Journal of Phonetics 83 (2020) 101016

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

A dynamic model of the change from pre- to post-aspiration in Andalusian Spanish

Johanna Cronenberg [a,*], Michele Gubian [a], Jonathan Harrington [a], Hanna Ruch [b]

[a] Institute of Phonetics and Speech Processing, Ludwig Maximilian University of Munich, Germany
[b] University Research Priority Program Language and Space, University of Zurich, Switzerland

A R T I C L E I N F O

Article history:
Received 15 January 2020
Received in revised form 11 September 2020
Accepted 19 September 2020

Keywords:
Sound change
Functional principal component analysis
Andalusian Spanish
Aspiration

A B S T R A C T

In Andalusian Spanish, there is a well-documented sound change in which pre- has become post-aspiration in sequences of /s/ followed by voiceless stops. Here we investigate acoustically its synchronic basis across two age groups and two different regions of Andalusia that differ in the degree to which the sound change has advanced. For this purpose, Functional Principal Component Analysis (FPCA) was applied to the probability of voicing and to the degree of closure that had been estimated from the speech signal extending between the two vowels on either side of the aspirated cluster. The first principal component derived from FPCA was mostly associated with changes to the timing of the closure. Earlier closures were characteristic of both younger and West Andalusian speakers and of alveolar stops. In the signals parametrised by the first PC score, post- and pre-aspiration were found to be acoustically inversely related to each other and predictable from closure timing. The general conclusion is that the sound change by which pre- evolves into post-aspiration is a derivative of resynchronising the closure relative to the voiceless interval that emerges after decomposing speech signals varying over a wide range of speakers into principal components of variation.

1. Introduction

The southern varieties of Spanish are derived from 13th century Castilian Spanish, when groups of speakers from different, yet mutually intelligible dialects came into contact with each other during the *Reconquista* (Villena-Ponsoda, 2008). The resulting dialects, i.e. the regiolects of Andalusia, Extremadura, the Canary Islands, and – because of emigration from Andalusia to South America – some American varieties of Spanish, have since undergone many changes, among them the lenition of the voiceless alveolar sibilant /s/ (Canfield, 1981; Villena-Ponsoda, 2008). This sound change, that dates back to the beginning of the 18th century according to Mondéjar Cumpián (2001) and Terrell (1980), manifests itself as /s/-debuccalisation and affects the sibilant in a wide range of positions: word-medially before consonants as in *este* /e^{h}te/ (engl. this), word-finally before consonants as in *las toman* /la^{h}toman/ (engl. they take them), word-finally before vowels as in *las alas* /la^{h}ala/ (engl. the wings), and in some cases even word-initially as in *sí* /^{h}i/ (engl. yes) (Momcilovic, 2009; Torreira, 2006). This type of reduction affects not only /s/ but also other fricatives in Andalusian Spanish. Mondéjar, for instance, shows that words like *ajo* (engl. garlic) have changed their place of articulation from uvular /χ/ to glottal /h/: /aχ o/ > /a^{h}o/ (Mondéjar Cumpián, 2001). Also, the interdental fricative /θ/ can be lenited to /h/ in syllable-final position in the South of Spain, such as in *voz* /voθ/ > /vo^h/ (engl. voice). The debuccalisation of fricatives is a fairly common process in the languages of the world (Solé, 2010; Terrell, 1980). Consider for example the existence of prefixes such as *super-/hyper-, sex-/hex-, semi-/hemi-* in present-day English. Such alternations between /s/ and /h/ derive from borrowings from Latin, which did not lenite /s/ to /h/, and Classical Greek which has developed away from its Proto-Indo-European (PIE) roots and has undergone /s/-debuccalisation, e.g. PIE *septm* > Cl. Greek *hϵptá* (engl. seven) (Ohala, 1993).

The debuccalisation of /s/ before voiceless plosives has led to the development of pre-aspirated plosives in Andalusian Spanish, e.g. in *esquina* /e^{h}kina/ (engl. corner) (Ruch & Harrington, 2014; Torreira, 2007). Romero (1994) links the

* Corresponding author at: Institute of Phonetics and Speech Processing, Ludwig Maximilian University of Munich, Schellingstr. 3 (VG, room 224), D-80799 Munich, Germany.

E-mail address: johanna.cronenberg@campus.lmu.de (M. Gubian).

https://doi.org/10.1016/j.wocn.2020.101016
0095-4470/

historical development of pre-aspiration to the production of /s/ with a laminal place of articulation in Andalusian as opposed to the post-alveolar, apical production of Castilian Spanish /s/. Romero (1994) also suggests that pre-aspiration may have been more likely to arise in Andalusian than in Castilian Spanish because, in comparison with an apical /s/, the gestures in producing a laminal /s/ are slower and less extensive as a result of which a laminal /s/ is more likely to be lenited.

The phonetic characteristics of the sibilant in /sC/ vary due to a number of factors such as stress, lexical frequency, and speech rate (Alvar, 1955; Parrell, 2012; Ruch & Peters, 2016; Terrell, 1980; Villena-Ponsoda, 2008; Torreira, 2006). There is mixed evidence about whether /s/-aspiration is accompanied by quality differences in the preceding vowel (see Herrero de Haro, 2017 for a review). Auditory impressions suggest that the preceding vowel has a more open quality if /s/ is aspirated or deleted (Navarro Tomás, 1938) leading to singular/plural distinctions in East Andalusian Spanish in final and principally mid vowels if the following /s/ is completely deleted, e.g. *paso* [ˈpaso] (engl. step) vs. *pasos* [ˈpasɔ] (engl. steps) (Hualde & Chitoran, 2016; Hualde and Sanders, 1995). Such differences in quality in mid vowels may also spread anticipatorily to the stem in a form of metaphony leading to distinctions such as *perro* [pe̝ro] (engl. dog) vs. *perros* [pe̞rɔ] (engl. dogs). However, further experimental data is needed to support these impressions (Torreira, 2007). Gerfen (2002) also provides some evidence of a greater duration of the consonant closure and of a correspondingly lesser duration of the vowel in pre-aspirated East Andalusian post-vocalic /sC/ words (e.g. *pasta* /pasta/ [paʰtːa], engl. pasta) than in corresponding singleton /C/ words (e.g. *pata* /pata/ [pata], engl. paw), but only on the assumption that vowel duration was defined to extend from its acoustic onset to the offset of pre-aspiration.

The phonetic characteristics of /s/ in Andalusian Spanish also vary by age and regional origin of the speakers. Pre-aspiration in word-medial /sC/ clusters fairly consistently occurs in the speech of older speakers from Granada (East Andalusia) and may be accompanied by breathy voice during the preceding vowel or a longer closure duration (Gerfen, 2002; Torreira, 2006). Younger speakers from Seville (West Andalusia), on the other hand, were found to produce pre-consonantal /s/ either as pre-aspiration, or as post-aspiration, e.g. /ekʰina/, or as a combination of both, e.g. /eʰkʰina/ (Ruch, 2013; Ruch & Harrington, 2014; Ruch & Peters, 2016; Torreira, 2012). These findings provide evidence for a sound change in progress from pre-aspirated to post-aspirated voiceless stops in Andalusian Spanish that, so far, has taken stronger hold in younger speakers from West Andalusia than in older speakers from East Andalusia.

Regular sound change is often directional due to the existence of a phonetic bias that promotes the change to work in one, but not the contrary direction (Harrington, Kleber, Reubold, Schiel, & Stevens, 2018; Labov, 1994). In the case of Andalusian Spanish, it is a faster speech rate that favours a decrease in pre-aspiration and an increase in post-aspiration. Parrell (2012) has shown that the /st/ cluster in the word *pastándola* (engl. grazing it) exhibits more post-aspiration and less pre-aspiration in fast than in slow speech. Similarly, it has been found for Cuban Spanish that the educated societal class retains /s/ as a sibilant in 22% of all cases in semiformal speech, but only in 3% in fast and informal speech (Terrell, 1980). The weaker perceptual salience of pre-aspiration in comparison with post-aspiration may be an additional bias (Bladon, 1986; Ruch, 2018; Ruch & Harrington, 2014). In the production of post-aspirated, but not pre-aspirated stops, there is a build-up of air-pressure behind the closure which results in an acoustic burst (Fant, 1973) and a rapid modulation of the acoustic signal when the stop is released. The strong perceptual salience of a post-closure release is demonstrated in perception experiments in which listeners' judgements of place of articulation in heterorganic stop consonant clusters C_1C_2 in VC_1C_2V are swayed far more by C_2 than by C_1 (Ohala, 1990). Moreover, in a perceptual experiment by Ruch and Harrington (2014), listeners of Argentinian Spanish who typically produce /sC/ clusters with pre-aspiration and no post-aspiration (Aleza & Utrilla, 2002; Torreira, 2006) were more inclined to perceive *pasta* in a continuum synthesised between singleton *pata* (engl. paw) and *pasta* (engl. pasta) when the continuum was created with a long (27 ms) as opposed to short (13 ms) duration of post-aspiration. Thus, Argentinian Spanish listeners are nevertheless influenced by post-aspiration as a cue to the distinction between post-vocalic /st/ and post-vocalic /t/, even though they produce pre- and not post-aspirated stops and even though (in contrast to Andalusian Spanish) there is no evidence of a sound change in progress in this variety in which pre- are becoming post-aspirated stops.

The synchronic basis for the sound change by which pre- has evolved into post-aspiration as proposed by Parrell (2012) is a resynchronisation of autonomous articulatory gestures that also typically occurs at faster rates of speech (cf. e.g. Beddor, 2009; Davidson, 2006). In this model, depicted in Fig. 1, the oral tract constriction gesture (solid) for the stop closure shifts to be in phase with the glottal opening gesture (dashed) at faster rates of speech. A direct consequence of the shift from anti- to in-phase timing is a decrease in pre-aspiration and an increase in post-aspiration strength: that is,

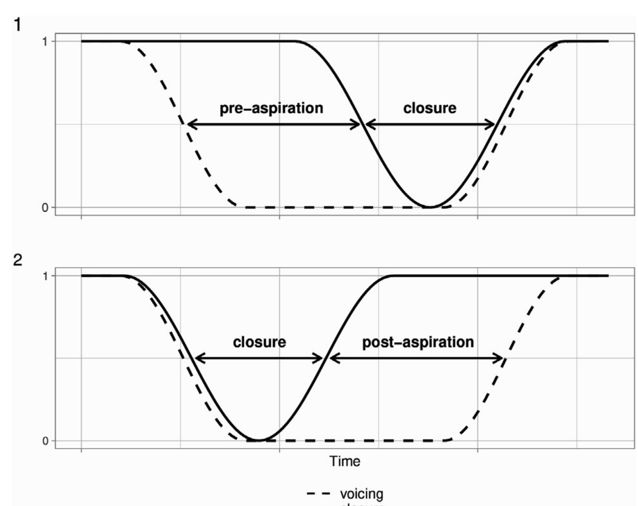

Fig. 1. Idealised scheme of resynchronisation of the closure with the voiceless interval in Andalusian Spanish /s/-aspiration. The solid line is the glottal gesture, where low values stand for an open glottis and hence voicelessness, the dashed line is the oral constriction gesture of the voiceless plosive, where the minimum of the curve indicates maximal closure.

a by-product of this resynchronisation is that the articulatory durations of pre- and post-aspiration stand in an inverse relationship to each other.

