Journal of Phonetics 71 (2018) 395–409

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Special Issue: Marking 50 Years of Research on Voice Onset Time, eds. Cho, Docherty & Whalen

A multimodal approach to the voicing contrast in Turkish: Evidence from simultaneous measures of acoustics, intraoral pressure and tongue palatal contacts

Özlem Ünal-Logacev [a],*, Susanne Fuchs [b], Leonardo Lancia [c]

[a] Istanbul Medipol University, Istanbul, Turkey
[b] Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Berlin, Germany
[c] Laboratoire de Phonétique et Phonologie (CNRS – Sorbonne nouvelle), Paris, France

ARTICLE INFO

Article history:
Received 3 December 2017
Received in revised form 6 October 2018
Accepted 7 October 2018
Available online 25 October 2018

Keywords:
Voicing contrast in Turkish
Stops
Tongue-palatal contacts
Intraoral pressure
GAMM

ABSTRACT

The aims of the study are to investigate acoustic, aerodynamic and supralaryngeal properties of the voicing contrast in Turkish and to better understand the relation between these factors in the maintenance and inhibition of phonetic voicing. For this purpose, simultaneous recordings were carried out using electropalatography, a piezoresistive pressure transducer and a microphone for six speakers of Turkish. The voiced /d, dʒ/ and voiceless /t, tʃ/ target sounds occurred in word-initial position in intervocalic context. Single time points were selected to study the voicing contrast and its corresponding properties. The most pronounced differences between voiced and voiceless consonants were the relative voicing during closure and the velocity maximum of intraoral pressure (Pio). Phonologically voiced stops showed a relatively long voicing portion, a negative VOT (for /d/) and a slower rise in Pio. Voiceless stops were realized with less voicing, positive VOT (for /t/) and a steep intraoral pressure rise. However, differences were not found for tongue-palatal contact patterns at full closure. The analysis of mutual dependence between articulatory and aerodynamic measures through Generalized Additive Mixed Model (GAMM) showed a linear relation between the two measures in voiced stops and a nonlinear relation for the voiceless. These results are discussed in light of laryngeal-oral coordination and cavity enlargement. Moreover, the different methodological approaches and their benefits are considered.

1. Introduction

In this work, we discuss various aspects of voicing contrast in Turkish stops. In particular, the relation between articulation and aerodynamics is investigated so as to discuss their interplay in the maintenance or disappearance of voicing during oral closure productions. This interplay is examined in light of motor equivalence, a basic principle in motor control describing the capacity of the motor system to achieve the same goal with different underlying mechanisms (Perrier & Fuchs, 2015). Maintaining voicing during oral closure as is the case in phonologically voiced stops requires a transglottal pressure drop between subglottal and intraoral pressure (e.g., Westbury, 1983). To guarantee such a pressure drop, the oral cavity must be enlarged to keep intraoral pressure low. Several cavity enlargement manoeuvres have been reported in the literature. If no cavity enlargement manoeuvres are realized or laryngeal-oral timing is changed, intraoral pressure rises quickly, i.e. with a steep slope, and voicing dies out. We carry out a multimodal analysis for Turkish, an under-investigated language for which preliminary evidence reveals a phonetic voiced-voiceless distinction (Öğüt, Kılıç, Engin, & Midilli, 2006).

The aims of the study are twofold: First, we aim to better understand the direct relation between intraoral pressure rise and supralaryngeal articulation to maintain or inhibit phonetic voicing during closure. Second, we wish to investigate acoustic, aerodynamic and supralaryngeal properties of the voicing contrast in Turkish. To do so, we use single time point analysis (with time points often suggested in the literature) and contrast this with an analysis of the mutual dependence of articulatory and aerodynamic measures through Generalized Additive Mixed Models (GAMMs).

---

* Corresponding author at: Istanbul Medipol University, School of Health Sciences, Department of Speech and Language Therapy, Kavacık, Istanbul, Turkey.
E-mail address: ologacev@medipol.edu.tr (Ö. Ünal-Logacev).

https://doi.org/10.1016/j.wocn.2018.10.002
0095-4470/

The originality of our approach lies in the combination of acoustic, articulatory and aerodynamic measures without sacrificing the comfort of the subject. Combining electropalatography with a piezoresistive pressure sensor is a powerful technique which allows for an investigation of the underlying mechanisms in the production of voicing. A similar approach has only been used for the study of voiceless obstruents (Fuchs & Koenig, 2009) while investigations of the phonological voicing contrast have mostly either focused on aerodynamic or articulatory measures or have drawn inferences on the basis of aerodynamic signals. Some major investigations and their results will be described in the following sections.

1.1. Acoustic properties of the voicing contrast in Turkish

Turkish as a language is particularly interesting, because it belongs to the group of languages which are under-investigated. The most comprehensive study in terms of sample size has been carried out by Öğüt et al. (2006). They investigated Voice Onset Time (VOT) in the production of word-initial monosyllabic stops /b, d, g/ versus /p, t, k/ in 30 speakers (15 females) of Standard Turkish. All words were meaningful. The word-initial stops were followed by eight different vowels and repeated three times. VOT was measured following the pioneering work of Lisker and Abramson (1964) with negative values corresponding to voicing lead and positive values corresponding to long or short voicing lag. An analysis of variance revealed significant differences between /b, d, g/ and /p, t, k/, an effect of place of articulation (velars are longer than dental and bilabial stops), but no effect of vowel context and sex. All phonologically voiceless stops were produced with a positive VOT. Results for phonologically voiced stops showed negative VOT values, with the exception of /g/. In /g/, positive VOT values were found in 40% of the cases and negative VOT values in 60%. The authors conclude that Turkish stops can be classified in the sense that phonologically voiced stops have voicing lead and phonologically voiceless stops have a long voicing lag.

The empirical evidence may change slightly when word-initial stops are preceded by an utterance. Feizollahi (2010) carried out an experiment recording four Turkish speakers reading words in sentences with word-initial plosives which were preceded by words with a final voiced consonant, a voiceless stop or a vowel. He hypothesized that if the word-final phoneme would be phonologically voiceless, voicelessness would also be found in the realization of the word-initial stop, no matter whether it is phonologically voiced or voiceless. Comparably, if the word–final phoneme would be phonologically voiced, voicing would be spread to the following word-initial position, no matter whether it is phonologically voiced or voiceless. Feizollah's findings only partially support these hypotheses. When the final phoneme was phonologically voiceless, three out of four speakers realized word-initial stops without voicing, even when the following word started with a phonologically voiced stop. However, this was not the case when the preceding final consonant was phonologically voiced. In this case voicing did not spread to the word-initial position with a phonological voiceless stop. Thus, there are contextual effects on the production of phonologically voiced stops in word-initial position. These sounds devoice when preceded by a phonologically voiceless stop. Phonologically voiceless stops, however, are relatively resistant and keep their phonetic voicelessness, even when preceded by a voiced segment.

The two studies reveal that VOT and voicing during closure are two acoustic parameters that can differentiate phonologically voiced from voiceless stops in Turkish.

1.2. Empirical evidence for the voicing contrast based on aerodynamics

Throughout the last century, a number of studies have been carried out which report larger intraoral pressure peaks in phonologically voiceless obstruents than in voiced ones. Most of these studies were carried out for American English speakers. For instance, Arkebauer, Hixon, and Hardy's (1967) findings revealed higher intraoral pressure peaks for voiceless stops and fricatives in comparison to voiced ones for children and adults, independent of position in the syllable, speech rate and intensity differences. Malécot (1970) attributed a particular role to the pressure differences. He claimed that intraoral pressure variation would lead to a speaker's synesthetic impression of either fortis (voiceless phonemes with higher intraoral pressure) or lenis sounds (voiced phonemes with lower intraoral pressure; see more recent experiments on the perception of aero-tactile feedback by Gick and Derrick, 2009).

Stathopoulos (1986) compared initial and final /p/ and /b/ in 20 adults and 20 children in comfortable, soft and loud speech of American English. She showed that intraoral pressure peaks were higher for /p/ than /b/, but syllable position only had an impact on /b/ not /p/. Higher values tended to occur more in initial than in final position, minimizing the intraoral pressure difference between /p/ and /b/ in syllable initial position and maximizing it syllable-finally. Subtelny, Worth and Sakuda (1966) analysed 10 males, 10 females and 10 children. Their findings revealed the expected differences in intraoral pressure peaks. Their results differed, however, with respect to age and sex. Males generally showed lower pressure peaks than females and females had lower pressure peaks than children. Warren and Wood (1969) investigated the production of phonologically voiced and voiceless obstruents in 20 speakers. They reported larger air flow peaks for voiceless obstruents and explained the corresponding larger intraoral peaks with respect to a larger air volume.

However, there are some investigations which found limited differences in intraoral pressure between voiced and voiceless sounds. For instance, Lisker (1970) questioned the speech material in other studies (consisting of very short, often monosyllabic words) and recorded an American English speaker producing /p, t, k/ and /b, d, g/ in various contexts (word-initial, medial and final, in unstressed and stressed positions). He writes, "Unless our sample is completely unrepresentative of American English stops, it must be significant that no more than about 15% of the stops measured have pressures so low that they can be classed with /b, d, g/ with certainty and that only a bare of 2% have pressure so great that they are unambiguously /p, t, k/. Thus, the overwhelming majority of the stops in our sample cannot be identified with confidence solely on the basis of peak pressures." (Lisker, 1970, p. 220). One limitation of Lisker's study, however, is that he only recorded one speaker. Flege (1983), who looked at six female American English speakers, found that the distinction between /p/ and /b/ in absolute utterance initial positions disappeared. Similarly, Fischer-Jørgenson and Hansen (1959) found only weak differences between /b/ and /p/ in peak intraoral pressure in Danish word internal stops.

Zygis, Fuchs, and Koenig (2012) reported language specific differences in terms of intraoral pressure in stops, affricates and fricatives during word-initial and medial productions of German and Polish speakers. Their results provide evidence for consistent differences in realizations of Polish speakers, with higher pressure peaks for phonologically voiceless obstruents in all positions. For German speakers, the pressure peak barely differed between phonologically voiced and voiceless items. The only significant effect was found for word-medial stops: Values for /t/ were higher than those for /d/.

In the light of these results it is hard to argue that phonologically voiced and voiceless obstruents can be distinguished on the basis of the intraoral pressure alone. However, if significant differences occur, there is a very high likelihood that phonologically voiceless obstruents have a larger pressure peak than voiced ones. Moreover, there may be cross-linguistic differences and one can expect these to occur in an utterance-, word-, or syllable- initial position.

