1), 1–48.

Boersma, P., and Weenink, D. (**2016**). "Praat: Doing phonetics by computer," Version 6.0.19 [computer program].

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. **84**(1), 115–123.

Halle, M., Hughes, W., and Radley, A. (**1957**). "Acoustic properties of stop consonants," J. Acoust. Soc. Am. **29**, 107–116.

Harrington, J., and Cassidy, S. (**1999**). *Techniques in Speech Acoustics* (Kluwer Academic Publishers, Dordrecht, Boston).

Jong Kong, E., Syrika, A., and Edwards, J. R. (**2012**). "Voiced stop prenasalization in two dialects of Greek," J. Acoust. Soc. Am. **132**(5), 3439–3452.

Kardach, J., Wincowski, R., Metz, D. E., Schiavetti, N., Whitehead, R. L., and Hillenbrand, J. (**2002**). "Preservation of place and manner cues during simultaneous communication: A spectral moments perspective," J. Commun. Disorders **35**(6), 533–542.

Kuhn, M. (**2016**). "caret: Classification and Regression Training," R package version 6.0-68.

Kuhn, M., Weston, S., Coulter, N., and code for C5.0 by R. Quinlan, M. C. C. (**2015**). "C50: C5.0 Decision Trees and Rule-Based Models," R package version 0.1.0-24.

Kuznetsova, A., Bruun Brockhoff, P., and Christensen, H. B. R. (**2016**). "lmerTest: Tests in Linear Mixed Effects Models," R package version 2.0-30.

Löfqvist, A. (**1992**). "Acoustic and aerodynamic effects of interarticulator timing in voiceless consonants," Lang. Speech **35**(1–2), 15–28.

Nissen, S. L., and Fox, R. A. (**2009**). "Acoustic and spectral patterns in young children's stop consonant productions," J. Acoust. Soc. Am. **126**(3), 1369–1378.

Quinlan, R. (**1993**). *C4.5: Programs for Machine Learning* (Morgan Kaufmann Publishers, San Francisco, CA).

R Core Team (**2016**). "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria.

Shadle, C. (**2012**). "Acoustics and aerodynamics of fricatives," in *The Oxford Handbook of Laboratory Phonology*, edited by A. C. Cohn, C. Fougeron, and M. K. Huffman (Oxford University Press, New York), pp. 511–526.

Themistocleous, C., Savva, A., and Aristodemou, A. (**2016**). "Effects of stress on fricatives: Evidence from Standard Modern Greek," in *Interspeech 2016*, San Francisco, CA (September 8–12, 2016), pp. 1026–1029.