It is possible – although so far undemonstrated for aspiration in Andalusian Spanish – that this inverse articulatory relationship forms the basis of a perceptual trading relationship by which listeners parse aspiration from the speech signal but might be agnostic about its temporal location (about whether aspiration occurs before the closure, after the closure, or both; Ruch & Harrington (2014)). There is a potential analogy to a different type of sound change involving the phonologisation of nasalisation studied by Beddor and colleagues in recent years (Beddor, 2009, 2012; Beddor, Coetzee, Styler, McGowan, & Boland, 2018). Their model is informed by at least four related sets of studies from speech perception. Firstly, listeners are sensitive to anticipatory coarticulation in perceiving speech (Alfonso & Baer, 1982; Martin & Bunnell, 1982): listeners perceive a speech sound close to its articulatory onset, i.e. from the time at which there is coarticulatory evidence for the speech sound in the acoustic signal (Fowler, 1984, 2005). Secondly, listeners weight differently the multiple cues to speech sounds that originate from overlapping and coproduced speech gestures (Boersma, Escudero, & Hayes, 2003; Clayards, 2018; Francis, Baldwin, & Nusbaum, 2000; Holt & Lotto, 2006; Idemaru, Holt, & Seltman, 2012). Thirdly, listeners have the capacity to re-weight cues (Boersma et al., 2003; Francis & Nusbaum, 2002; Harmon, Idemaru, & Kapatsinski, 2019; Idemaru & Holt, 2011) as also shown by studies of perceptual learning (McQueen, Norris, & Cutler, 2006; Norris, McQueen, & Cutler, 2003; Reinisch & Holt, 2014, see Samuel & Kraljic, 2009 for a review). Cue reweighting in perception has also recently been shown to carry over to speech production (Lehet & Holt, 2017). Fourthly, the cues can enter into a so-called trading relationship (Beddor, 2009; Best, Morrongiello, & Robson, 1981; Haggard, Summerfield, & Roberts, 1981; Kingston, Diehl, Kirk, & Castleman, 2008; Kirby, 2014; Repp, 1982; Whalen, Abramson, Lisker, & Mody, 1990) in which listeners can pay more attention to a secondary cue if the primary cue to a phonological contrast is compromised. In many types of sound change such as the development of contrastive vowel nasalisation, metaphony (Savoia & Maiden, 1997; Torres-Tamarit, Linke, & Oostendorp, 2016), and tonogenesis (Hagège & Haudricourt, 1978; Hombert, Ohala, & Ewan, 1979), secondary cues that were initially brought about by coarticulation become primary while the primary cues are downweighted and typically completely disappear (thus leading to e.g. the development of present-day German Füße /fysə/ (engl. feet) from old High German /fotiz/ in which there is no trace left of /i/ which initially caused via VCV coarticulation the fronting of the high back vowel in the first syllable; see Kiparsky, 2015; Twaddell, 1938; Penzl, 1949). This stage, in which the secondary cues become primary and primary cues have all but vanished is when the sound change is in the process of phonologisation (Bermúdez-Otero, 2015; Bermúdez-Otero & Trousdale, 2012; Hyman, 2013; Kiparsky, 2015; Ramsammy, 2015). Beddor showed that the precursor to phonologisation is the development of a perceptual trading relationship (Beddor, 2009, 2012). When the cues for a feature trade, then listeners perceive the feature but without necessarily associating it with the coarticulatory source or effect: thus, in American English *send*, they hear nasalisation somewhere in the rhyme without parsing the nasalisation with either of the rhyme's constituents (the vowel or the coda /n/).

The studies by Beddor and colleagues further suggest that the physiological basis of such a trading relationship is an inverse relationship between the coarticulatory source and effect. As far as nasalisation is concerned, this is modelled as the temporal sliding of a nasal gesture of constant articulatory duration into the gesture for the vowel (Beddor, 2009; Beddor, Brasher, & Narayan, 2007): the more the two gestures overlap, the greater the extent of vowel nasalisation and the smaller the degree of articulatory prominence of the post-vocalic nasal.

The purpose of the present study is to determine from an acoustic analysis whether there is any evidence of this type of gestural sliding in the production of /s/-aspiration in Andalusian Spanish that could in turn form the basis of a perceptual trading relationship as proposed by Beddor. There are evidently similarities in the proposed physiological model that underlies both types of sound change. Thus, in both Beddor's model of the production of nasalisation and in that proposed by Parrell (2012) for /s/-aspiration, the consequence of sliding the gesture of constant duration in the direction that eventually leads to the sound change is that, as one of the cues wanes (post-vocalic /n/; pre-aspiration) the other becomes more prominent (coarticulatory vowel nasalisation; post-aspiration).

One of the main aims of the present study is to analyse /s/-aspiration in Andalusian Spanish in order to determine whether there is any evidence for the type of model proposed by Parrell (2012) in which pre- and post-aspiration stand in an inverse relationship to each other as a consequence of the resynchronisation of articulatory gestures of relatively constant duration. Neither studies by Ruch and colleagues (Ruch, 2013; Ruch & Harrington, 2014; Ruch & Peters, 2016) nor Torreira (2007) have found much evidence in support of such a relationship. Moreover, Ruch and Harrington (2014) showed that the closure duration, far from being stable, was influenced by the age and regional origin of the speakers. On the other hand, none of these studies nor indeed Parrell's (2012) are necessarily appropriate tests of the model in Fig. 1 because they are based on vertical segmentations of the speech signal (as defined in van der Kooij & van der Hulst, 2005: p. 167) into pre- and post-aspiration and a closure instead of modelling more directly how glottal and supraglottal gestures are aligned and potentially resynchronised. Segmentations are often unreliable and arbitrary given the dynamic nature of speech (Fowler & Smith, 1986), so it might be the case that a trading relationship can only be identified by considering a representation of speech that is a closer representation of the model in Fig. 1. Moreover, the type of trading relationship that has been identified in perception in connection with sound changes in progress might come about only after a listener has experienced talkers that differ in their use of the acoustic cues that are traded in perception. For example, a listener might interpret perceptual equivalence between V□ and VN only after having been exposed to many talkers that differ in the extent to which they nasalise a vowel in a VN context. If this is so, then a trading relationship might be related to variations regulated by dynamic relations between two cues that characterise an entire population of speakers.

The present study addresses these issues by computing two time-varying signals from acoustic data of Andalusian Spanish, one being a proxy for the glottal gesture, the other a proxy for the oral constriction gesture. According to articulatory phonology (Browman & Goldstein, 1986, 1992), the alignment of these two gestures with respect to each other is responsible for the presence or absence of aspiration around the voiceless plosive C in sequences like V_1sCV_2. A central concern in Section 2 will be to test whether there is any evidence across the population of speakers that pre- and post-aspiration are negatively related through the phasing of the oral closure. The aim in Section 3 is to investigate whether age and regional origin of the speakers condition the timing of the oral closure with reference to the voiceless interval. If so, this would provide a link between the synchronic model in Fig. 1 and a sound change in progress by which pre-aspiration wanes and post-aspiration increases in Andalusian Spanish /sC/ clusters. In the last part of this paper, we will discuss how the dynamic analysis of speech can be beneficial to the study of sound change and outline a new cognitive-computational model of sound change which is based both on principles from articulatory phonology and episodic models of speech.

2. Relation between pre- and post-aspiration

There were two main issues of concern here. The first was to build a general model of the synchronisation of the closure with the glottal opening (Fig. 1) based on the entire database of V_1sCV_2 sequences across speakers of different age groups and regions. Functional Principal Component Analysis (FPCA) was used for this purpose since it can model two (or more) time-varying signals simultaneously (Gubian, Harrington, Stevens, Schiel, & Warren, 2019; Gubian, Torreira, & Boves, 2015). The second was to test in this general model the extent to which pre- and post-aspiration are predictable from closure phasing.

2.1. Method

2.1.1. Speakers and materials

The data analysed in this study was taken from a larger speech database collected by the fourth author of this paper (Ruch, 2013). Parts of this database have been analysed elsewhere (Ruch, 2013; Ruch & Harrington, 2014; Ruch & Peters, 2016). The focus of the present study was on 48 speakers of Andalusian Spanish who produced words containing /sC/ clusters in a V_1sCV_2 context (C = /p, t, k/). The speakers were equally divided between two age groups (younger: 20–36 years; older 55–79 years) and two regions (East and West): thus, there were 12 speakers for each of the four possible age group × region combinations. The clusters occurred in the words listed in Appendix A. The words in the available corpus had been constructed as far as possible to include /s/-aspiration clusters in several mostly high frequency real words for three places of articulation /sp, st, sk/ combined with one of the three vowel types /i, u, a/. The majority (just over 80%) of words had a paroxytonic lexical stress pattern with primary lexical stress on the penultimate syllable (e.g. *espía* /esˈpia/). Other stress patterns (e.g. *pasta* /ˈpasta/) or non-words (e.g. *bestiando*) were used when there were an insufficient number of real words for these place × vowel combinations.

The productions were elicited using a prompt in which each word appeared individually or in a short sentence on a computer monitor (see Ruch & Harrington, 2014 for further details of the recording procedure). There were up to three repetitions per speaker per word, thus giving a potential maximum of 48 (speakers) × 52 (words) × 3 (repetitions) = 7488 word tokens. However, not all speakers produced three repetitions. Tokens which included errors of production (in particular the production of the wrong target word or of a false start) as well as any productions of target words with standard Spanish /s/ in the /sC/ cluster were removed from further consideration following the procedure in Ruch and Harrington (2014). This left 6393 word tokens. Furthermore, 446 productions in which the voicing probability peak in V_1 and/or V_2 was too low to indicate voicing were also discarded since it was then not possible to identify reliably the existence of pre- or post-aspiration. Another 82 tokens were excluded because the PEFAC algorithm (see Section 2.1.2) erroneously computed the voicing probability peaks to occur during the closure, not during the neighbouring vowels. A further 18 tokens were excluded from the analysis because the target words were not fully recorded, i.e. the recording was clipped at the end. Two tokens had to be excluded because of technical issues during the automatic speech processing. The final count of the analysed word tokens was 5845, distributed across 48 speakers and 52 word types that included 20, 17, 15 word types containing /sp, st, sk/ clusters respectively.

2.1.2. Acoustic parameters

The synchronisation of the closure with the glottal opening was estimated acoustically by means of two separate parametrisations of the acoustic signal. The first of these, designed to model the glottal opening, was the voicing probability (henceforth VP) that was computed by applying the PEFAC algorithm (Gonzalez & Brookes, 2014) to the original audio files with a 5 ms frame shift and otherwise default settings. The second, which was used to model the supralaryngeal closure, was derived from the high frequency energy (henceforth HF) in the speech signal. This was obtained by double-differencing the audio signal to give 12 dB boost per octave (i.e. pre-emphasis) resulting in sharper transitions between fricated and closure phases of the signal (Harrington & Cassidy, 1999). These double-differenced signals were then high-pass filtered at 3 kHz, from which the logarithmic root mean square energy was computed with a window length of 20 ms and a frame shift of 5 ms. The resulting signal was smoothed with a 20 Hz Butterworth low-pass filter (Butterworth, 1930), then normalised such that 0 dB was set to the minimum of this signal during the supralaryngeal closure (point *M* in Fig. 2). An example of the two resulting signals, HF and VP, for one utterance of the word *estado* (engl. state) by an older West Andalusian speaker can be seen in panels 2 and 3 of Fig. 2, respectively. The interval that was processed within

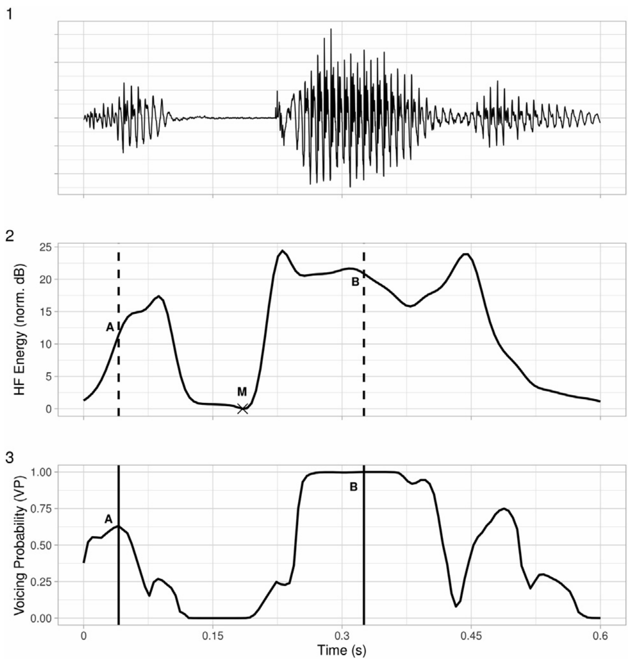

Fig. 2. From upper to lower panels: Waveform, HF (amplitude-normalised high frequency energy), and VP (voicing probability) signals over time for a production of *estado* (engl. state) by an older West Andalusian speaker. *A* & *B* are the voicing probability maxima in V_1 and V_2, respectively. *M* is the point of lowest amplitude during the closure. The interval that was analysed extended between *A* and *B*.