In their seminal work on intraoral pressure profiles, Müller and Brown (1980) took the analysis a step further. They did not only look at one particular time point, the intraoral pressure peak, but tried to provide a metric that permits characterizing pressure profiles. First, they graphically inspected the data of five speakers and noted that especially the closure part of the pressure profiles could be separated into concave, convex, linear, bimodal and delayed shapes (see Fig. 12 in Müller and Brown, 1980, p. 337). In 70% of all cases, voiceless stops had a convex shape while voiced stops had a concave shape. These two shapes were then further quantitatively assessed by the difference of two slopes. For the convex shape, the initial slope from the baseline rose quickly to a plateau. This steep initial slope was subtracted from the second slope determining the slowly rising pressure during the plateau up to the pressure maximum. The concave shape in phonologically voiced stops was determined similarly. However, the initial slope from the baseline to a turning point rose slowly, while the second slope corresponded to a quicker rise up to the pressure maximum. Note that in voiced stops, no pressure plateau was present. These shapes were discussed with respect to the underlying articulatory mechanisms. In particular, the quickly rising initial pressure slope in voiceless stops was analysed as a result of glottal aperture and increased pulmonary airflow leading to a fast increase in intraoral pressure while the slowly rising initial slope in voiced stops was associated with cavity enlargement manoeuvers, preventing a fast decrease of the transglottal pressure difference. Since then, these measures have been used by Koenig and Lucero (2008) for children (5 and 10 years old) and women (for each group, eight speakers were recorded) producing /p/ and /b/ in word-initial and medial positions. Differences in intraoral pressure shapes with respect to the voicing contrast (convex and concave) have been found for all women, some ten-year old children, but only a few 5 year olds. Koenig and Lucero (2008) suggest limited aerodynamic control in the production of voicing, at least for the 5 years old children. Some other authors have only partially adopted Müller and Brown's (1980) measures by looking in particular at the initial slope (slowly rising for voiced and quickly rising for voiceless). For example, Zygis, Fuchs, and Koenig (2012) were able to distinguish phonologically voiced and voiceless in initial and medial obstruents for German and Polish on the basis of this measure. Polish turned out to be a particularly special case, because the pressure rose only very slowly in the voiced phonemes so that the authors supposed that specific cavity enlargement strategies were at work. However, no articulatory data were reported. The underlying articulatory mechanisms could be manifold.

1.3. Empirical evidence for cavity enlargement based on articulatory studies

Cavity enlargement refers to some mechanisms in which the size of the oral cavity is increased during the oral closure of a stop. This enlargement is carried out to prevent intraoral pressure from rising quickly with the closure of the vocal tract (for a modelling approach see Westbury, 1983). Moreover, an enlarged oral cavity allows a transglottal pressure difference between subglottal and intraoral pressure to be sustained, a necessary requirement for phonation. Our focus here lies primarily on supralaryngeal articulation, without questioning that glottal closure or aperture have an impact on intraoral pressure changes and the evolution of the transglottal pressure differences as well. Early work by Kent and Moll (1969) using lateral cinefluorography supported the hypothesis that supralaryngeal articulation is involved in the voicing contrast, even if in phonology the contrast is often exclusively defined at the level of the larynx. Three speakers of American English were recorded with stop series in different contexts. Results of this experiment consistently showed a larger oral cavity for voiced than for voiceless stops. In particular, the hyoid bone was depressed (lowered) with greater distance between the back of the tongue and the posterior pharyngeal wall. Westbury (1983) also used high-speed cinefluorographic films to analyse /b, d, g/ versus /p, t, k/ productions for one speaker of American English. He found different strategies involved in the larger oral cavity in voiced stops. "If it is more important during voiced stops to control whether (rather than how) the vocal folds oscillate, then all cavity enlargement manoeuvres whose magnitude and duration satisfy the boundary conditions necessary for oscillations can be equally well-suited for that behavioural goal" (Westbury, 1983, p. 1333). What Westbury describes may be subsumed under the term *motor equivalence* (Perrier & Fuchs, 2015). Motor equivalence can be defined as the capability to achieve the same result through different approaches to a given task. In Westbury's study, the maintenance of voicing by means of an enlarged oral cavity was realized by a lowered larynx (for medial /b/ and /d/) and an advanced tongue root (for /d/ and /g/). The author notes "it would be of great interest to know whether and to what extent voicing related behaviour might vary for the same stop, repeated many times by the same speaker, in the same phonetic environment. . . .. such data might provide insights into optimization criteria. . ." (Westbury, 1983, p. 1334).

A number of studies investigating different mechanisms for cavity enlargement will be described here. The evidence for laryngeal lowering as a potential strategy is not uniform. By means of a thyroumbrometer, Ewan and Krones (1974) recorded the vertical laryngeal movements of intervocalic voiced and voiceless stops in six English speakers, one French speaker, one Thai speaker and one Hindi speaker. According to their results, voiceless stops have a higher larynx position than their corresponding voiced stops, in particular at the end of oral closure. For three Danish speakers, Petersen (1983) found a lower larynx position for voiced stops. However, since nasals showed the lowest laryngeal position, he assumed that the lower larynx position could hardly be responsible to preserve a sufficient pressure drop to guarantee voicing for the nasal consonants. Riordan (1980) recorded two speakers and though he found some small difference in laryngeal height, he suggested that this effect is so small that it cannot account for cavity enlargement on its own.

Nasal leakage has been proposed as an additional strategy and has been found more frequently in French and Spanish than in English (Solé, 2011) with some between-speaker variation. Furthermore, Perkell (1969) as well as Bell-Berti and Hirose (1972) found a higher velum for voiced stops in comparison to voiceless stops. Additionally, Bell-Berti and Hirose (1971, 1972) and Bell-Berti (1975) investigated whether cavity enlargement would be passive, i.e. due to reduced vocal tract compliance, or active. Three American English speakers were recorded by means of EMG and no consistent results were found. Tongue displacement in /m, b, p/ in relation to intraoral pressure estimates was observed by Svirsky et al. (1997). Both measurements were used to assess the validity of a tongue compliance model. Based on their results, they concluded that the tongue should be actively stiffened for voiceless stops. However, relaxation of the tongue for voiced stops did not explain all the observed changes results. Hence, the authors proposed a combination of intentional relaxation of tongue muscles with an active displacement for the voiced stops.

Using an X-ray microbeam system, Fujimura and Miller (1979) recorded three American speakers producing /d/ and /t/ in syllable and word-final position. Their results were mostly consistent for the jaw and provided evidence that /d/ was produced with a lower jaw position and a lower velocity compared to /t/. For /t/, the jaw moved more vigorously. These results could explain the production of a salient burst in /t/ due to a high jaw position (Mooshammer et al., 2003).

Different tongue placements, as measured with electropalatography (EPG), have also been described in the literature, though with different results. Dagenais, Lorendo, and McCutcheon (1994) recorded 10 American English speakers with EPG and showed more alveolar midline contacts for phonologically voiced stops compared to voiceless ones, averaged over all speakers. He explained this difference with a relaxation of the tongue at the palate for the voiced stops and a stiffening of the tongue with less contacts for voiceless stops. The opposite was found by Moen and Simonsen (1997) and Moen, Simonsen, Huseby, and Grue (2001) for /d/ versus /t/ in Norwegian (1997, 2001) and English (1997). For both languages, they reported a tendency for a greater amount of contact for /t/ than for /d/ during oral closure, but no statistics were provided. Fletcher (1989), who recorded American English speaking children, found no significant differences between voiced and voiceless alveolar stops. Dixit (1990) studied voiced and voiceless dental stops and retroflexes in Hindi and found that voiceless stops generally showed a significantly greater overall contact compared to the voiced ones. It is possible that all these studies differ because they used different speech material in different languages. However, they may also differ, because only a single time point was chosen for which tongue-palatal contacts were measured, most often the maximum amount of contact during oral closure.

1.4. Combining aerodynamics and articulation

Combining aerodynamic and articulatory measures in a comfortable way for the participants of a study is quite a challenging endeavour. Therefore, most studies concentrate either on aerodynamic or articulatory data and derive inferences about the other aspect. There are, however, a few exceptions: e.g. Lubker and Parris (1970) who combined lip contact, labial EMG and intraoral pressure; Fuchs and Koenig (2009), who worked on voiceless obstruents only and Searl and Evitts (2013), who investigated conversational versus clear speech. To some extent, inferring articulatory and aerodynamic properties may be appropriate when describing a general behaviour and under the assumption that there is a linear relation between aerodynamics and articulation. A linear relation would for example exist in the following case: Let us say that the tongue touches the palate while two electrodes are active in the EPG palate, leading to a rise in intraoral pressure by a specific amount. Then, if two additional contacts are activated, the pressure should rise to twice the level it was before. However, if the relation between intraoral pressure and number of contacts is nonlinear, we need time series of pressure and tongue-palatal contact values, since picking out a single time point could be misleading when attempting to describe an overall relation.

Even if only aerodynamics or articulatory measures are considered, selecting the time point most conducive to understanding one particular measure may be difficult. As was discussed earlier with respect to the intraoral pressure peak, several studies provided evidence that this peak might be a good measure while others have shown that phonologically voiced and voiceless obstruents do not differ in this respect. Nevertheless, clearly, this does not allow us to derive that there are no differences in the aerodynamics. Taking all samples of larger time windows into account while comparing different segments may give us a better idea of where or where not to expect differences in which temporal frames.

Vatikiotis-Bateson, Barbosa and Best (2014) wrote about this issue: “The inevitable and even desirable presence of fluctuations has several important implications for research on spatiotemporal behaviour. Importantly, it means that we cannot simply disregard measured variability as irrelevant noise, as has been done so often in psychological and linguistic research, because variance conflates notions of noise and error with mandatory, healthy fluctuations in patterned behaviour. Implicitly, then, the behaviour of the system must be examined dynamically as it unfolds through time – certainly, snapshot, magic moment measures will not suffice” (p. 168). We generally agree with this notion, although we adopt a less radical stance based on the idea that a careful inspection of the data, informed by the knowledge of the processes at work, may be sufficient for certain topics and less time consuming and computationally complex than analyses of all samples.

An important work in line with "all sample analysis" is for instance presented in Koenig, Lucero, and Perlman (2008) using Functional Data Analysis registration, a method for non-linear time warping. The method allowed them to decompose amplitude and time related variability of all samples and calculate an amplitude and a warping index for the time series, which were then fit into an ANOVA.

Another approach for looking into time varying behaviour and the voicing contrast has been proposed by Shih, Möbius, and Narasimhan (1999) who developed the so called "voicing profiles". For this purpose, the closure duration of stops and affricates were time-normalized and divided into 10 equidistant intervals. Based on several repetitions of the same phoneme in a certain context, the probability of the occurrence of voicing at each time step was calculated, showing the maintenance or disappearance of voicing over time. These voicing profiles have been calculated for various corpora and languages. They allow investigating the gradual changes of voicing probability in a given context and normalized time interval.

A relatively new statistical approach in the speech domain is based on the application of General Additive Mixed Models (GAMMs, see Wood, 2006). By using GAMMs, it is possible to statistically model nonlinear relations between continuous time series (more details are given in Section 2.5.2).