V_1sCV_2 was defined as extending between points *A* and *B* in Fig. 1. Point *A* is the time at which the voicing probability first attains its maximum value in V_1 working backwards in time from the time of maximum closure, *M*. Point *B* is the time at which the voicing probability first attains its maximum value in V_2 working forwards in time from *M*. The other parts of the signal from the acoustic onset of V_1 to *A* and from *B* to the acoustic offset of V_2 were excluded from the analysis.

2.1.3. Data analysis

Functional Principal Component Analysis (FPCA) (Gubian, Torreira, & Boves, 2015; Ramsay & Silverman, 2010) was used in order to find the main dimensions of variation in the $N = 5845$ pairs of $HF_i(t)$ and $VP_i(t)$ curves, $i = 1, \ldots, N$. There were three pre-processing steps prior to applying FPCA. Firstly, in order to obtain a functional representation from the time-sampled curves, standard smooth interpolation techniques were applied using B-splines as function basis (see Gubian, Torreira, & Boves, 2015; Ramsay & Silverman, 2010 for details). Secondly, the signals were linearly time-normalised between the times of the voicing probability maxima (time points *A* and *B* in Fig. 2; see Appendix E for details of the effect of time normalisation). Thirdly, the HF signals were also amplitude re-scaled by dividing each curve by the 75% quantile of all HF values. This was done to ensure that HF and VP signals both spanned approximately the same range between 0 and 1.[1] As a consequence, HF and VP were a closer representation of the model in Fig. 1. This re-scaling also prevented the principal components from being unduly influenced by large amplitude values of one of the signals with respect to the other.

The FPCA parametrisation was expressed by the following pair of equations applied between time-normalised values 0 and 1 (e.g. points *A* and *B* in Fig. 2):

$$HF_i(t) \approx \mu_{HF}(t) + \sum_{k=1}^{K} s_{k,i} \cdot PCk_{HF}(t) \tag{1a}$$

$$VP_i(t) \approx \mu_{VP}(t) + \sum_{k=1}^{K} s_{k,i} \cdot PCk_{VP}(t) \tag{1b}$$

where $\mu_{HF}(t)$ and $\mu_{VP}(t)$ are the mean signals, e.g. $\mu_{HF}(t) = \frac{1}{N}\sum_i HF_i(t)$, $PCk_{VP}(t)$ and $PCk_{HF}(t)$ are *K* pairs of Principal Components (PCs), $k = 1, \ldots, K$, which are based on the entire data set, and $s_{k,i}$ are weights or *scores*, which modulate each PC*k* differently for each signal pair $(HF_i(t), VP_i(t))$. Formally, Eq. (1) follow the same structure of ordinary PCA in which any input signal is approximately decomposed into a linear combination of *K* PCs added to the mean. The main difference in comparison with PCA is that in FPCA the input, mean, and PCs are functions of time as opposed to vectors of real numbers. Crucially, the linear combination expressed by PC scores modulates the PCs for both dimensions *together*, i.e. $s_{1,i}, s_{2,i}, \ldots, s_{k,i}$ are the same in Eqs. (1a) and (1b), which is essential for capturing systematic co-variations across dimensions. Using the R package `fda`[2], we computed the first $K = 3$ PCs for the set of 5845 smoothed curve pairs of $HF(t)$ and $VP(t)$ which explained 31.5%, 24.3%, and 14.5% of the variance in the curve shapes respectively (70.3% combined). Only the first PC is considered in the remainder of this paper since the second and third added very little to explaining phonetic variation in the data that was relevant to the sound change at hand (see Appendix D for more details on PC2 and PC3).

The resulting PCs and PC scores can be used to reconstruct individual pairs of signals. The reconstructed curves tend towards an ever closer approximation to the original signals as the number of PCs that are used in the reconstruction is increased (i.e. for increasing values of *K* in Eq. (1); see Fig. 4 in Gubian et al., 2015 for an example). This operation is demonstrated in Figure 3 for one clearly post-aspirated and one clearly pre-aspirated instance of the word *despide* (engl. he/she/it fires) produced by two different speakers. The normalised time points 0 and 1 correspond to the beginning and end respectively of the sequence of interest V_1sCV_2 (cf. points *A* and *B* in Fig. 2). Row 2 in this plot shows HF (solid) and VP (dashed) which were obtained as described in Section 2.1.2. Row 3 instead shows the HF and VP signals which were reconstructed using only PC1 and the corresponding PC score s_1 (0.46 for the post-aspirated token, -0.48 for the pre-aspirated token) in Eq. (1). It is clear from Fig. 3 that the reconstructed curves in row 3 are similar in shape, but smoother versions of the raw HF and VP signals in row 2 that were derived directly from the speech signal. Rows 2 and 3 of the left column of Fig. 3 show the in-phase timing of the articulatory gestures represented by HF and VP, in which the closure of the plosive and the start of the voiceless interval are approximately synchronous, resulting in post-aspiration. By contrast, HF and

[1] The empirical choice of the 75% quantile as a normalising factor was justified by the observation that the distribution of HF values has a long right tail; thus an exact normalisation based on the maximum would have compressed the HF signals to the extent that the VP signal would dominate (VP > HF) most of the time.

[2] Version 2.4.0 was used here. More recent versions (5.1.5.1 being the current one at the time of writing) can be used as well, provided that PC scores are summed across dimensions.

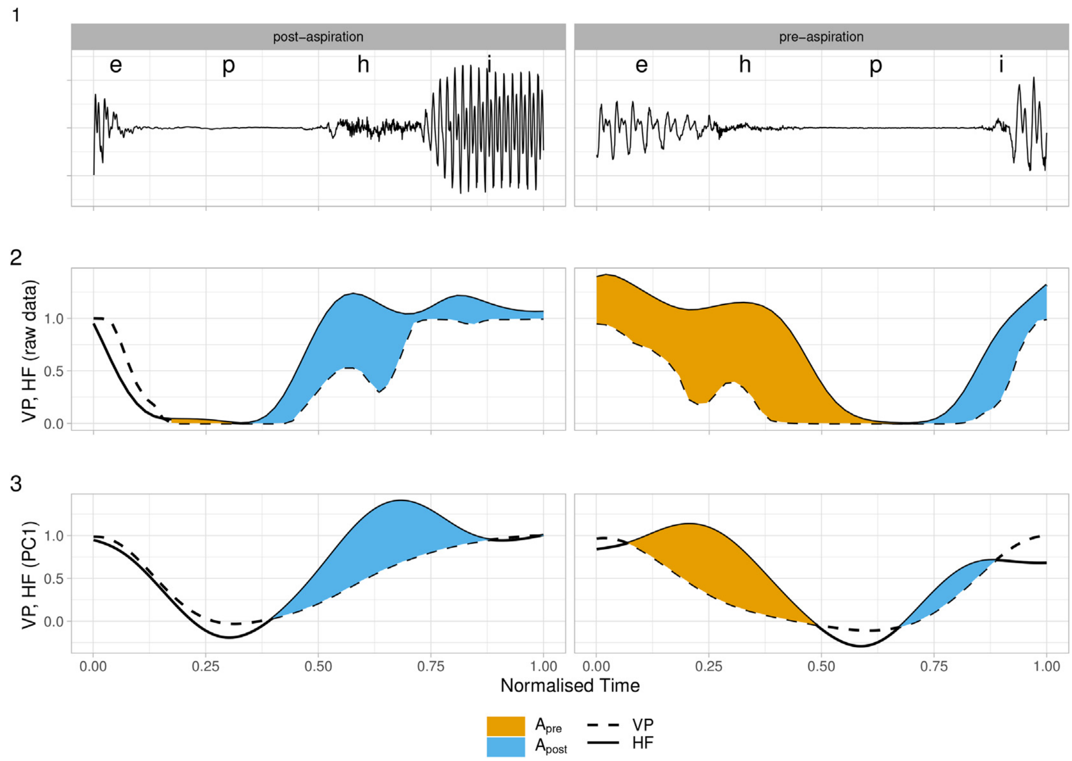

Fig. 3. Example of a post-aspirated and a pre-aspirated token extending between points *A* and *B* (see Fig. 2) in V_1sCV_2 in the word *despide* (engl. he/she/it fires). The post-aspirated token (left column) was produced by a young West Andalusian speaker, the pre-aspirated token (right column) was produced by an older East Andalusian speaker. The first row shows the waveforms, the second row shows HF (solid) and VP (dashed) calculated on the raw speech signals as described in Section 2.1.2, the third row shows HF and VP as reconstructed based solely on PC1 using Eq. (1). The yellow areas are A_{pre} and the blue areas are A_{post} as in Eq. (2).

VP in the right column of Fig. 3 show an anti-phase timing in which the closure is delayed relative to the start of the voiceless interval, resulting in pre-aspiration. Recall that we expect there to be a closure when both HF and VP show very low values, i.e. when there is very little energy as well as hardly any or no voicing. Overall, the HF and VP signals matched the waveforms very well. While there can be some imprecisions due to erroneous VP values or residual noise in the HF signal, a visual inspection of several signal pairs showed that the curves represented quite consistently and accurately the pre- and post-aspiration phrases.

2.1.4. Estimating aspiration

In order to quantify the extent of aspiration before and after the closure, the area A_{tot} enclosed between the two curves HF and VP for HF > VP was measured. The reason why A_{tot} is an appropriate measure for aspiration is that aspiration occurs whenever VP as the proxy for the glottal gesture is low (i.e. no voicing) and HF as the proxy for the oral constriction gesture is high (i.e. no closure), as suggested by Fig. 1. The influence of pre- and post-aspiration that precede and follow the closure was respectively estimated from the areas A_{pre} and A_{post} defined in the same way as A_{tot} but spanning the normalised time intervals up to and beyond the minimum of HF, respectively (hence $A_{tot} = A_{pre} + A_{post}$). Formally:

$$A_{pre} = \int_{(HF>VP)\cap(0\leqslant t\leqslant t_M)} (HF(t) - VP(t))\, dt \tag{2a}$$

$$A_{post} = \int_{(HF>VP)\cap(t_M\leqslant t\leqslant 1)} (HF(t) - VP(t))\, dt \tag{2b}$$

where $t_M = \text{argmin}_t HF(t)$ (cf. point *M* in Fig. 2, panel 2). The constraint HF > VP was introduced in order to exclude negative areas which, in contrast to the positive areas which we take as an estimate of aspiration strength, would not have any meaning in this context.

This method for inferring aspiration strength from area calculations is entirely independent of the functional data analysis, i.e. Eq. (2) does not specify how HF and VP were obtained. Examples of the area calculation both on raw (row 2) and reconstructed data (row 3) can be seen in Fig. 3. That is, wherever HF > VP, A_{pre} (yellow) was computed as the area between the curves from normalised time point 0 to the point of maximal closure, and A_{post} (blue) was computed as the area between the curves from the point of maximal closure to time point 1. Fig. 3 shows that A_{post} is large while A_{pre} is very small for the post-aspirated token on the left, whereas the opposite is true for the pre-aspirated token on the right. This was so both when HF and VP were obtained from the raw speech signals and when they were reconstructed using only PC1. It can also be seen that the areas partially extend into the vocalic parts of the segments. This is because aspiration (and in particular pre-aspiration) could often overlap with the vowel in a breathy voice production. This type of temporal overlap can be expressed by the area measurements proposed here, but is far more difficult to represent using durations extracted from vertical segmentations of the acoustic speech signals into the vowel and aspiration.[3]

[3] Appendix F shows how the methods which were chosen for data analysis and the quantification of aspiration compare to various other methods, including the use of more conventional segmental durations.