In the following section, we will describe our methodology in which we used both single time points analysis and GAMMs to investigate the relation between intraoral pressure and tongue-palatal contacts.

2. Methodology

2.1. Participants

Three males and three females ranging in age from 25 to 38 years took part in the study. All participants were native speakers of Standard Turkish. Two of the speakers lived in Berlin for two years, while the other four participants lived in Turkey and came to the phonetics laboratory at ZAS in Berlin for the purpose of the experiment. For each of them, a custom-made artificial palate was made. None of the participants had any known speech, language, or hearing disorders.

2.2. Experimental set-up

Three different systems were used simultaneously: (i) the acoustic signals were recorded on DAT (Tascam DA 20 MK II) at a sampling rate of 48 kHz via a Sennheiser MKH 20 P48 microphone, (ii) the EPG data were recorded by a Reading EPG 3 system at a sampling rate of 100 Hz, (iii) the intraoral pressure signal was recorded with a pressure sensor (Endevco 8507C-2) attached to the posterior end of the EPG palate (cf. Fig. 1). The sensor measured the difference between atmospheric pressure and intraoral pressure via a small tube passing through the teeth outside the oral cavity. The intraoral pressure signal was sampled with 6000 Hz.

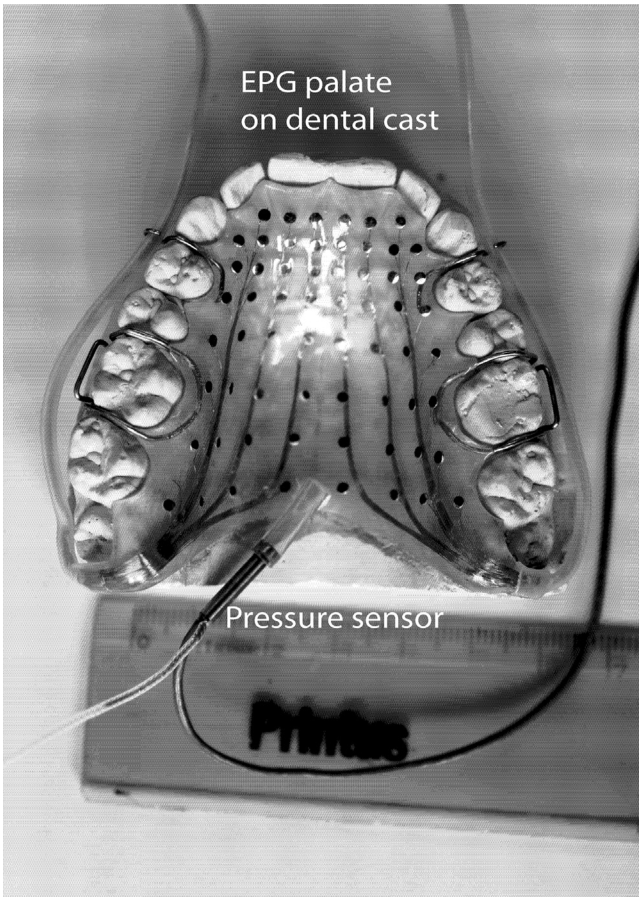

Fig. 1. Intraoral pressure sensor attached to the posterior end of an EPG palate.

2.3. Speech stimuli and procedures

This study was conducted as part of a larger experiment that investigated speech production in Turkish. Over the course of the experiment, participants read five randomized lists with 53 sentences. That is, each sentence was read five times in different positions in the list. Eight sentences which contained the alveolar /t, d/ and postalveolar sounds /tʃ, dʒ/ in each list were part of the present study. Bilabial stops were not included, because the production of bilabial closure cannot be measured with EPG. For a similar reason, velar stops were not included, because some closures may occur behind the end of the artificial palate and are therefore not detectable. Besides the alveolar stops, the affricates were included, because phonologically they belong to the stop category. Each of these target sounds was followed by either vowel /a/ or vowel /i/ in different words, following Koenig, Fuchs, and Lucero's (2011) experimental design. All sounds occurred in word-initial position of bisyllabic words and these words were placed in a carrier phrase, as illustrated in example (1). The target words occurred in the second position to avoid list and declination effects in repetitions of successive single target words.

(1) Arda *çabuk* anlamlı bir sözcüktür dedi.
(Arda said (that) '*quick*' is a meaningful word)

Every participant wore a custom-made artificial palate with an attached pressure sensor. Participants wore their palate for at least 30 min before the experiment. Once they became familiar with the artificial palates, they were instructed to read each sentence aloud at their normal speech rate.

The target sentences were displayed via PowerPoint on a computer screen. The experimenter used a pointer to change from one slide to the next, following the participant's pace.

2.4. Data labelling and pre-processing

In total, we recorded 240 tokens (6 speakers × 4 target stops × 2 following vowels × 5 repetitions). Each token was analysed separately in terms of acoustics, tongue palatal contacts and intraoral pressure.

The acoustic data were analysed manually using Praat (Boersma and Weenink, 2013; version 5.3.53) by labelling the onset of the target sound as the end of the preceding vowel (end of pronounced second formant), the offset of the target sounds as the beginning of the following vowel (beginning of pronounced second formant), the offset of voicing and the burst (see Fig. 2). The following parameters were calculated on the basis of these measures:

(1) Consonant duration = target offset–target onset.
(2) Closure duration = release–target onset.
(3) Voicing duration = voicing off–target onset.
(4) Percentage of voicing into closure = voicing duration *100/closure duration.
(5) VOT for /d, t/ = phonation onset – burst (in case of fully voiced stops, the onset of phonation was defined at the end of the preceding vowel).

Subsequently, we imported the acoustic landmarks into mview (Tiede, 2005), a MATLAB based tool to annotate the EPG recordings. With the help of this tool, we determined two landmarks: (a) the earliest time point after the end of the preceding vowel at which two additional EPG electrodes were activated. This landmark corresponded to the onset of closure; (b) the earliest time point at which the speaker produces full closure in the anterior region of the EPG palate. Based on these time landmarks we calculated the overall percentage of contact (PC), the percentage of contact in the anterior region (i.e. the percentage of contact in the first four rows of the artificial palate) and the centre of gravity (COG, a weighted index in the front-back dimension giving more weight to the anterior rows than the posterior ones; see Hardcastle, Gibbon, & Nicolaidis, 1991).

Before the intraoral pressure data could be annotated, they were filtered using a Kaiser window, with 40 Hz passband and 100 Hz stopband edges to remove vocal fold oscillations. Based on the filtered signal the first derivative (velocity) was calculated in MATLAB. Fig. 3 shows the raw and filtered intraoral pressure data and the two landmarks which were obtained. Both landmarks, the intraoral pressure peak (Pio Max) and the velocity peak (Vel Max) were annotated in the filtered data.

For the analyses of all data points using GAMMs, we considered all data points (i.e. all filtered intraoral pressure data and all PC values for EPG) from the end of the preceding vowel, determined by the acoustic signal, to the maximal intraoral pressure.

2.5. Statistical analyses

Prior to statistical analyses, we standardized each predictor variable by participant (centred and divided by one standard deviation). This permitted better estimates of the effects tested in our models. Statistical analyses are divided in two parts. In the first one, we will focus on selected measures taken at single time points and the second one refers to an all point analysis using GAMMs.

2.5.1. Single time point analyses

For the single time point analyses we used linear mixed-effects models (Baayen, 2008; Gelman & Hill, 2007; Pinheiro & Bates, 2000) as developed in the lme4 package (Bates et al., 2013) for the R software (R Core Team, 2013).

In order to test the effects of the continuous factors separately and to avoid multicollinearity issues, we ran several mixed effects models. Each model, except the one for VOT, incorporated as predictors the articulation manner (plosive

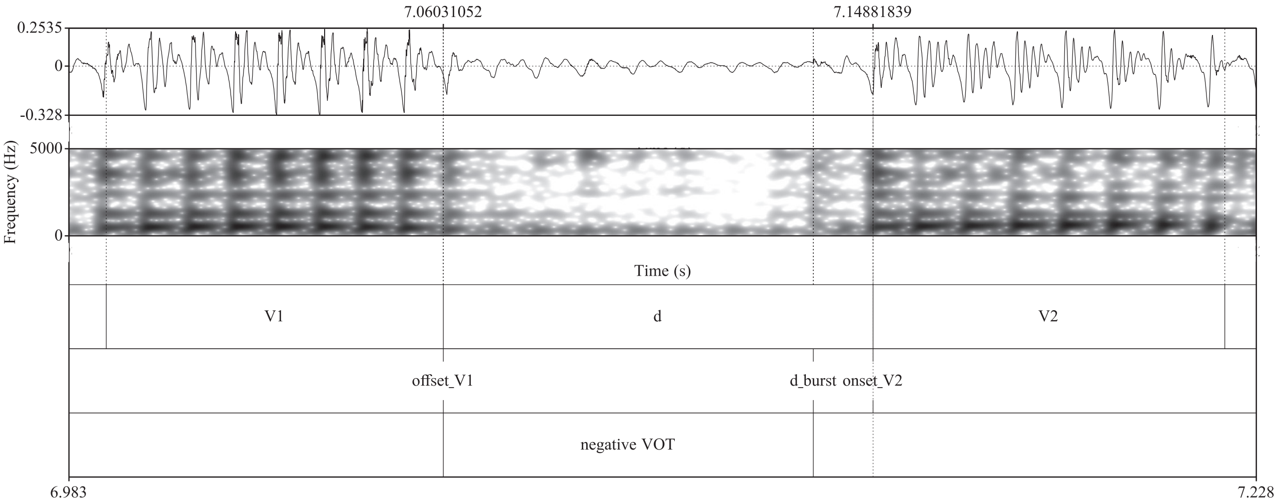

Fig. 2. The acoustic landmarks of a voiced /d/ annotated in Praat.