2.2. Results

We consider firstly the variation in PC1 derived from an applicaton of FPCA to the whole data set; and secondly an analysis of the how pre- and post-aspiration are related based on area calculations.

Fig. 4 shows the relationship between quantitative changes in the first PC score and qualitative changes in the two signals HF and VP. The middle panel contains the mean signals $\mu_{HF}(t)$ (solid) and $\mu_{VP}(t)$ (dashed) across all input signals. These mean curves change, however, when s_1 is set to positive (right column) or negative values (left column), and all other scores to zero (recall Eq. (1)). Thus, the panels from left to right show how the shapes of the VP and HF signals are modified as the first PC score changes from negative to positive values. We chose these representative PC score values to be $\pm\sigma_{s_1}$, i.e. we added to, or subtracted from, each mean curve only the PC1 curve multiplied by 0.28, the standard deviation of the first PC score s_1.

PC1 was closely related to the dynamic changes predicted by the model in Fig. 1. A comparison of the left, central, and right panels of Fig. 4 shows that PC1 modelled a phase shift relative to the VP signal of an HF-minimum that corresponds to the instant when the maximal constriction of the vocal tract is attained. More specifically, variations from negative to positive s_1 resulted in a shift of the signal associated with the closure from late to early. Note, however, that it is not a necessary consequence of FPCA that the first PC describes a phase shift, nor that it coincides with the shape variation that is most relevant for the analysis it was employed for, as no prior information on the relevance of a phase shift was introduced as input to FPCA. The first PC simply explains (by definition) the largest amount of variance in any given data set. In the case reported here, it is coincidental that the most relevant kind of variation was captured by PC1 (the reader is referred to Appendix D for an analysis of PC2 and PC3).

A test was then made of whether the closure phasing expressed by PC1 conditioned the extent of aspiration before or after the closure. To this end, HF and VP were derived from Eq. (1) using only PC1 (cf. Fig. 4, or Fig. 3, row 3) and the areas A_{pre}, A_{post}, and $A_{tot} = A_{pre} + A_{post}$ were computed using Eq. (2). Fig. 5 shows these areas as a function of the first PC score s_1. The steep and opposite trends of the lines for $A_{pre}(s_1)$ (yellow) and $A_{post}(s_1)$ (blue) indicate that pre-aspiration and post-aspiration were indeed related through the phase shift of the closure expressed by changes in s_1. More specifically, when s_1 was negative (e.g. left panel in Fig. 4), the closure in the HF signal was in an anti-phase relationship with the opening of the glottis, thus leaving more time for pre-aspiration (large A_{pre}) and less for post-aspiration (small A_{post}) to occur, and vice versa for positive s_1 (see e.g. the right

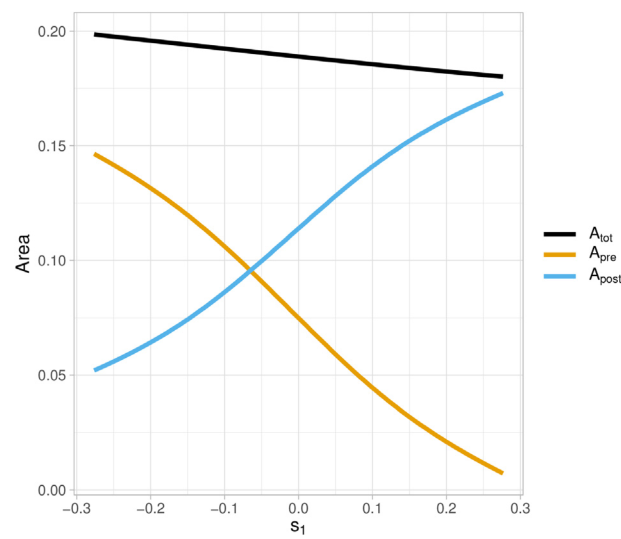

Fig. 5. A_{pre} (yellow), A_{post} (blue), and $A_{tot} = A_{pre} + A_{post}$ (black) computed by using Eq. (2) as a function of s_1, when the signals HF and VP are defined as in Eq. (1) using only PC1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

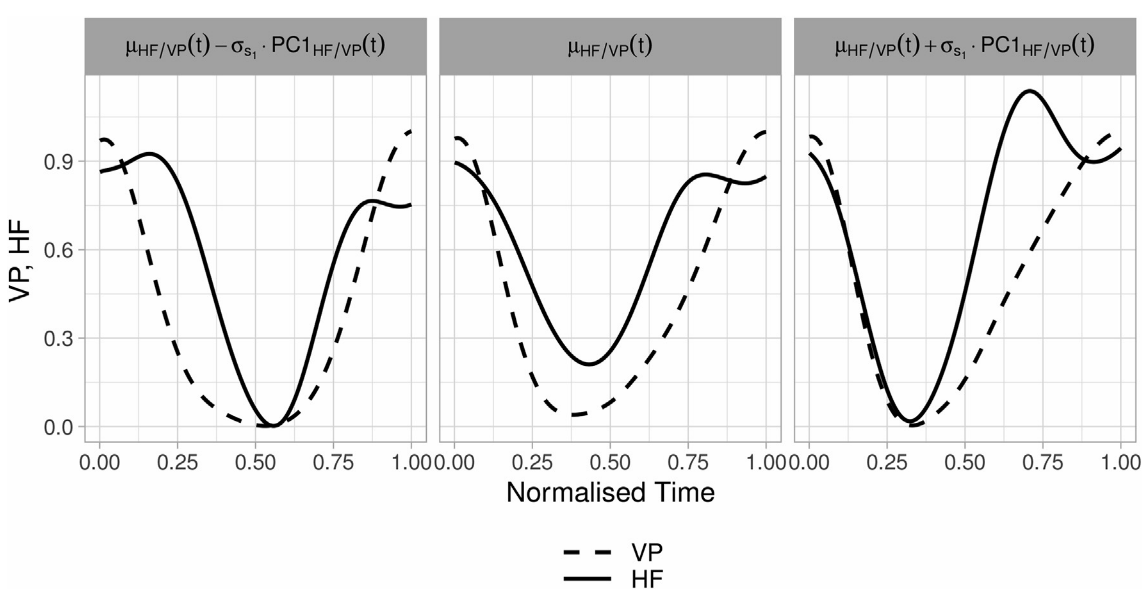

Fig. 4. Variation expressed by PC1. The middle panel shows the mean curves $\mu_{HF}(t)$ and $\mu_{VP}(t)$, which were modified by adding to (right panel) or subtracting from (left panel) each mean curve the PC1 curve multiplied by the standard deviation of s_1. The exact formulae are given in the panel headings.

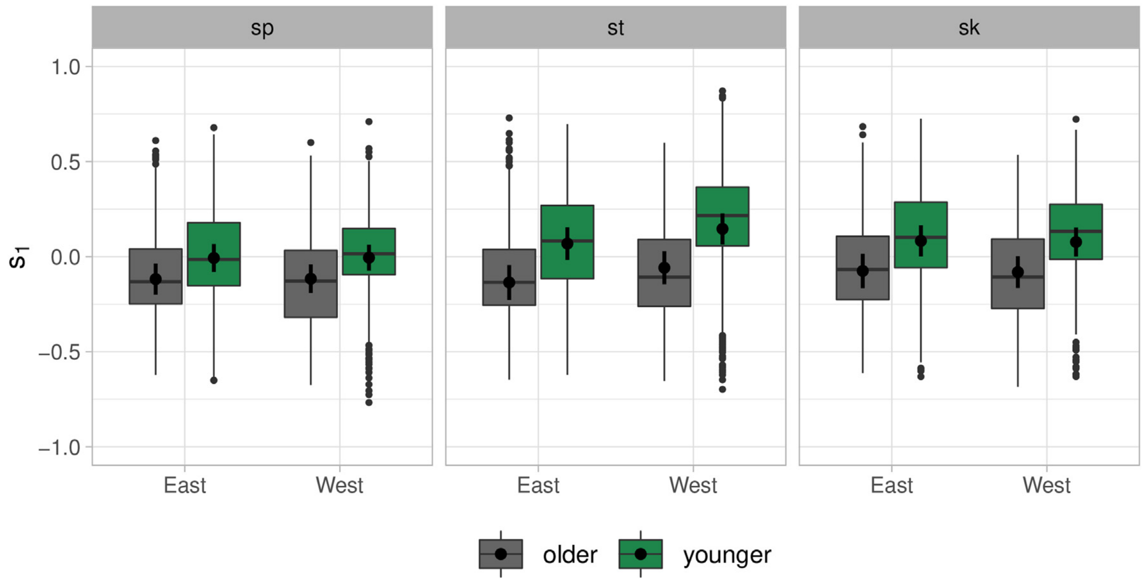

Fig. 6. Boxplots of s_1 values as well as estimated marginal means of s_1 (black dots within the boxes) with related confidence intervals (black vertical bands around the dots) based on Eq. (3). Younger speakers are shown in green, older ones in dark grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

panel in Fig. 4 in which a closure is represented by low values in both HF and VP). The overall area $A_{tot}(s_1)$, i.e. the total amount of aspiration, remained stable across all values of s_1. Appendix B provides proof that the steepness of the lines $A_{pre}(s_1)$ and $A_{post}(s_1)$ is high, while $A_{tot}(s_1)$ does not significantly depart from a flat line.

2.3. Discussion

The aim of this section was to identify whether there was a relation between pre- and post-aspiration strength, following the model in Fig. 1. When FPCA was applied to HF and VP that had been extracted from the speech signals and that represent the oral constriction and glottal gesture respectively, it was found that the most important dimension for explaining the variance in these Andalusian data was the relative alignment of the closure which was captured by PC1 and modulated by score s_1. Moreover, a reconstruction of HF and VP using only PC1 in Eq. (1) (thereby eliminating all other sources of variation) showed a clear relationship between closure timing, pre-, and post-aspiration. Thus, later closures were associated with more extensive pre-aspiration and less extensive post-aspiration while for earlier closures pre-aspiration diminished and post-aspiration increased.

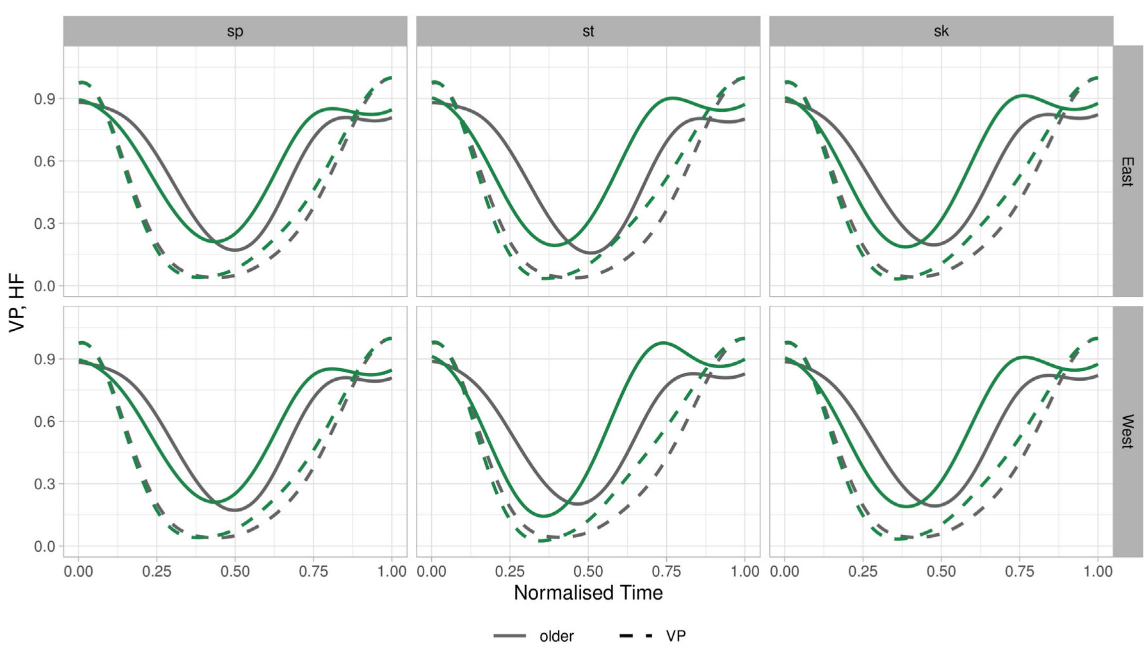

Fig. 7. Reconstruction of $HF(t)$ (solid) and $VP(t)$ (dashed) using Eq. (1). For each factor combination, EMMs (estimated marginal means) are used for s_1, while $s_k = 0$ for $k > 1$. The curves for the younger age group are green, the ones for the older group are dark grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