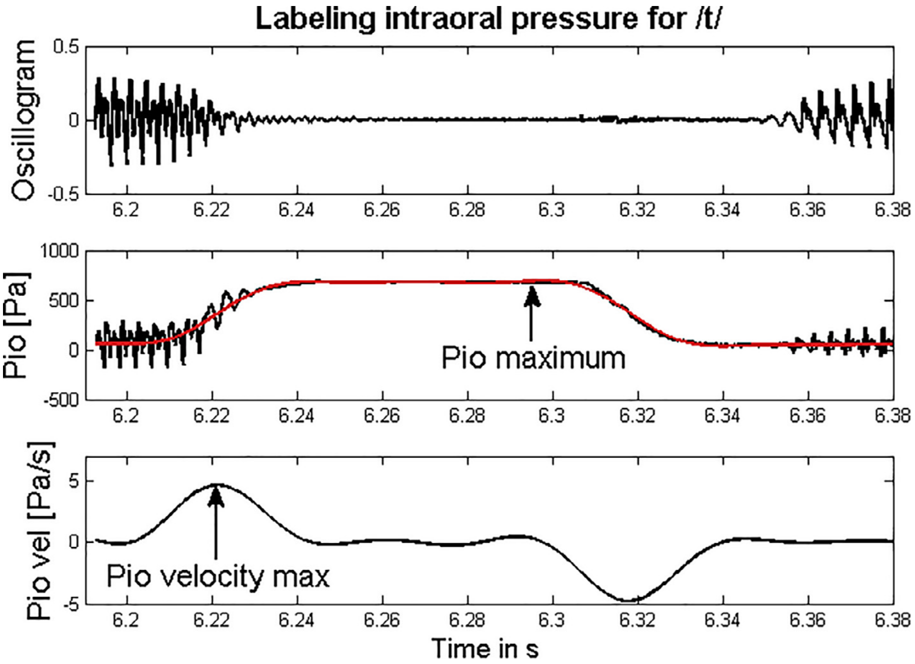

Fig. 3. Upper track: Acoustic signal; middle track: intraoral pressure (raw data in black and filtered data in grey (colour online), annotation of intraoral pressure maximum (Pio Max); lower track: intraoral pressure velocity of the filtered data with velocity maximum (Vel Max) during oral closure.

vs. affricative, reference level: plosive), the nature of the vowel (/a/ vs. /i/, reference level: /a/), voicing contrast (reference level: voiced) and their two- and three-way interactions. For each model, one of the following dependent variables were selected: the duration of the consonantal target (denoted as Target Dur), the percentage of voicing into oral closure (rel Voi) the percentage of anterior contact observed during full closure (Ant), the percentage of contact over the whole palate during full closure (PC), the Centre of Gravity at full closure (COG), the maximum of intraoral pressure (Pio Max) and the velocity maximum during the build-up of pressure when an oral closure is produced (Vel Max). All models had the same random effects structure including a speaker specific random intercept and a speaker specific random slope for each fixed factor. After running each model, non-significant interactions that did not contribute to improve the model fit (assessed by comparing the model residuals obtained with and without the interaction by Chi-square tests) were removed. For the model using VOT as the dependent variable, only the /d, t/ data was included. The predictors of this model were voicing contrast (reference level: voiced) and vowel (/a/ vs. /i/, reference level: /a/), as well as their interaction. Random effects were determined in a similar way as in the other models.

The *p*-values were obtained by Shatterwise approximation separately for each model via the lmerTest package for R (Kuznetsova, Brockhoff, & Christensen, 2015) and submitted to False Discovery Rate correction (Benjamini, Yoav, & Yekutieli, 2001).

2.5.2. All time point analyses

The second aim of our analyses was to estimate the nonlinear relation between intraoral pressure rise and tongue-palatal contacts. This analysis was conducted by means of a General Additive Mixed Model (GAMM, Wood, 2006) through which we predicted the values of intraoral pressure during pressure rise (from the end of the preceding vowel, defined in the acoustics, to the maximal intraoral pressure) for different manner and voicing conditions (see Appendix for details). Before describing the models designed for the current study, we wish to introduce a few basic concepts which can better facilitate the interpretation of the result obtained by fitting a GAMM. Since the approach followed in this work is that described by Wood (2006) and implemented in the MGCV package for R (Wood & Wood, 2017), the reader is referred to these works for details concerning the content of the next section.

2.5.2.1. General additive mixed models. GAMM differ from common Linear Mixed Models in regards to their potential to model nonlinear effects of continuous factors on observed variables. In a linear model, the values of an observed variable are predicted by multiplying the values of some fixed factors by the appropriate coefficients' values. In a GAMM, the observed variable can be predicted by multiplying some (or all) coefficients by smooth functions of the relative factors. A smooth function corresponds to a curve that represents the nonlinear effect of a predictor on the observed variable. The curve is obtained by linearly combining several simpler nonlinear functions of the predictor (basis functions) in such a way that the resulting curve is continuous and appears smooth. For example, if the smooth function is approximated via a cubic spline, the basis functions are cubic polynomials (see Fig. 4). A cubic polynomial is the lowest order polynomial displaying inflection points and it can be shown that the smoothest possible curve joining *n* points can be obtained by connecting the points through an equal number of cubic polynomials as done in Fig. 4. The basis adopted to build the curve in the figure has a strongly local character as each different polynomial approximates a different stretch of curve (the portion joining

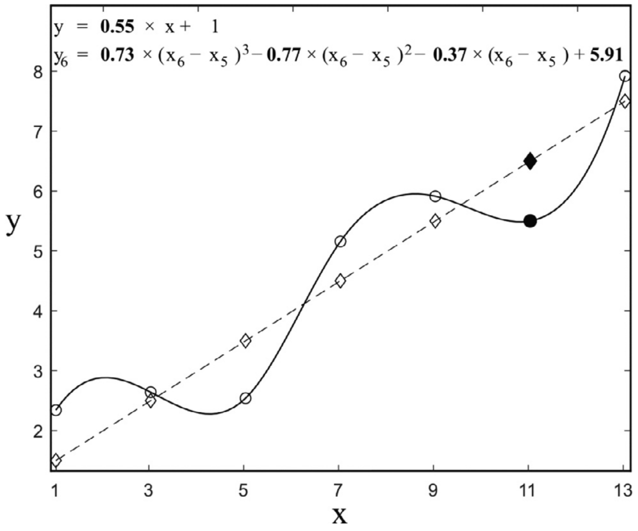

Fig. 4. Comparing how a linear model and a cubic spline model approximate the relation between two variables x and y. Empty circles: data points. Continuous line: spline model. Dashed line: linear model. Empty diamonds: linear predictions of the values of y given the values of x. These are obtained by applying the topmost formula to all values of x (the coefficients of the linear model, displayed in bold typeface in the formula, are equal for all values of x). The bottommost formula predicts the value of y corresponding to the 6th value of x (x = 11) according to the spline model (the coefficients of the spline model, bold typeface in the formula, change across values of x, because a different polynomial connects each pair of consecutive values of y). Filled circle: spline model prediction of y corresponding to x = 11. Filled diamond: linear prediction of y corresponding to x = 11.

two consecutive observed data points). This feature is not optimal in a regression strategy as it makes model comparison harder. This problem is addressed by adopting thin-plate regression splines (Wood, 2003). These smoothing functions based on cubic polynomials allow for low-rank approximations that permit maintaining a reasonable degree of model complexity even in the case of multiple interacting covariates.[1]

The following features differentiate a GAMM from Linear Mixed models:

- Several kinds of smooth functions can regulate the degree of smoothness of the modelled curve through a parameter usually denoted as λ. In order to determine the most appropriate value of this parameter a generalized cross validation approach is adopted. Once the smoothing parameter is determined, the model coefficients can be computed. Due to the computation of the smoothing parameter prior to model fitting, p-values in GAMM models are usually underestimated and particular care should be taken in their interpretation.
- In order to avoid over fitting, due to the generally high number of coefficients, GAMM coefficients are usually estimated by penalised likelihood maximization with the penalties suppressing wiggly estimates of the smooth function. Due to penalization, some coefficients play a small role or play no role at all in shaping the behaviour of the dependent variable. The number of coefficients required to model the effect of a predictor corresponds to the effective number of its degrees of freedom. This quantity is usually estimated because it indicates the complexity of the effect modelled and it is useful to determine if an effect is linear. Indeed, a linear effect is expected to have an estimated number of degrees of freedom equal to one.
- One of the core assumptions of linear modelling is the independence of the observations. This is not often true in GAMMs, because contiguous data points are usually correlated. To deal with autocorrelation of residuals, the degree of autocorrelation of the model residual is estimated and accounted for.

As linear mixed models, GAMMs can have both random intercepts and random slopes. However, in a GAMM, smooth functions can also be included in the random effects structure. Therefore, a specific smooth function can be used to model a nonlinear effect that is specific to the level of a random factor (as for example the speaker identity).

2.5.2.2. GAMM modelling for the relation between intraoral pressure and percentage of contact. In order to investigate the relation between intraoral pressure and percentage of tongue-palatal contacts we implemented a GAMM in which the values of Pio depend on a combination of categorical variables and smooth factors. The categorical variables were: manner (affricate vs. plosive, reference level: plosive), voicing contrast (voiced vs. voiceless, reference level: voiced) and their interaction. We also included a smooth predictor for PC (accounting for the effect of PC on Pio at the reference levels of the other factors), one smooth predictor for the combined effect of PC and manner (accounting for the differences between the effect of PC in plosives and affricates) and one smooth predictor for the combined effect of PC and voicing contrast (accounting for the differences between the effect of PC on voiced and voiceless stops). The random effects structure included a random intercept per participant (allowing for participant-specific reference Pio values at the mean PC, in voiced plosives), a random smooth for participant and voicing (introducing a random effect of PC on Pio for each combination of participant and level of the voicing factor), a random smooth per participant and manner (introducing a random effect of PC on Pio for each combination of participant and level of the manner factor).

3. Results

3.1. Single time point analyses

Fig. 5 provides a general overview of the measured variables in the acoustic, articulatory and aerodynamic domains and illustrates how they differ with respect to voiced and voiceless plosives and affricates. Note that though a further separation into different vowel contexts has not been included, so as to keep the figure clear and understandable, vowel context also affected the acoustic and aerodynamic data (see Table 1). At first glance, the most extreme differences and robust results between phonologically voiced and voiceless stops can be found in the percentage of voicing into closure (acoustics), in the velocity maximum of the intraoral pressure (aerodynamics) and VOT (for /t, d/, acoustics).

Turkish speakers produce voicing during the entire closure in almost all cases for the phonologically voiced /d/ resulting in a negative VOT and almost 100% of voicing during closure, while VOT is positive for /t/ and voicing during closure is limited in the measured phonologically voiceless consonants. These results are coherent with the rate of intraoral pressure rise, measured as the maximum velocity peak. In phonologically

---

[1] Note however that using this kind of smoothing functions is not appropriate when modeling interactions between continuous covariates defined on different scales.