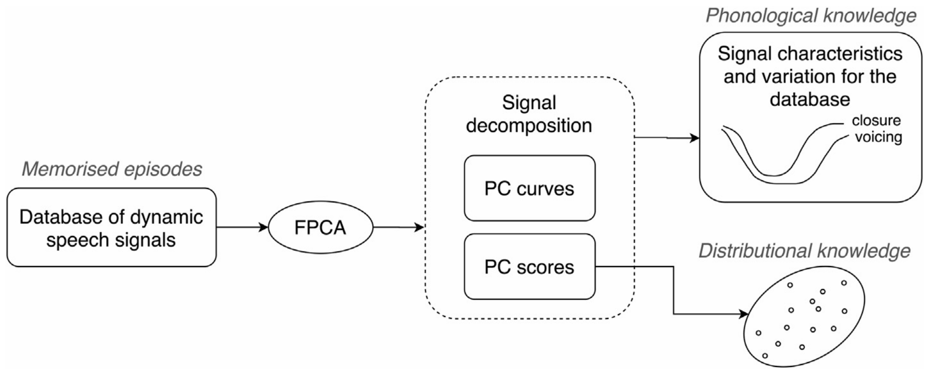

Fig. 8. A schematic outline of the outputs after applying FPCA to the database in this study and their potential association (in italics) to components of a cognitive-computational model of sound change. The PC-based signal decomposition model gives rise to both summary signal characteristics over the entire database as well as distributional information.

The results of this first part therefore provide some support for the model in Fig. 1 based on a gestural model of speech production in which the extent of pre- and post-aspiration are inversely related to each other as a consequence of how the closure is timed relative to the glottal opening. A major difference between the acoustic model presented here and the schematic outline based on articulatory gestures is that the former but not the latter takes amplitude into account. Thus, the model in Fig. 1 is based only on timing considerations and not on the size of vocal tract opening during the intervals when the vocal tract is predominantly given over to aspiration. On the other hand, in the present study the areas between the acoustic signals of high frequency energy and voicing probability that were taken as a proxy for the extent of aspiration were influenced not only by the phasing of the closure, but also by the amplitude of energy in the HF signal. That is, Fig. 4 has shown that PC1 did not only capture a phase shift of HF relative to VP, but also that post-aspiration had a higher energy peak than pre-aspiration. This functional analysis thereby presents more information on aspiration than would be possible using duration measurements based on a segmentation of the speech signals. However, in order for the amplitude (and hence areas) to be comparable across tokens, global amplitude differences which were most likely caused by speaker-dependent variations in the amplitude of the signal (caused e.g. by speaking more quietly or softly and/or as a result of different distances from the microphone) had to be factored out (see Appendices C and D for further details).

3. Influence of speakers' age and region on closure phasing

The analysis suggesting that pre- and post-aspiration are predictable consequences of resynchronising the closure with the voiceless interval has been based so far on a model applied to the entire data set across all speakers and repetitions. The concern of this section is to test whether s_1, which was shown to capture variations in closure timing (Section 2), also distinguishes between age and regional origin of the speakers. The prediction is that it should do so, given that older and East Andalusian speakers have been shown to have greater pre-aspiration and less post-aspiration than their younger and West Andalusian counterparts (Ruch & Harrington, 2014; Ruch & Peters, 2016). If the prediction holds, then this would establish a link between synchronic variation and a sound change in progress.

3.1. Method

A linear mixed effect regression model was constructed for the first PC score s_1 as response variable and age (two levels), region (two levels), and cluster type (three levels) as fixed factors, while word (52 levels) and speaker (48 levels) were added as random factors. All statistical results reported below for s_1 also approximately hold for A_{post} (and for A_{pre} with negative sign), since the areas are near-linearly related to the first PC score (cf. Fig. 10 and see Appendix B for further mathematical details). While the areas were helpful in demonstrating the existence of a relationship between pre- and post-aspiration in Andalusian Spanish, s_1 was chosen as response variable here because it directly expresses the closure phasing which, according to the model in Fig. 1, is responsible for this relationship.

The full LMER model is given in Eq. (3) below (R notation):

$$s_1 \sim (age + region + cluster)^3 + (cluster|speaker) + (age + region|word) \tag{3}$$

where the term $(age + region + cluster)^3$ indicates the presence of three fixed factors plus all the possible two- and three-way interaction terms formed by them, while the random factor *speaker* is modulated by *cluster* and *word* is modulated by *age* and *region*. The model was pruned using the R package `lmerTest` (version 3.1.2) in order to remove all non-significant factors and factor combinations. After pruning, all fixed and random terms were retained apart from the three-way interaction of the fixed factors and the two-way interaction between age and region. All post hoc tests were computed using the R package `emmeans` (version 1.4.6).

In order to translate the predicted s_1 values back into HF and VP curves, the resulting estimated marginal means (EMMs) were substituted into Eq. (1), where only s_1 was used and the other scores were set to zero. For instance, $HF(t)$ for /st/ produced by young West Andalusians was represented by $\mu_{HF} + s_1 \cdot PC1_{HF}(t)$, where s_1 took the EMM value for that particular factor combination.

3.2. Results

The boxplots in Fig. 6 provide a graphical impression of how age, region, and cluster type affect the first PC score. Compatibly with the production of an earlier closure accompanied by post-aspiration, s_1 was higher in Fig. 6 for younger speakers

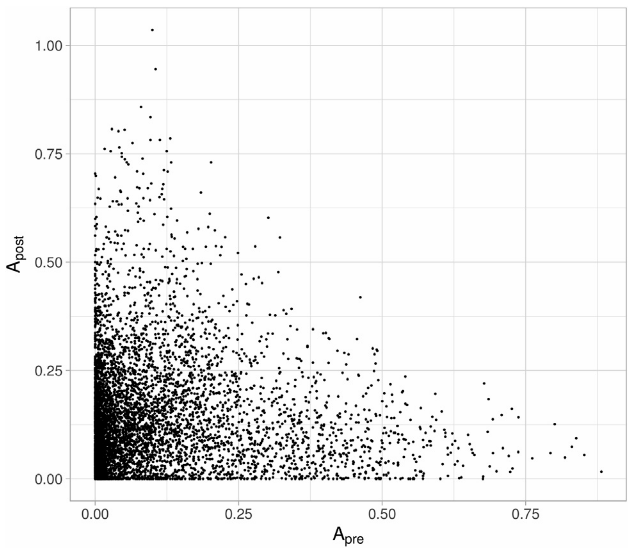

Fig. 9. Area A_{post} against A_{pre} as in Eq. (2) when $HF(t)$ and $VP(t)$ are the (smoothed and time-normalised) curves obtained directly from the speech signal without FPCA transformation.

in all conditions. Fig. 6 suggests only few region differences except for alveolars. The results of the mixed model with the fixed and random factors given in Section 3.1 showed a significant main influence on s_1 of age (F[1, 46.9] = 24.6, $p < 0.001$) but not of region nor of cluster type. There was however a significant two-way interaction between age and cluster (F[2, 54.4] = 6.6, $p < 0.01$) as well as region and cluster (F[2, 53.3] = 5.7, $p < 0.01$). Post-hoc alpha-adjusted Tukey tests showed that there were significant differences between older and younger speakers for all three places of articulation (/sp/: $t = 3.8$, $p < 0.001$; /st/: $t = 5.0$, $p < 0.001$; /sk/: $t = 4.6$, $p < 0.001$). The post hoc tests revealed no significant region differences. However, there was a trend towards a significant difference between speakers from East and West Andalusia producing the alveolar cluster ($t = 1.9$, $p = 0.06$). For young West Andalusian speakers there was a significant difference between labial and alveolar clusters ($t = 3.9$, $p < 0.001$).

Fig. 6 also shows the estimated marginal means (EMMs) for each combination of age group × region × cluster type (black dots within the boxes), with their respective confidence intervals (vertical bands), which are the values of s_1 that the model predicts for each combination of the fixed factors. As described in Section 3.1, these EMM values for specific factor combinations were used to reconstruct the HF and VP curves. This reconstruction (Fig. 7) shows that the closure (represented by the HF signal) occurred earlier for younger (green) than for older (dark grey) speakers for all three places of articulation and both regions. These differences between the age groups are the consequence of (i) $s_1(\text{young}) > s_1(\text{old})$ as found by the LMER model (cf. Fig. 6) and (ii) s_1 mainly modulating a phase shift of HF, where the shift is towards the left of the time axis for increasing s_1 (cf. Fig. 4). Additionally, the HF signals for /st/ and young speakers differed from each other between East and West, the latter being more left shifted and reaching higher energy values in the second part of the signal. Although this difference was not found to be significant, it is clearly visible in the reconstructed curves.

3.3. Discussion

There were age- and (to a lesser extent) region-dependent variations in the PC score s_1 derived from FPCA. These differences between the speaker groups together with the evidence that PC1 models a phasing of the closure suggest that, consistently with various other studies (Moya Corral, 2007; O'Neill, 2010; Ruch, 2013; Torreira, 2007; Torreira, 2006), there is a sound change in progress in /sC/ clusters in Andalusian Spanish. Based on the analysis and conclusions in Section 2, the direction of change is such that closure is timed to occur earlier in younger than in older speakers as a consequence of which younger speakers produce these clusters with more post-aspiration and less pre-aspiration than their younger counterparts.

The East Andalusian variety is more conservative as far as this sound change is concerned than its Western counterpart, as others (O'Neill, 2010) have shown. A comparison between the two regions can therefore provide some clues about how sound change is affected by phonetic context. The analysis conducted in this section suggests that alveolar contexts might lead the sound change while labial clusters seem to be the last ones to be affected by the sound change. This is because, firstly, there was a trend towards a significant difference between East and West Andalusian speakers only for /st/ and secondly, there was a significant difference between /sp/ and /st/ for the most advanced speaker group, namely young West Andalusian speakers. This finding is compatible with the suggestion in Ruch & Peters, (2016) that the sound change by which pre- evolves into post-aspiration first originates in the alveolar context before spreading to the velar and, lastly, to the labial context.

4. General discussion

The first part of this study showed that pre- and post-aspiration strength in Andalusian Spanish are inversely related to each other and a predictable consequence of how the closure is phased with respect to the voiceless interval in V_1sCV_2 sequences. This finding is consistent with the model based on articulatory phonology proposed by Parrell (2012) in which closure rephasing is at the core of the change from pre- to post-aspiration at a faster speech rate in the production of /st/ in *pastándola*. It also extends this model by suggesting that the inverse relationship between pre- and post-aspiration as a consequence of closure re-phasing depends not just on timing but also on scaling, i.e. on the relationship between the amplitude of aspiration noise before and after the closure.