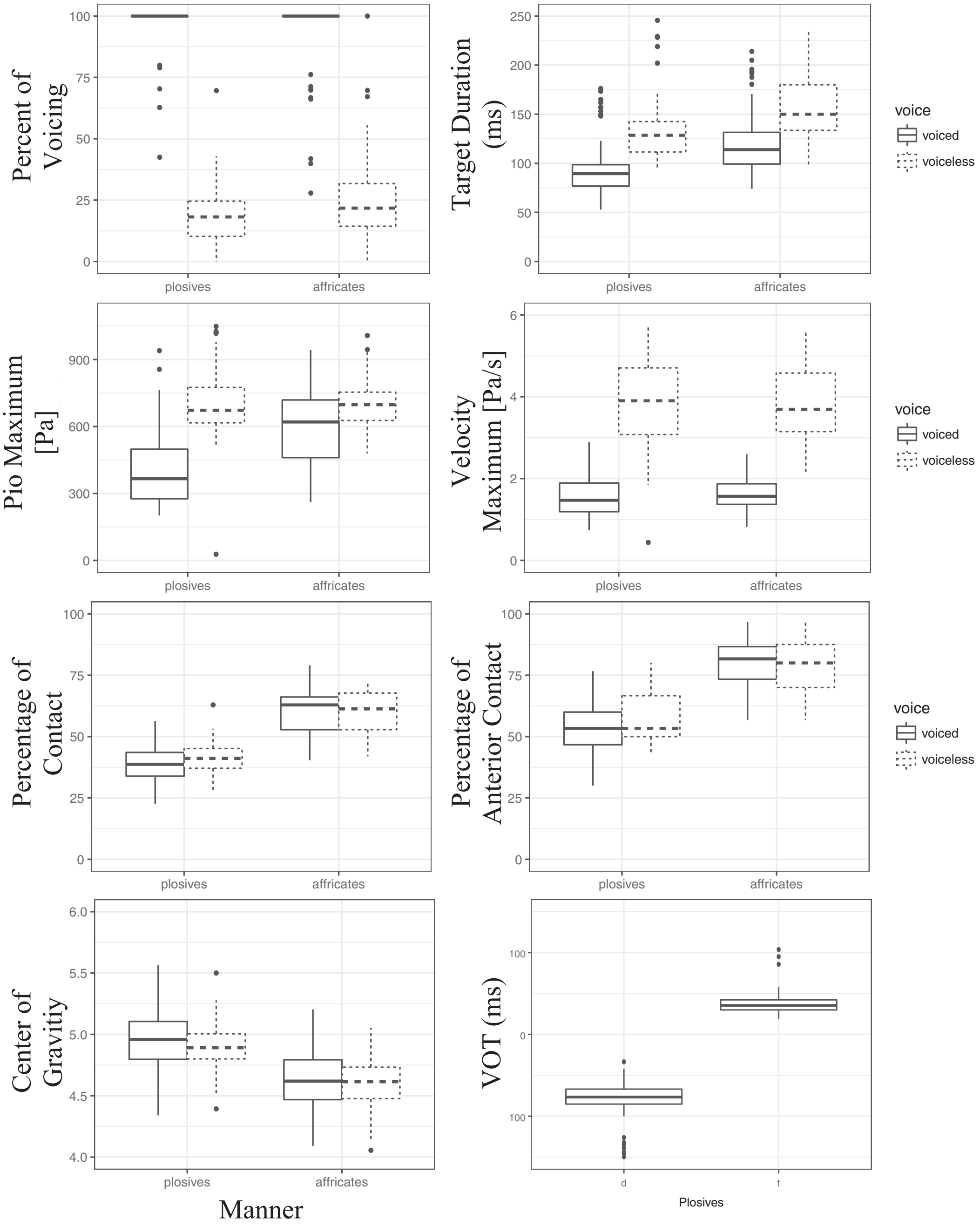

Fig. 5. Boxplots for measured dependent variables (y-axes) and different manner of articulation (affricates versus stops, x-axes). Phonologically voiced phonemes are represented by the continuous line plots, while the dashed line plots represent the phonologically voiceless phonemes. Data from all speakers have been collapsed.

voiceless consonants, intraoral pressure rises substantially faster than in voiced stops.

Fig. 5 also provides some evidence that the differences in the measured articulatory data regarding the voicing contrast are subtle. Table 1 provides a more in-depth analysis based on the linear mixed effect models.

Results for VOT show that as expected the voiced plosive /d/ has a negative VOT ($\beta = -77.063$, $t = -8.394$) while the voiceless plosive /t/ has a positive VOT ($\beta = 118.084$, $t = 8.394$). No other significant effects were observed. Our findings reveal that in the context of vowel /a/ phonologically voiceless plosives have a significantly longer overall duration ($\beta = 1.61$, $t = 14.23$), a smaller percentage of relative voicing during oral closure ($\beta = -1.996$, $t = -29.71$), a higher intraoral pressure peak ($\beta = 1.63$, $t = 7.51$) and a higher pressure velocity maximum ($\beta = 1.89$, $t = 21.98$). The three parameters for tongue palatal contact patterns (PC, Ant, COG) did not reveal a main effect regarding the voicing contrast. In the context of vowel /a/ these parameters differed between voiced stops, showing consistently larger percentage of anterior contacts ($\beta = 1.97$, $t = 9.59$) and overall contacts ($\beta = 2.35$, $t = 15.92$) as well as more posterior placement (COG) ($\beta = -0.79$, $t = -5.60$) in voiced affricate /dʒ/ than in voiced plosive /d/. These results may well be explained with the anticipatory preparation of an oral constriction for the production of frication after closure. We did not expect the following vowel context

Table 1
Results of the linear mixed effect models conducted to estimate the effect of voicing contrast (column VI), manner (column IV) and vowel context (column V) and their interactions (columns VII and IX). For each model estimates of the effects, *t*-values and adjusted *p*-values are given. Significant *p*-values (<0.05) are marked by bold typeface. Results for different coefficients are arranged in different columns. Results from different models are separated by empty rows.

I Dependent variable	II	III Intercept	IV Manner (ref.: plos.)	V Vowel (ref.: /a/)	VI Voicing contrast (ref.: voi.)	VII Manner × vowel	VIII Manner × voicing contrast	IX Vowel × voicing contrast
VOT	Estimate	−77.063		−5.380	118.084			2.564
	t value	−8.394		−0.603	8.398			0.254
	p value	**<0.01**		0.56	**<0.01**			0.799
TargetDur	Estimate	−1.008	0.329	−0.13	1.611	1.193	−0.335	
	t value	−6.508	1.539	−1.2	14.230	8.138	−2.287	
	p value	**<0.01**	0.705	1.000	**<0.01**	**<0.01**	0.128	
RelVoicDur	Estimate	1.032	−0.083	−0.01	−1.996		0.236	0.162
	t value	16.160	−1.253	−0.15	−29.711		3.046	2.084
	p value	**<0.01**	1.000	1.000	**<0.01**		**0.019**	0.205
PC	Estimate	−1.082	2.355	0.516	0.297	−0.353	−0.396	
	t value	−7.821	15.916	4.398	2.289	−2.772	−3.110	
	p value	**<0.01**	**<0.01**	**0.011**	0.247	**0.038**	**0.016**	
Ant	Estimate	−0.651	1.974	0.402	0.214	−0.313	−0.379	
	t value	−4.170	9.585	3.768	2.242	−2.639	−3.194	
	p value	**0.007**	**<0.01**	**0.019**	0.215	0.054	**0.013**	
COG	Estimate	1.019	−0.790	−0.34	−0.088	0.324	1.019	
	t value	9.042	−5.595	−4.08	−1.157	3.029	9.042	
	p value	**<0.01**	**0.001**	**0.007**	1.000	**0.019**	**<0.01**	
Pio Max	Estimate	−0.567	0.923	−0.58	1.632	0.514	−1.113	0.326
	t value	−3.178	5.482	−4.62	7.509	3.652	−7.908	2.314
	p value	**0.032**	**<0.01**	**<0.01**	**<0.01**	**0.003**	**<0.01**	0.123
Vel Max	Estimate	−0.397	−0.091	−0.44	1.892	0.296		
	t value	−3.818	−0.727	−5.25	21.976	2.520		
	p value	**0.010**	1.000	**<0.01**	**<0.01**	0.073		

(high versus low vowel) to already affect the mechanisms involved in the closure. However, in voiced plosive /t/ we observed a larger percentage of tongue palatal contact patterns at oral closure in /i/ than /a/ context (higher PC: $\beta = 0.52$, $t = 4.40$), a larger percentage of anterior contacts ($\beta = 0.40$, $t = 3.77$) and a more posterior articulation (lower COG values: $\beta = -0.34$, $t = -4.08$). Moreover, a lower pressure maximum was reached ($\beta = -0.58$, $t = -4.62$), pressure velocity values were lower ($\beta = -0.44$, $t = -5.25$) in /i/ than in /a/ context.

Besides the main effects, some significant interactions were also observed. Specifically, the manner × voicing contrast interaction revealed that the effect of the voicing contrast on relative duration of voicing into closure is weakened in affricates ($\beta = 0.24$, $t = 3.05$). However, both effects, the effect of manner on the percentage of contact over the whole palate and on the percentage of contact in the anterior region decrease significantly in voiceless stops ($\beta = -0.40$, $t = -3.11$ for the first interaction and $\beta = -0.38$, $t = -3.19$ for the second). Similarly, the effect of manner on COG is weaker in voiceless stops ($\beta = 1.019$, $t = 9.042$). This suggests that the fronting observed in the voiced affricate /dʒ/ in contrast to the voiced plosive /d/ is reduced in the voiceless affricate /tʃ/ in comparison to the voiceless plosive /t/. Finally, the effect of manner on the maximum intraoral pressure of stops is weaker in affricates ($\beta = -1.11$, $t = -7.91$).

The single time point analyses including acoustic, aerodynamic and articulatory data revealed that the Turkish voicing contrast affected the selected variables to different degrees. Specifically, robust differences could be found in the acoustic domain with respect to VOT and voicing during closure and in the aerodynamic domain concerning the velocity maximum of pressure rise. Selected data obtained from EPG at full oral closure did not show an involvement of supralaryngeal articulation in the voicing contrast of Turkish. However, this may simply be the consequence of the selected time point. We subsequently carried out an all point analysis.

3.2. All time point analyses: the relation between intraoral pressure and percentage of contacts

To provide a general overview, average trajectories for articulatory and aerodynamic data have been provided in Fig. 6. In these plots, data were time-normalized to 10 points and then averaged with respect to the phoneme, following vowel context and manner of articulation. The dashed lines depict the standard deviations. While intraoral pressure constantly rises in phonologically voiced stops, in phonologically voiceless stops the rise is steeper, i.e. it changes faster at the beginning (up to time step 4). Tongue-palatal contact patterns also show nonlinear behaviour over time, but the percentage of contact rises faster at the beginning than at the end of both phonologically voiced and voiceless stops. Thus, different relations between articulation and aerodynamics with respect to the voicing contrast can be expected.

These general observations are confirmed by the results of the statistical analysis. When interpreting the results of a GAMM, categorical predictors are analysed separately from smooth terms, because only categorical predictors are fully represented by the estimates of the model parametric coefficients. Results for the categorical predictors are displayed in Table 2.

The non-significant effect of manner concerns voiced stops. This means that although at the reference level of percentage

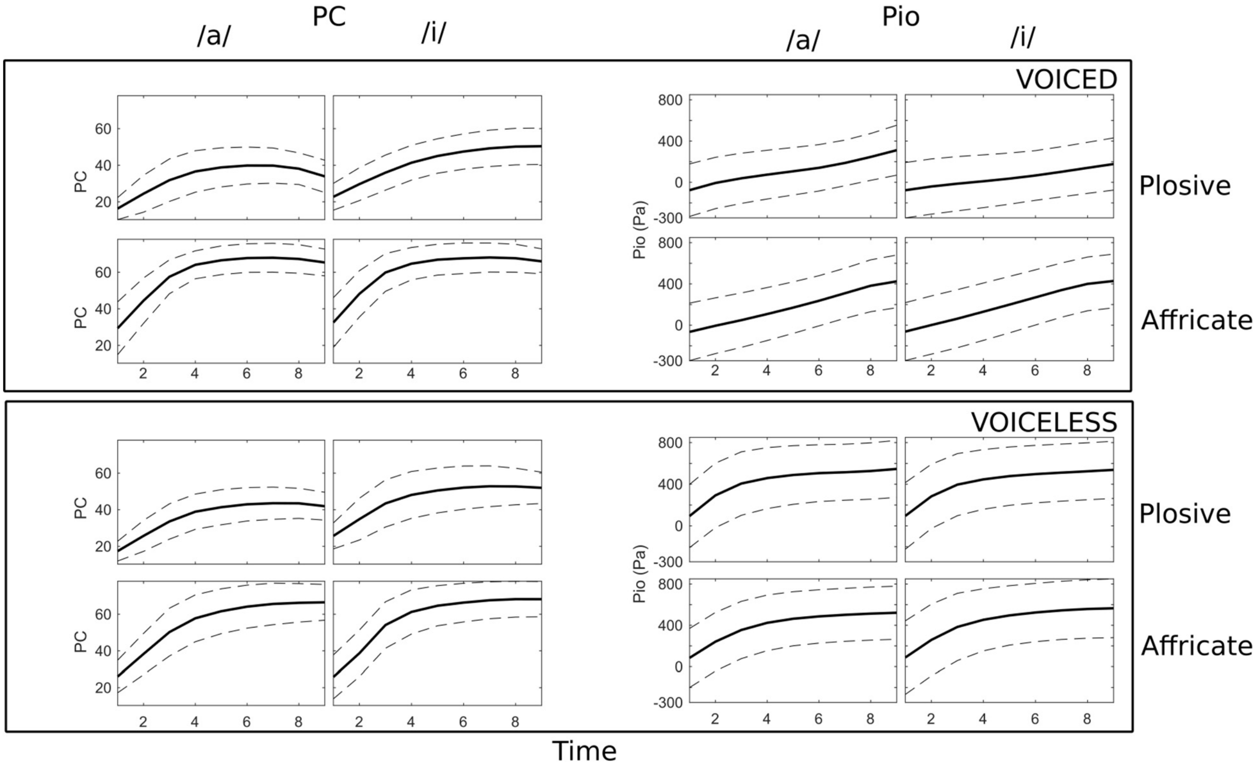

Fig. 6. Trajectories for intraoral pressure (3rd and 4th column) and percentage of tongue-palatal contacts (1st and 2nd column) for phonologically voiced stops (1st row), voiced affricates (2nd row), voiceless plosives (3rd) and voiceless affricates (4th row). All data are averaged from all speakers and time-normalized. The solid line displays the mean while the dashed lines correspond to the standard deviations. Vowel context is given in different columns (1st and 3rd column for /a/ and 2nd and 4th column for /i/).

Table 2
GAMM results for parametric coefficients with estimates, standard error (SE), *t*-values. Stars determine the level of significance with ***: $p < 0.001$, **: $p < 0.001$, *: $p < 0.05$.

	Estimate	SE	*t*-value
Intercept	273.18	72.79	3.75^{***}
Manner (reference: plosives)	−33.65	17.16	−1.96
Voicing contrast (reference: voiced)	166.21	33.31	4.99^{***}
Manner × Voicing contrast	−50.47	14.36	-3.51^{***}

Table 3
Results from smooth and random terms (random terms in italics) with the estimated degrees of freedoms (second column), the nominal degreed of freedom as determined by the model's coefficients (third column) and the F ratios (fourth column). All results are highly significant with $p < 0.001$: ***.

	Estimated df	Ref. df	*F*
PC	0.9925	9	13.70^{***}
PC: Manner (reference: plosive)	2.2933	5	2.04^{***}
PC: Voicing (reference: voiced)	3.5864	5	10.43^{***}
PC by participant and manner	14.45	35	1.44^{***}
PC by participant and voicing	21.04	35	4.47^{***}
Participant	4.87	5	37.30^{***}

of contacts (at the mean value of PC) intraoral pressure values tend to be smaller in phonologically voiced affricate /dʒ/ than in voiced plosive /d/, this tendency is not significant. However, we obtained clear effects for the voicing contrast. When PC is equal to its mean value, intraoral pressure is significantly higher in voiceless than in voiced plosive ($\beta = 166.21$, $t = 4.99$). Finally, the interaction between voicing contrast and manner is significant, too. This means that the difference in intraoral pressure, observed between the voiced and voiceless plosives when PC is equal to its mean value, is smaller in affricates ($\beta = -50.47$, $t = -3.51$).

Table 3 integrates data from all smooth and random terms. Note that the estimated number of degrees of freedom for the smooth term modelling the effect of PC in voiced plosive /d/ is practically one, indicating a linear relation during the production of /d/. In order to better evaluate the effects of these smooth terms on the dependent variable, we plotted the model predictions in Fig. 6 using the R package *itsadug* (van Rij, Wieling, Baayen, & van Rijn, 2015).

Fig. 7 shows a linear relation between percentage of contact and intraoral pressure in voiced stops as well as a nonlinear relation with a concave shape for voiceless stops. In voiced affricates, the relation is nonlinear and has a slightly convex shape. Fig. 8 provides a complementary picture, showing the estimated difference between different conditions, similarly to the way one would subtract the continuous curve from the dashed one (voiceless-voiced in Fig. 7) for the respective comparison. Moreover, the bold black line of the x-axes in Fig. 8 refers to the samples where the relation between intraoral pressure and percentage of contact differs significantly between the compared pairs.

This reveals that the relation between intraoral pressure and percentage of contact is rather similar in the very beginning of the stop, but when the difference in pressure reaches approximately 100 Pa, phonologically voiced and voiceless stops show a different relation. Affricates also differ in the first few milliseconds of oral closure when some tongue-palatal contacts are already made, but pressure has not yet built up.

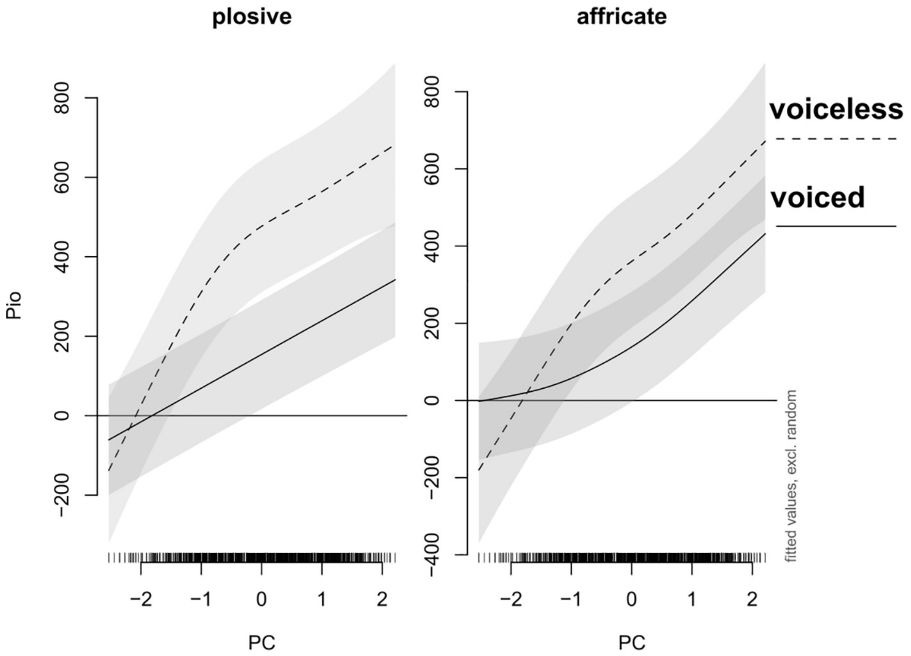

Fig. 7. GAMM predictions with intraoral pressure (y-axes) and PC on the x-axes. Line type refers to voiced (continuous) or voiceless (dashed) stops. Left: relation between voiced and voiceless stops, Right: relation between voiced and voiceless affricates.

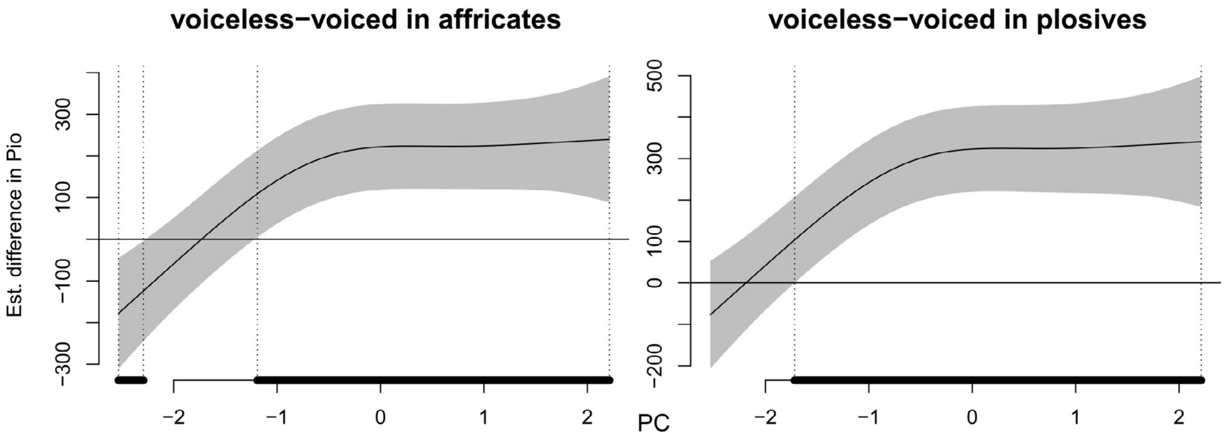

Fig. 8. Estimated differences between intraoral pressure and percentage of contact for the respective comparison (see subtitles) obtained by the GAMM. The thick lines overlapped on the x-axes depict the temporal windows in which the relation between the two variables differs.

4. Discussion

Our paper examines the relation between intraoral pressure and supralaryngeal articulation. Our aim was in particular to gain a deeper insight into the potential involvement of tongue motion in cavity enlargement mechanisms in Turkish, an as of yet under-investigated language. We carried out our work by means of a relatively unique experimental set-up combining acoustics, electropalatography and a piezoresistive pressure transducer.

Furthermore, following different approaches in the literature, we selected specific time points in the acoustic, aerodynamic and articulatory data, which have provided evidence for significant differences between phonologically voiced and voiceless plosives and affricates. Since single time point analysis includes the risk of missing some important information, especially if the choice of the time point was not appropriate, we added an all sample analysis based on GAMMs.