The second part of the study showed that the closure phasing and hence the inverse relationship between pre- and post-aspiration are conditioned by age and to a lesser extent by region. As far as age is concerned, these results are compatible with evidence showing a sound change in progress in Andalusian Spanish (O'Neill, 2010; Ruch & Harrington, 2014; Ruch & Peters, 2016) given that the closure was found to be timed earlier for younger than older speakers. The change in progress was more advanced for speakers from West than from East Andalusia when producing the alveolar cluster. The new approach in this study is that these age, region, and place of articulation differences have been established based on analyses of pairs of time-varying signals requiring no segmentation of the closure and aspiration intervals. This approach has a methodological advantage given that these intervals (in particular V_1 and pre-aspiration) overlap with each other, thereby often making vertical segmentation unreliable and inconsistent (Fowler & Smith, 1986). Speech is inherently dynamic involving the overlapping of autonomous articulatory gestures (Fowler & Saltzman, 1993) and regular sound change almost always arises out of dynamic processes such as coarticulation (Ohala, 1993) and undershoot (Lindblom, Guion, Hura, Moon, & Willerman, 1995). The method proposed in this study based on FPCA is appropriate for modelling these dynamic aspects of speech and the sound changes that they give rise to precisely because it provides a way of categorising speech signals and of quantifying change without having to enforce an often arbitrary vertical segmentation of the speech signals (Fowler, 1984; Fowler & Smith, 1986). Although FPCA does not produce a dynamic model in the strict sense (e.g. a set of differential equations), it provides a way of isolating the variation in the pair of signals $HF(t)$ and $VP(t)$ (as shown by PC1) corresponding to the gestural dynamics sketched in Fig. 1. In other words, the evidence that the main variation in those signals is due to a phase shift emerges from the statistical FPCA of the signal shapes.

Fig. 8 is a summary of the dynamic approach to analysing the Andalusian database and its potential association to components of a cognitive-computational model of sound change (Ettlinger, 2007; Stevens, Harrington, & Schiel, 2019; Todd, Pierrehumbert, & Hay, 2019; Wedel, 2006). In this study, FPCA (Section 2.1.3) was applied to a database of dynamic episodes of speech derived from acoustic speech signals in order to obtain a signal decomposition model that consists of PC curves and PC scores. There are two different kinds of information that follow from the decomposition: First, the model shows how the time-varying signals that represent articulatory gestures systematically vary in shape and phasing. Second, the decomposition assigns every dynamic speech episode i its specific PC scores $s_{1,i}, s_{2,i}, \ldots, s_{k,i}$. All PC scores of all tokens in the database form a distribution in an abstract, multidimensional space. The location of the scores in this space can depend on speaker-specific properties like their age or regional origin, as shown in the second part of the study (Section 3).

The cognitive-computational architecture of speech processing that is proposed in this model adopts the idea from exemplar models that phonological categories stand in a stochastic relationship to remembered speech signals (Pierrehumbert, 2003; Pierrehumbert, 2006; Johnson, 2006; Johnson, 1997). The central idea here is that in human speech processing, individuals derive both phonological and distributional information after applying a transformation such as FPCA to multidimensional, remembered, time-varying episodes of speech. The derived phonological knowledge is analogous to separate tiers in articulatory phonology containing independent gestural dynamics showing (as in Fig. 4) how the shape and phasing of gestures that characterise the phonological category vary across the population of speakers. The distributional information is a cloud of points derived from the remembered episodes for the same phonological category (cf. Fig. 6).

The transformation of memorised episodes of speech (e.g. FPCA in Fig. 8) can result in a large amount of dimensions along which signals of the same phonological category can vary (e.g. PCs). The further issue to be considered is how individuals learn which kinds of variation are most relevant for a given phonological category. We suggest this might be guided by both phonetic and phonological criteria. The *phonetic* criterion is that a dimension of variation represents how the tokens of the phonological category are actually produced. In the case of the Andalusian database presented here, the most relevant variation for the /sC/ category was how the opening/closure phase of the vocal tract was variably timed with respect to a voiceless interval (i.e. PC1). The *phonological* criterion for identifying a dimension of variation as relevant is that it represents a group-level characteristic of the phonological category. That is, for the Andalusian data, the level of aspiration in episodes of /sC/ must be high regardless of how the closure is phased relative to the voiceless interval. This is evident in the black line in Fig. 5 which shows that the total amount of aspiration (expressed by A_{tot}) is more or less constant.

Such knowledge is functionally useful because it is likely to be a feature that distinguishes aspirated from non-aspirated clusters. Although we have not investigated unaspirated stops in the present study due to sparse data for /p, t, k/, our earlier investigations (Ruch & Harrington, 2014; Ruch & Peters, 2016) showed that the extent of pre- and post-aspiration is much less in unaspirated than aspirated clusters. Thus, whereas we find high levels of aspiration irrespective of how the closure is timed in the aspirated cluster of e.g. *pasta* (engl. pasta), the corresponding black line in Fig. 5 is likely to be much lower for the unaspirated /t/ in *pata* (engl. paw). It is from this perspective that it is of functional value to choose a dimension of variation that is likely to provide categorical information for distinguishing between aspirated, i.e. /st/, and unaspirated, i.e. /t/, plosives. The actual classification i.e. distinction between an aspirated and unaspirated cluster would be accomplished not directly by this phonological information but instead by calculating the probability of class membership to the cloud of data points that are also derived by FPCA. Thus the model in Fig. 8 shows, compatibly with exemplar models, that there is a stochastic relationship between phonological knowledge and speech signals: the new angle proposed here is that a transformation analogous to FPCA is intermediary between the two.

A sound change in progress can sometimes be characterised by a perceptual trading relationship between the coarticulatory source and effect. Beddor and colleagues (Beddor, 2009; Beddor, 2012; Beddor et al., 2018) have investigated the phonologisation of nasalisation in the vowels of words exemplified by American English *send* and *sent* from this perspective. They show that there is an inverse relationship in speech production between the extent of vowel nasalisation and the duration of the following /n/ that gives rise to coarticulatory nasalisation. In perceiving *send*, they also show that listeners often identified nasalisation from the signal without associating or parsing it explicitly with either the vowel or following nasal consonant. This trading relationship is an appropriate strategy in perception for such variation, given that nasalisation in the American English variety that they investigated could be manifested in the vowel, in the following nasal consonant, or both in speech production (and variably so between listeners in speech perception).

There is a degree of commonality between Beddor's findings and those in the present speech production study of Andalusian Spanish in which there was shown to be an inverse relationship between pre- and post-aspiration (Fig. 5). This inverse relationship in our study only emerges, however, following the application of an FPCA transformation across speech signals from several speakers that differed in the extent to which they produced /sC/ clusters with pre- or post-aspiration. Without this FPCA transformation, there is no such inverse relationship. This is demonstrated by Figure 9 in which the areas A_{pre} and A_{post} were calculated from Eq. (2) where HF and VP were directly obtained from the speech signals, without applying FPCA. The general conclusion from Fig. 9 is that the inverse relationship between pre- and post-aspiration is not directly manifested in the acoustics of any (or several) /sC/ clusters. It is perhaps for this reason that there has been scant evidence for such a relationship from other studies (Ruch, 2013; Ruch & Harrington, 2014; Ruch & Peters, 2016; Torreira, 2007). The inverse relationship exists instead at a more abstract level that is a consequence of modelling the acoustic speech of multidimensional, time-varying speech signals over many words and repetitions and above all over many speakers of which some have predominantly pre- and others predominantly post-aspiration in producing these clusters (see Appendices C and F for more details).

An individual who has abstracted the phonological and distributional knowledge from the type of data investigated in this study is likely to classify pre- or post-aspiration equivalently. This is apparent in Fig. 5 which shows that, irrespective of whether the closure is late or early, the quantity of aspiration stays more or less the same. Here there are parallels once again to Beddor's findings showing that, at least for some listeners, it does not matter whether the nasalisation occurs in the vowel or the following consonant: both are treated equivalently as being [+nasal]. Whether or not Andalusian listeners actually exhibit such trading relationships is not something that we have yet investigated. Following analogous nasalisation studies by Beddor (2009) and Zellou (2017), we would expect a considerable degree of variation across Andalusian listeners in whether or not such a trading relationship is manifested. The further prediction from the type of model in Fig. 8 is that experience conditions whether or not a listener demonstrates such a trading relationship. Listeners who have been exposed predominantly to old, East Andalusian speakers (who typically pre-aspirate) or those exposed mostly to young, West Andalusian speakers (who typically post-aspirate) are predicted to show much less evidence of a trading relationship than those listeners exposed to both these types of speakers.

The model in Fig. 8 brings together insights from articulatory phonology and episodic models of speech in order to relate synchronic variation to diachronic change. Articulatory phonology has provided great advances in understanding speech dynamics, but given its historical emphasis on articulatory invariants (Fowler, 2003), has been not so easily adaptable to the findings in the last 20 years or so that the relationship between phonological knowledge and speech is a stochastic one. By contrast, exemplar theory has provided great advances in modelling this stochastic relationship. With few exceptions (e.g. Kirchner, Moore, & Chen, 2010), there has, however, been an almost complete neglect in explaining quite how these stochastic phonological categories are derived and associated with multidimensional speech signals that change in time. The idea in the present study that users of the language may extract dimensions across their remembered speech signals brings together these important insights from these separate models. This unified model shares with episodic models of speech that there is no sharp distinction either between synchronic variation and the resulting diachronic change nor between phonological knowledge and the (remembered) speech signals out of which such knowledge is constructed.

CRediT authorship contribution statement

Johanna Cronenberg: Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Michele Gubian:** Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Jonathan Harrington:** Conceptualization, Data curation, Writing - original draft, Supervision, Project administration, Funding acquisition. **Hanna Ruch:** Resources, Data curation, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was supported by European Research Council Grant No. 742289 "Human interaction and the evolution of spoken accent" (2017–2022).

Appendix A. Word List

/sp/
caspa – dandruff
espada – sword
espalda – back
España – Spain
espanto – fright
despide – he/she/it fires
espía – spy
espina – thorn
respira – he/she/it breathes
despierta – awake (fem.)
espiaba – I/he/she/it spied
espiando – spying
disputa – argument
espuma – foam
esputo – sputum
después – later
espuela – spur
/st/
vestuario – wardrobe
hasta – until
pistolín – small pistol
resto – rest
estaba – I/he/she/it was

estado – state
estanco – kiosk
pestaña – eyelash
destino – fate
estima – he/she/it respects
estío – summertime
pestiño – type of pastry
bestial – bestial
bestiando – (pseudoword)
destiempo – untimeliness
estuche – case
estufa – stove
estuve – I was
estuela – (pseudoword)
pasta – pasta
/sk/
vasca – Basque (fem.)
escama – scale
escapa – he/she/it escapes
escaso – insufficient
pescado – fish
cosquillas – tickle (noun)
esquía – he/she/it skies
esquife – skiff
esquina – corner
esquiando – skiing
escucha – he/she/it listens
escudo – shield
escupe – he/she/it spits
escuela – school
escueto – concise

Appendix B. Mathematical details on A_{pre} and A_{post}

In Section 2.2 it was stated that the steepness of the curves $A_{pre}(s_1)$ and $A_{post}(s_1)$, defined in Eq. (2) and shown in Fig. 5, is extreme and not due to chance, while their sum $A_{tot}(s_1)$ does not depart significantly from a flat line. Here we provide a proof of those statements.

Before proceeding with the actual proof, we apply two modifications to the definitions in Eq. (2) that in combination produce linear approximations of $A_{pre}(s_1)$ and $A_{post}(s_1)$. This is necessary in order to obtain properly defined slopes, and also it makes the proof more manageable. First we lift the $HF > VP$ constraint, thus including integration intervals where the area between $HF(t)$ and $VP(t)$ is negative. Then we substitute the integration boundary $t_M = \text{argmin}_t HF(t)$, which varies with s_1 around the middle of the time interval, with the fixed value $t = 0.5$. Looking at Fig. 4 we can see that the impact of these modifications on the shape of $A_{pre}(s_1)$ and $A_{post}(s_1)$ will be rather modest. The inclusion of intervals where $HF < VP$ results in the inclusion of the two small intervals at the beginning and at the end of the normalised time interval where $VP(t)$ is above $HF(t)$. In those intervals, the areas delimited by the curves are small compared with the positive areas and are roughly constant at varying s_1. We can then expect that the main effect of this first approximation step is going to be a downward shift of $A_{pre}(s_1)$ and $A_{post}(s_1)$, as negative areas are going to subtract some (roughly constant) amount from the positive areas. The second approximation step is going to have a clearer, yet modest impact on the shapes of $A_{pre}(s_1)$ and $A_{post}(s_1)$, namely a mitigation of the divergence of the two curves, especially for $s_1 > 0$. For example, in the right panel of Fig. 4 we can see that fixing the demarcation between A_{pre} and A_{post} at $t = 0.5$ allows more area to be assigned to A_{pre} and less to A_{post}; the opposite occurs in the left panel.