Our findings show that in accordance with previous work on Turkish (Öğüt et al., 2006; Feizollahi, 2010), VOT shows a clear distinction between phonologically alveolar voiced and voiceless plosives (/t, d/) in word-initial position, with a negative VOT for the voiced and a positive VOT for the voiceless and further emphasis the importance of this acoustic measure that has been used extensively in different languages. Moreover, even in word-initial position, in which devoicing is generally probable (Fuchs, 2005; Pape, Mooshammer, Hoole, & Fuchs, 2006), Turkish speakers produce phonologically voiced stops with almost full voicing during closure. However, voicing disappears quickly in phonologically voiceless stops. The findings are different from Kallestinova (2004), who reported on the basis of two Turkish speakers that /d/ would be produced as a voiceless unaspirated stop. The reasons for these differences might be manifold. We interpret the differences in light of different phonemic contexts, following Feizollahi (2010). His findings show that the preceding phonemic context affects phonologically voiced stops in word-initial position. In our study, the preceding word ended with a vowel and may have caused the maintenance of voicing in phonologically voiced stops. Hence, the surrounding context may be crucial for the phonetic realization of these voiced stops. Phonologically voiceless stop and affricates, on the other hand, resisted to contextual effects and voicing diminished very quickly, as was also found by Feizollahi (2010). We suppose that this resistance is a consequence of glottal abduction. Furthermore, we would like to mention that even on a speaker-specific basis (which was not explicitly discussed here) the difference in voicing was very robust and stable in our dataset.

These results gave us a perfect testbed for investigating the relation between intraoral pressure rise and supralaryngeal involvement for cavity enlargement. As Westbury (1983) described, many different possibilities exist to increase the size of the oral cavity to keep the intraoral pressure relatively low, a principle called motor equivalence (Perrier & Fuchs, 2015). A number of researchers provided evidence for different strategies, also involving different supralaryngeal articulators, e.g. laryngeal lowering, jaw lowering, nasal leakage, hyoid bone depression, an advanced tongue root, higher velum and different tongue compliance, as summarized earlier. None of these strategies was found for all places of articulations, different contexts, prosodic positions or across many speakers. In the present study, we were interested in potential involvement of the tongue to keep intraoral pressure low. Measurements of tongue motion were obtained indirectly, by means of tongue-palatal contact patterns, reflecting the different stages in which the tongue touches the palate during oral closure. This technique has the advantage of making tongue-palatal contacts visible not only in the mid-sagittal plane, but also distributed over the entire hard palate. It has, however, only a temporal resolution of 100 Hz, which is rather low in comparison to other articulatory techniques.[2] Our findings provide evidence for a linear relation between intraoral pressure evolution and tongue palatal contact patterns in voiced plosives in Turkish. Thus, the slow increase in intraoral pressure correlates with a gradual slow increase in tongue-palatal contacts. In voiced affricates, an additional strategy might be at play, because, especially in the beginning of the closure, a nonlinear relation between the aerodynamics and articulation can be found. However, this difference is short and subtle, as we did not find significant differences in the relation between voiced stops based on GAMMs.

In phonologically voiceless stops, the relation between aerodynamics and tongue-palatal contacts is non-linear. It begins as a linear relation, but we found a sort of turning point after which intraoral pressure no longer increases to the same extent as tongue palatal contacts increase. We interpret these findings in accordance with Müller and Brown's (1980) convex pressure profiles in voiceless stops. The beginning of the pressure increase may be primarily driven by the supralaryngeal articulators closing the vocal tract and the opening glottis, while the second phase in which intraoral pressure does not rise to a large extent any more may be explained with respect to a closed vocal tract and widely open glottal configuration. A limited amount of air can still be delivered from the lungs through the open glottis into the closed oral cavity (e.g. Fuchs, 2005 for an overview on coordinated actions). Hence, one way of interpreting the non-linear relation between intraoral pressure and tongue-palatal contacts may be to relate it to the different coordinated actions involved in the production of the voicing contrast. Further analyses for different languages and datasets are required to come to more definite conclusions.

The selected time point analyses gave us further clues pertaining to differences in the production of phonologically voiced and voiceless stops in Turkish. Our findings revealed clear differences in temporal patterns of the acoustic signal, i.e. in VOT, relative voicing duration and in overall target duration, with the former being more pronounced than the latter. Intraoral pressure data, i.e. the intraoral pressure peak and the velocity maximum (reflecting the slope of pressure rise), also showed the expected patterns with higher peaks and a higher velocity maximum for phonologically voiceless stops. These results for Turkish are in agreement with some earlier work on other languages (e.g. American English: Arkebauer et al., 1967; Stathopoulos, 1986; Subtelny et al., 1966; Müller & Brown, 1980; Koenig & Lucero, 2008; Danish: Fischer-Jørgenson, and Hansen, 1959; Polish: Zygis, Fuchs and Koenig, 2012). Tongue palatal contacts, however, did not show any difference, neither over the whole palate, nor in the anterior portion, nor in the frontness in place of articulation (measured as the COG parameter). These results for Turkish are thus similar to the ones reported by Fletcher (1989) on American English. They differ from findings reported by Moen and Simonsen (1997) and Moen et al. (2001) for Norwegian and Dixit (1990) for Hindi. Hence, how much the amount of tongue-palatal contacts differs may be language specific and also highly depend on the speech material and the prosodic structure of the respective language. In addition, tongue-palatal contacts are also affected by the height of the following vowel (with more percentage of contact in high vowel contexts) and whether the alveolar is a stop or affricate (more contacts for the affricate in preparation of the oral constriction phase).

Taking into account all of our results for acoustics, aerodynamics and articulation, there is an obvious mismatch between the different domains. We think that these differences are primarily a consequence of the choice of the time point. Tongue-palatal contacts may show different behavior for phonologically voiced and voiceless stops at a very early stage, starting when the first contacts of closure are made. However, these early time points are often not quantitatively assessed in the literature. Therefore, we conclude that a more global analysis may provide more reliable findings. That should not imply that every researcher must now apply GAMMs to their data. Even the visual inspection of the relevant time series already constitutes an important step towards making choices about which time point to choose and why. However, when there are reasons to believe that different variables are related in a nonlinear way, GAMMs are a powerful tool to characterize their relations. Despite this important feature, the

---

[2] We ran a few tests to combine the piezoresistive pressure sensor with Electromagnetic Articulography (EMMA), but since the sensor is made of metal and very close to the tongue coils, it yielded several errors and artefacts of the tongue coils.

application of GAMMs strongly depends on the data at hand. They still require a lot of expertise and understanding to parameterize the analysis and especially to interpret the results. Finally, the status of the significance values obtained by the application of GAMMs is not completely clear, suggesting that using GAMMs for hypothesis testing is a hazardous practice.

We would also like to mention that our work has some limitations. The dataset we analyzed is limited due to the constraints of the EPG palate on speech articulation and due to the costs of the custom made artificial palates. In our study, we focused on phonologically voiced and voiceless stops/affricates in intervocalic position across a word boundary (V#CV) and all findings may be specific to that particular position. It is quite possible that phonologically voiced stops devoice when they are preceded by a phonologically voiceless obstruent. Moreover, for the single time point analysis, we have taken a selection of measures which are often described in the literature, but it would have been possible to have included others. Nevertheless, we strongly believe that our work was one of the necessary steps opening new avenues for future research on the interplay between the different interacting processes underlying the production of the voicing contrast.

Acknowledgements

We thank our informants for participating in the experiment, Joerg Dreyer for technical expertise and Olivia Maky for proofreading. We also thank Mark Tiede for invaluable MATLAB scripts he shared with us and the maintainer Mark's Speechblog for the Praat scripts. This research has been supported by the Council of Higher Education in Turkey to Özlem Ünal-Logacev and by a grant (01UG1411) from the Ministry for Education and Research (BMBF, Germany) as well as the Leibniz Association (Germany) to Susanne Fuchs. Leonardo Lancia's work, carried out within the Labex EFL (ANR-10-LABX-0083) and BLRI (ANR-11-LABX-0036) and the Institut Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government, managed by the French National Agency for Research (ANR), under the program "Investissements d'Avenir" and the Excellence Initiative of Aix-Marseille University (A*MIDEX). We thank all researchers who have been spending enormous time and efforts in developing various tools for the general research community for free. Without their continuous work, our study would not have been possible in such a time period.

Appendix: Models' specification

A: LMMs

Eq. (1) represents the initial linear mixed model fitted to estimate the effects of the factors *manner*, *voicing contrast* and *vocalic context* on the values of the following scalar variables: *Target Dur*, *rel Voi*, *PC*, *COG*, *Pio Max*, *Ant* and *Vel Max*. After fitting each model, non-significant interactions that did not improve the model fit (according to a Chi-square test comparing the residuals obtained with and without each interaction) were removed.

$$\begin{aligned} Y_i = {} & \beta_0 + b_{0s} + (\beta_1 + b_{1s})X_1 + (\beta_2 + b_{2s})X_2 + (\beta_3 + b_{3s})X_3 \\ & + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3 \ + e_i, (b_{0s}, b_{1s}, b_{2s}, b_{3s}) \\ & \sim MVN(0, U), e_i \sim N\left(0, \sigma^2\right), \end{aligned}$$

$$U = \begin{bmatrix} \tau_0^2 & \rho_{01} & \rho_{02} & \rho_{03} \\ \rho_{01} & \tau_1^2 & \rho_{12} & \rho_{13} \\ \rho_{02} & \rho_{12} & \tau_2^2 & \rho_{23} \\ \rho_{03} & \rho_{13} & \rho_{23} & \tau_3^2 \end{bmatrix}. \tag{1}$$

Y_i is the ith observation of the continuous dependent variable.

X_{ji} (with $\in \{1, \ldots, 3\}$) represents a cell in the model design matrix with factors arranged column-wise according to the following order: *manner*, *vocalic context*, *voicing contrast*.

The β_j terms $(\beta_0, \ldots,\ \beta_6)$ are the coefficients of the fixed effects.

The b_{ks} terms $(b_{0s}, \ldots, b_{3s})$ are the random coefficients for the speaker-specific intercept (b_{0s}) and slopes ($b_{1s}, \ldots, b_{3s}$, with $s \in \{1, \ldots, m\}$, where m is the number of speakers). These coefficients are jointly drawn from a multivariate normal distribution with variance parameters τ_j and covariance parameters ρ_{mn} (with $m, n \in \{0, \ldots, K\}$, where K is the number of random slopes).

e_i is a random term drawn from a normal distribution with mean 0 and standard deviation equal to σ.

The model used to estimate the effects of voicing contrast and vocalic context on VOT is the same as the model in (1) but with only two predictors and therefore with only one interaction.