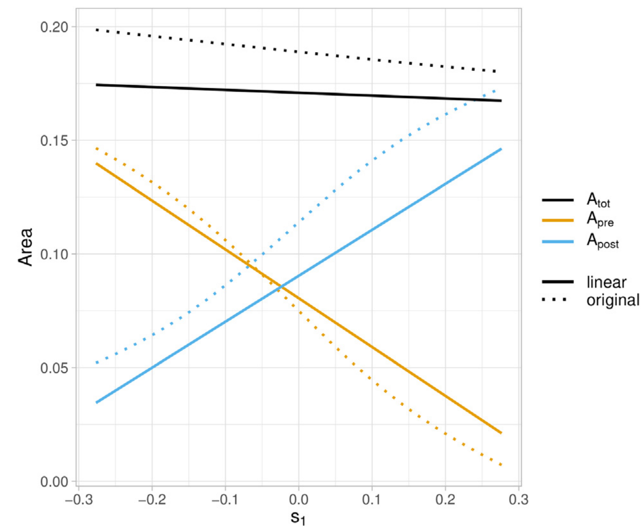

Fig. 10. A_{pre} (yellow), A_{post} (blue), and $A_{tot} = A_{pre} + A_{post}$ (black) as functions of s_1, computed by using Eq. (2) (dotted lines, the same as in Fig. 5) or its approximation Eq. (4) (solid lines), when signals HF and VP are defined as in Eq. (1) using only PC1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Applying the above modifications to Eq. (2) we obtain:

$$A_{pre} \approx \int_0^{0.5} (HF(t) - VP(t))\, dt \tag{4a}$$

$$A_{post} \approx \int_{0.5}^{1} (HF(t) - VP(t))\, dt \tag{4b}$$

Fig. 10 compares A_{pre}, A_{post}, and A_{tot} based on Eq. (2), i.e. the original definition in the main text, to the approximate versions based on Eq. (4), where we can see that the approximate curves (solid) are indeed linear and also close enough to the original ones (dotted), especially as far as their slope is concerned. The fact that $A_{pre}(s_1)$ and $A_{post}(s_1)$ in Eq. (4) are linear in s_1 when $HF(t)$ and $VP(t)$ are computed only on the basis of $PC1$ follows from the linearity of FPCA and the linearity of the definite integral. To make this explicit, the following steps derive an expression for A_{pre}:

$$\begin{aligned} A_{pre}(s_1) &\approx \textstyle\int_0^{0.5} (HF(t) - VP(t))\, dt \\ &= \textstyle\int_0^{0.5} (\mu_{HF}(t) + s_1 \cdot PC1_{HF}(t) - \mu_{VP}(t) - s_1 \cdot PC1_{VP}(t))\, dt \\ &= \underbrace{\int_0^{0.5} (\mu_{HF}(t) - \mu_{VP}(t))\, dt}_{M_{pre}} + s_1 \cdot \underbrace{\int_0^{0.5} (PC1_{HF}(t) - PC1_{VP}(t))\, dt}_{P_{pre}} \\ &= M_{pre} + P_{pre} \cdot s_1 \end{aligned} \tag{5}$$

where the first step is Eq. (4a), the second step is the application of Eq (1) using only s_1, and the rest is term rearrangement and convenient definitions of constants M_{pre} and P_{pre}. Similarly we find $A_{post}(s_1) \approx M_{post} + P_{post} \cdot s_1$ and $A_{tot}(s_1) = A_{pre}(s_1) + A_{post}(s_1) \approx (M_{pre} + M_{post}) + (P_{pre} + P_{post}) \cdot s_1$. The four constant values are: $M_{pre} = 0.08, M_{post} = 0.09, P_{pre} = -0.21$ and $P_{post} = 0.20$.

Having derived linear approximations to $A_{pre}(s_1)$, $A_{post}(s_1)$, and $A_{tot}(s_1)$, we focus on the corresponding slopes P_{pre}, P_{post}, and P_{tot}. We want to prove that P_{pre} and P_{post} are extreme (opposite) values, while P_{tot} is not significantly different from zero. To this end we construct a reference distribution for slopes by extending Eq. (4) to allow for any arbitrary partition of the time interval $(0,1)$ in two complementary subsets on which the two integrals are computed. In this way, the partition $\{(0,0.5),(0.5,1)\}$ is the special case defining A_{pre} and A_{post}, while $\{(0,1),(1,1)\}$ defines A_{tot} (and a null area). In practice we want to show that only by setting the subdivision between A_{pre} and A_{post} in the middle of the total time interval we get a clear trade-off relationship between the two, i.e. steep and opposite slopes P_{pre} and P_{post}, while partitioning the interval between, say, the central and remaining part or any other arbitrary partition would not provide any significant complementary relation. Operationally, we sliced the interval $(0,1)$ in $N_{int}=20$ slots of equal size, i.e. $(0,0.05),(0.05,0.10)$, etc., then defined the corresponding N_{int} sub-areas a_i as:

$$a_i = \int_{\frac{i-1}{N_{int}}}^{\frac{i}{N_{int}}} (HF(t) - VP(t))\, dt,\ i = 1,\dots,N_{int} \tag{6}$$

and then assigned a random subset of them to A'_{pre} and the remaining to A'_{post}, the generalisations of A_{pre} and A_{post} for arbitrary integration limits.

Fig. 11 shows an example of randomly partitioning the integration interval among the ones allowed by the discretisation imposed by Eq. (6). Each random partition will produce different $A'_{pre}(s_1)$ and $A'_{post}(s_1)$, which results in different constant terms M'_{pre}, P'_{pre}, M'_{post}, and P'_{post} defined as in Eq. (5). We are interested in the distribution of P'_{pre} and P'_{post}. Since those have complementary definitions, their distributions are identical, hence we will look at P'_{pre} only. There are $2^{N_{int}}$ possible partitions, that is $2^{20}=1,048,576$ when $N_{int}=20$, each producing a different slope P'_{pre}. We treat this large yet deterministic set of values as a population and describe it empirically as a random distribution. We estimate it by computing 10,000 randomly chosen values of P'_{pre}.

Fig. 12 shows the Empirical Cumulative Distribution Function (ECDF) of P'_{pre}. We can immediately notice that the particular values $P_{pre}=-0.21$ and $P_{post}=0.20$ from Eq. (5), which descend from (a linear approximation of) our quantification of pre- and post-aspiration, are indeed extreme (opposite) values for P'_{pre}, while their sum $P_{tot}=-0.012$ lies inside the central area of the distribution, which includes the value zero, i.e. the slope of a flat line. The distribution of P'_{pre} is quite symmetric (skewness is 0.0058), the median and mean are both around -0.005, its kurtosis is 2.78, very close to 3, which

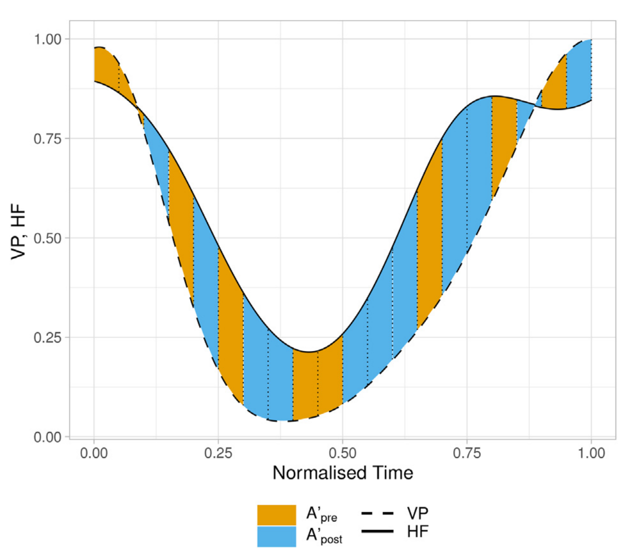

Fig. 11. Random partition of the integration interval obtained by setting $A'_{pre}=a_1+a_2+a_4+a_6+a_9+a_{10}+a_{14}+a_{17}+a_{19}$, and $A'_{post}=A_{tot}-A'_{pre}$, where a_i's are defined in Eq. (6). The particular curves shown here are $HF(t)=\mu_{HF}(t)$ and $VP(t)=\mu_{VP}(t)$, i.e. $s_1=0$.

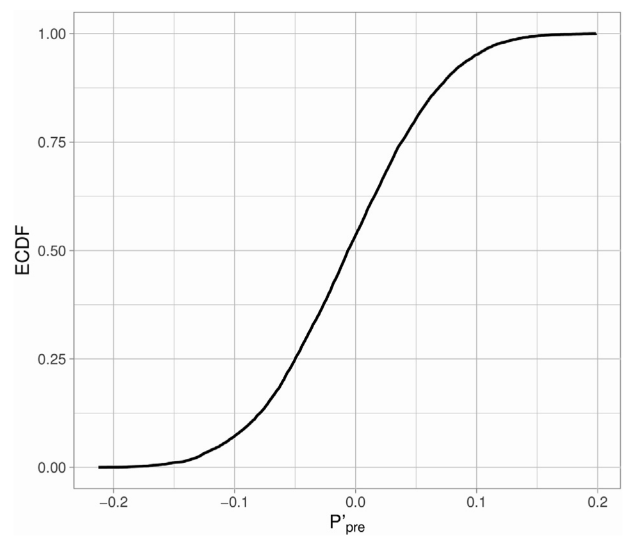

Fig. 12. Empirical Cumulative Distribution Function (ECDF) of P'_{pre}, based on 10,000 random values out of 1 million.

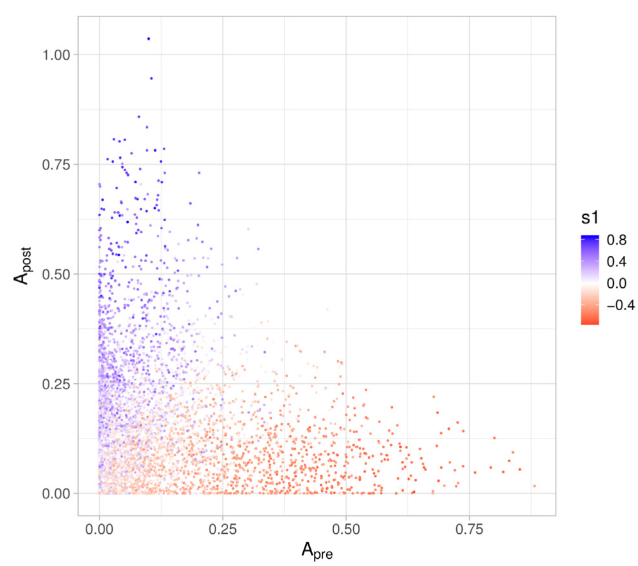

Fig. 13. Area A_{post} against A_{pre} as in Eq. (2) when $HF(t)$ and $VP(t)$ are the curves obtained directly from the speech signal without FPCA transformation. In comparison to Fig. 9, the colour-coding in this plot indicates the values of PC score s_1 for each data point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

allows us to treat it as a Gaussian. The empirical quantile $q_{2.5\%}$ is -0.13, much greater than P_{pre}, and $q_{97.5\%}$ is 0.12, much smaller than P_{post}, i.e. P_{pre} and P_{post} are indeed extremes. On the other hand, the symmetric interval centred around the mean and spanning one standard deviation is $(-0.069, 0.059)$, which includes both zero and P_{tot}, hence supporting the evidence that the curve $A_{tot}(s_1)$ does not depart significantly from a flat line, i.e. the total amount of aspiration is approximately constant throughout the data set we have collected.