B. GAMMs

In order to estimate the relation between *PC* and *PIO* we fitted the two GAMM models in Eqs. (2) and (3). The differences between the models are limited to the residual term e_i.

$$\begin{aligned} Y_i = {} & \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + f_1(X_{3i}) + X_{1i} f_2(X_{3i}) \\ & + X_{2i} f_3(X_{3i}) + f_{0S} + X_{1i} f_{1S}(X_{3i}) + X_{2i} f_{2S}(X_{3i}) + e_i \end{aligned}$$

$$e_i \sim N\left(0, \sigma^2\right) \tag{2}$$

$$\begin{aligned} Y_i = {} & \beta_0 + b_{0S} + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + f_1(X_{3i}) + X_{1i} f_2(X_{3i}) \\ & + X_{2i} f_3(X_{3i}) + X_{1i} f_{1S}(X_{3i}) + X_{2i} f_{2S}(X_{3i}) + \rho e_{i-1} + \epsilon_i \end{aligned}$$

$$\epsilon_i \sim N\left(0, \sigma^2\right) \tag{3}$$

Y_i is the *i*th observation of the variable *Pio*.

X_{ji} represents a cell in the model design matrix with factors arranged column-wise according to the following order: *manner*, *voicing*, *PC*.

The β_j terms $(\beta_0, \ldots,\ \beta_6)$ are the coefficients of the scalar effects.

The smooth functions of the *PC* variable are represented by the $f_j(X_{3i})$ terms.

The random (speaker-specific) intercept is represented by the b_{0S} term, while the speaker specific smooth functions of *PC* are represented by the $f_{jS}(X_{3i})$ terms.

e_i is a random term drawn from a normal distribution with mean 0 and standard deviation equal to σ.

In (3) ρ is the lag-1 correlation between the residuals of the model in (2).

ϵ_{si} is a random variable drawn from a normal distribution with mean 0 and standard deviation equal to σ.

References

Arkebauer, H. J., Hixon, T. J., & Hardy, J. C. (1967). Peak intraoral air pressures during speech. *Journal of Speech Language and Hearing Research, 10*(2), 196–208.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Bates D., Maechler M., Bolker B. (2013). lme4: Linear-mixed Effects Models Using S4 Classes. Retrieved from: http://CRAN.R-project.org/package=lme4 (R package version 0.999999-2).

Bell-Berti, F., & Hirose, H. (1971). Velopharyngeal function in oral/nasal articulation: an electromyographic study. In Meeting of the American speech-language-hearing association, Chicago, Ill. November 1971. Asha, 13,528. (A).

Bell-Berti, F., & Hirose, H. (1972). Velar activity in voicing distinction: A simultaneous fiberoptic and electromyography study. *Haskins Laboratories Status Report on Speech Research, SR-31/32*, 223–230.

Bell-Berti, F. (1975). Control of pharyngeal cavity size for English voiced and voiceless stops. *Journal of the Acoustical Society of America, 57*(2), 456–461.

Benjamini, Yoav, & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics, 29*(4), 1165–1188.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer [Computer program]. Version 5.3.53 retrieved 04 October 2013 from http://www.praat.org/).

Dagenais, P. A., Lorendo, L. C., & McCutcheon, M. J. (1994). A study of voicing and context effects upon consonant linguapalatal contact patterns. *Journal of Phonetics, 22*, 225–238.

Dixit, R. P. (1990). Linguotectal contact patterns in the dental and retroflex stops of Hindi. *Journal of Phonetics, 18*, 189–201.

Ewan, W. G., & Krones, R. (1974). Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics, 2*, 327–335.

Feizollahi, Z. (2010). *Two case studies in the phonetics phonology interface: Evidence from Turkish voicing and Norwegian coalescence* (Unpublished doctoral dissertation). Georgetown University.

Fischer-Jørgensen, E., & Hansen, A. T. (1959). An electrical manometer and its use in phonetic research. *Phonetica, 4*(1), 43–53.

Flege, J. E. (1983). The influence of stress, position, and utterance length on the pressure characteristics of English /p/ and /b/. *Journal of Speech Language and Hearing Research, 26*(1), 111–118.

Fletcher, S. (1989). Palatometric specification of stop, affricate, and sibilant sounds. *Journal of Speech and Hearing Research, 32*, 736–748.

Fuchs, S. (2005). *Articulatory correlates of the voicing contrast in alveolar obstruent production in German* (Ph.D. thesis). Queen Margaret University College.

Fuchs, S., & Koenig, L. L. (2009). Simultaneous measures of electropalatography and intraoral pressure in selected voiceless lingual consonants and consonant sequences of German. *Journal of the Acoustical Society of America, 126*(4), 1988–2001.

Fujimura, O., & Miller, J. E. (1979). Mandible height and syllable-final tenseness. *Phonetica, 36*, 263–272.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature, 462* (7272), 502–504.

Hardcastle, W. J., Gibbon, F., & Nicolaidis, K. (1991). EPG data reduction and their implications for studies of lingual coarticulation. *Journal of Phonetics, 19*, 251–266.

Kallestinova, E. (2004). Voice and aspiration of stops in Turkish. *Folia linguistica, 38*(1–2), 117–144.

Kent, R. D., & Moll, K. L. (1969). Vocal-tract characteristics of the stop cognates. *Journal of the Acoustical Society of America, 46*(6), 1549–1555.

Koenig, L. L., Fuchs, S., & Lucero, J. C. (2011). Effects of consonant manner and vowel height on intraoral pressure and articulatory contact at voicing offset and onset for voiceless obstruents. *Journal of the Acoustical Society of America, 129*(5), 3233–3244.

Koenig, L. L., & Lucero, J. C. (2008). Stop consonant voicing and intraoral pressure contours in women and children. *Journal of the Acoustical Society of America, 123* (2), 1077–1088.

Koenig, L. L., Lucero, J. C., & Perlman, E. (2008). Speech production variability in fricatives of children and adults: Results of functional data analysis. *Journal of the Acoustical Society of America, 124*(5), 3158–3170.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2015). Package 'lmerTest'. R package version 2.0.

Lisker, L. (1970). Supraglottal air pressure in the production of English stops. *Language and Speech, 13*(4), 215–230.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384–422.

Lubker, J. F., & Parris, P. J. (1970). Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates /p/ and /b/. *Journal of the Acoustical Society of America, 47*(2B), 625–633.

Malécot, A. (1970). The lenis-fortis opposition: Its physiological parameters. *Journal of the Acoustical Society of America, 47*(6B), 1588–1592.

Moen, I., & Simonsen, H. G. (1997). Effects of voicing on /t, d/ tongue/palate contact in English and Norwegian. In *Proceedings of Eurospeech* 97, Greece (pp. 2399–2402).

Moen, I., Simonsen H., Huseby, M., & Grue, J. (2001). The relationship between intraoral air pressure and tongue/palate contact during the articulation of Norwegian /t/ and /d/. In *Proceedings of Eurospeech*, Aalborg, Denmark, September 3–7, 2001.

Mooshammer C., Geumann A., Hoole P., Alfonso P., van Lieshout P., & Fuchs S., (2003). Coordination of lingual and mandibular gestures for different manners of articulation. In *Proceedings of the 15th International Congress of Phonetic Sciences*. 81–84.

Müller, E. M., & Brown, W. S. (1980). Variations in the supraglottal air pressure waveform and their articulatory interpretation. In N. Lass (Ed.). *Speech and language: Advances in basic research and practice* (Vol. 4, pp. 318–389). Madison, WI: Academic.

Öğüt, F., Kılıç, M. A., Engin, E. Z., & Midilli, R. (2006). Voice onset times for Turkish stop consonants. *Speech Communication, 48*(9), 1094–1099.

Pape, D., Mooshammer, C., Hoole, P., & Fuchs, S. (2006). Devoicing of word-initial stops: A consequence of the following vowel? In J. Harrington & M. Tabain (Eds.), *Speech production: Models, phonetic processes, and techniques*. New York: Psychology Press.

Perkell, J. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Cambridge, MA: MIT Press.

Perrier, P., & Fuchs, S. (2015). Motor equivalence in speech production. In M. Redford (Ed.), *Handbook of Speech Production* (pp. 225–247). John Wiley & Sons, Inc.

Petersen, N. R. (1983). The effect of consonant type on fundamental frequency and larynx height in Danish. *Annual Report of the Institute of Phonetics, 17*, 55–86. University of Copenhagen.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

R Core Team (2013). *R: A Language and Environment for Statistical Computing [Computer software manual]*. Vienna: Austria. Retrieved from: (Version 3.0.0) http://www.R-.org/.

Riordan, C. J. (1980). Larynx height during English stop consonants. *Journal of Phonetics, 8*, 353–360.

Searl, J., & Evitts, P. M. (2013). Tongue-palate contact pressure, oral air pressure, and acoustics of clear speech. *Journal of Speech Language and Hearing Research, 56* (3), 826–839.

Shih, C., Möbius, B., & Narasimhan, B. (1999). Contextual effects on consonant voicing profiles: A cross–linguistic study. In *Proceedings of the 14th International Congress of Phonetic Science*, San Francisco, CA (Vol. 2, pp. 989–992).

Solé, M.-J. (2011). Articulatory adjustments in initial voiced stops in Spanish, French and English. In *Proceedings of the 17th International Congress of Phonetic Science* (Vol. 17, pp. 1878–1881).

Stathopoulos, E. T. (1986). Relationship between intraoral air pressure and vocal intensity in children and adults. *Journal of Speech Language and Hearing Research, 29*(1), 71–74.

Subtelny, J., Worth, J., & Sakuda, M. (1966). Intraoral pressure and rate of flow during speech. *Journal of Speech and Hearing Research, 9*, 498–518.

Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S., & Wilhelms-Tricarico, R. (1997). Tongue surface displacement during bilabial stops. *Journal of the Acoustical Society of America, 102*(1), 562–571.

Tiede, M. (2005). *MVIEW: Software for visualization and analysis of concurrently recorded movement data*. New Haven, CT: Haskins Laboratories.

Vatikiotis-Bateson, E. V., Barboso, A. V., & Best, C. T. (2014). Articulatory coordination of two vocal tracts. *Journal of Phonetics, 44*, 167–181.

Warren, D. W., & Wood, M. T. (1969). Respiratory volumes in normal speech: A possible reason for intraoral pressure differences among voiced and voiceless consonants. *Journal of the Acoustical Society of America, 45*(2), 466–469.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America, 73*(4), 1322–1336.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65*(1), 95–114.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Florida: Chapman and Hall/CRC.

Wood, S., & Wood, M. S. (2017). Package 'mgcv'. R package version, 1-7.

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2015). itsadug: Interpreting time series and autocorrelated data using GAMMs. R package version, 1(1).

Żygis, M., Fuchs, S., & Koenig, L. L. (2012). Phonetic explanations for the infrequency of voiced sibilant affricates across languages. *Laboratory Phonology, 3*(2), 299–336.