Appendix C. Effects of FPCA-based signal decomposition

In Section 2.2 it was shown for the examined data set that pre- and post-aspiration are related through the phasing of the closure when the selected acoustic signals HF and VP were derived from Eq. (1) using only PC1. The same relation, however, was not found when HF and VP were derived from the raw data, as shown by Fig. 9. In this section, we will provide a more detailed explanation for this phenomenon.

Acoustic signals contain a virtually infinite amount of variation, parts of which transmit attributes of the speaker to the listener. Therefore, variations that are important in a possible trade-off between pre- and post-aspiration in Andalusian Spanish can easily be masked by other sources of variation, as shown in Fig. 13. This plot was constructed in the same way as Fig. 9, but every data point is now coloured according to its s_1 value. Recall from Fig. 4 that positive values of s_1 were associated with earlier closures accompanied by post-aspiration, and negative values of s_1 with later closures that leave an interval for pre-aspiration. Even though the areas in Fig. 13 were calculated on the raw HF and VP signals, they bear a strong connection to the s_1 values of the data points: s_1 is positively correlated with A_{post} and negatively with A_{pre} (cf. Fig. 5). When a token is post-aspirated according to its s_1 value (blue), it has a small A_{pre}, but can have any A_{post}, and vice versa for pre-aspirated (red) tokens. A_{post} in post-aspirated tokens can take low or high values because of global amplitude variations in the HF signal, but importantly, A_{pre} is always small in these cases. There is, therefore, a (barely discernible) trade-off between pre- and post-aspiration in the raw

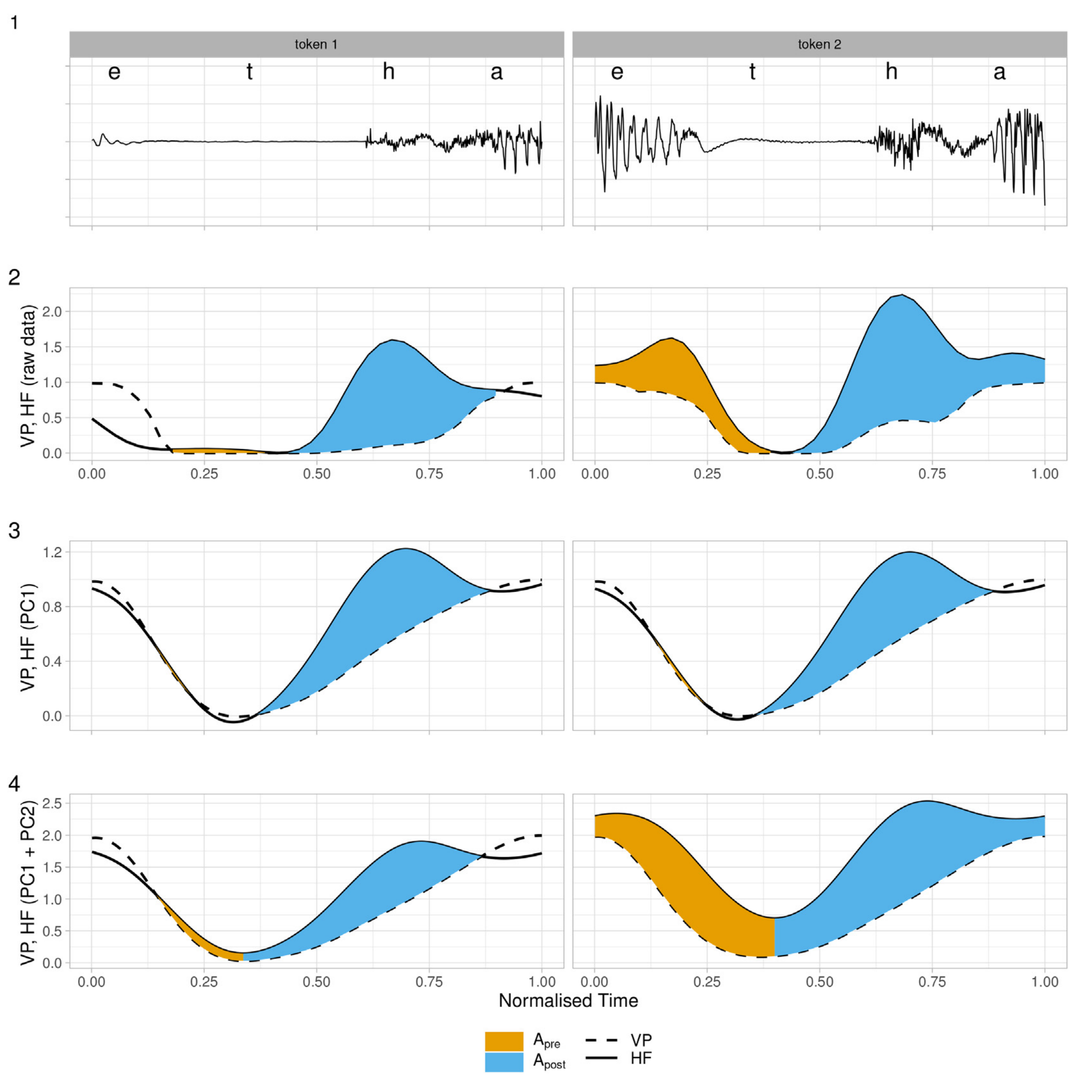

Fig. 14. Two post-aspirated tokens of the word *estanco* (engl. kiosk) produced by different young West Andalusian speakers. When reconstructing HF (solid) and VP (dashed) curves based on only PC1 (row 3), the areas A_{pre} (yellow) and A_{post} (blue) are similar for both tokens, but they are markedly different from each other when the reconstruction of the curves is based on both PC1 and PC2 (row 4). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

data (as shown by the association of the raw data points to s_1 and consequently to the areas computed on signals reconstructed using PC1), but the trade-off is buried beneath many other kinds of variation (see e.g. Appendix D). This might also be the reason why other studies were unable to identify a trading relationship (e.g. Ruch, 2013; Ruch & Harrington, 2014; Ruch & Peters, 2016; Torreira, 2007).

FPCA is a powerful tool for disentangling sources of variation in data sets of continuous signals so that previously unnoticeable, systematic variations can be examined while excluding other influences. This methodological advantage of FPCA is demonstrated in Fig. 14 which shows the raw data (row 2) as well as signals derived from PC1 (row 3), and both PC1 an PC2 together (row 4) for a production of *estanco* (engl. kiosk) by two different speakers. The areas A_{pre} (yellow) and A_{post} (blue) between VP and HF were computed for all panels using Eq. (2). The pairs of signals (and hence: the areas between them) are very similar to each other in row 3, i.e. both panels correctly show that there is more post- than pre-aspiration in the two tokens. That is because they have a very similar closure phasing (PC score s_1 is 0.34 for token 1 and 0.32 for token 2). However, when the signals are derived from both PC1 and PC2, a large difference between the two emerges: the HF signal of token 1 is shifted downwards ($s_2 = -0.12$), resulting in overall smaller areas, whereas the HF signal of token 2 is shifted upwards ($s_2 = 0.59$) which has the opposite effect on the areas. The latter signal reconstructions (row 4) are a closer approximation to the originals (row 2) than those based on PC1 only (row 3), and they still correctly indicate more post- than pre-aspiration for both tokens ($A_{post} > A_{pre}$). However, A_{post} and A_{pre} are now rendered incomparable across tokens, e.g. A_{post} in token 2 is considerably larger than A_{post} in token 1, because they are affected by a kind of variation that is irrelevant to the trade-off between pre- and post-aspiration, namely the global energy level in /sC/ clusters. From another perspective, factoring out the variation expressed by PC2 (and all other PCs except PC1) provides a way of standardising the area measurements with the consequence that the trade-off between pre- and post-aspiration emerges from the remaining amplitude differences in the PC1-derived HF signal.

Appendix D. Variation explained by PC2 and PC3

When applying FPCA to the input HF and VP signals, the kind of variation that is most relevant to the sound change from pre- to post-aspiration in /sC/ clusters was found to be captured by PC1. As shown in Fig. 4, PC1 expresses a phase shift of the HF minimum relative to a voiceless interval, which supports the model of this sound change given in Fig. 1 and suggests that there is a trading relationship between pre- and post-aspiration in Andalusian Spanish. However, we also computed the second and third principal component. Here we will analyse these further components and explain why we excluded them from the main part of this study.

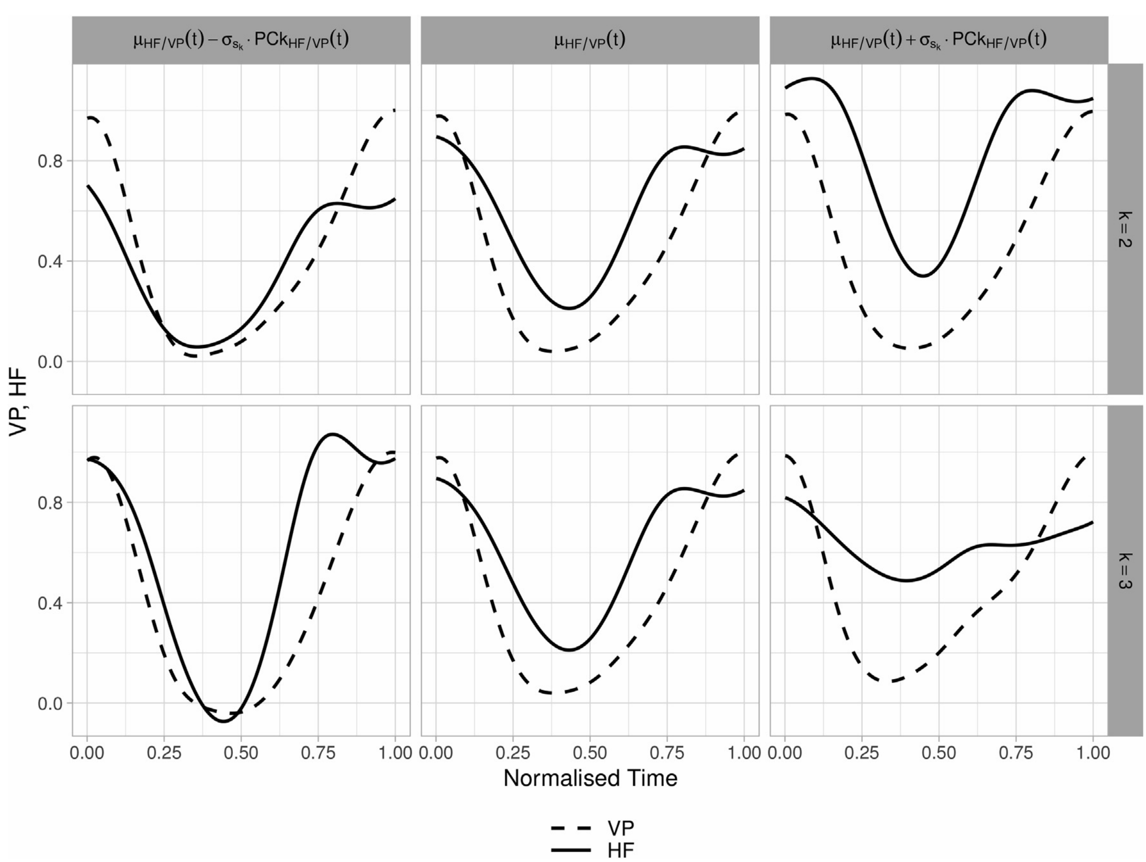

Fig. 15. Modification of the mean curves $\mu_{HF}(t)$ and $\mu_{VP}(t)$ (middle panel) by adding to (right column) or subtracting from (left column) each mean curve only one PC*k* curve multiplied by the standard deviation of its corresponding score (0.24 for s_2, 0.19 for s_3). The top row corresponds to PC2, the bottom row to PC3. VP curves are dashed, HF curves are solid lines.

Fig. 15 was constructed in the same way as Fig. 4, but for the second and third PC. That is, the mean HF and VP curves are shown in the middle column; they are the same for all PCs. These mean curves were then modified by adding to (right column), or subtracting from (left column), each curve only one PC*k* curve multiplied by the standard deviation of its corresponding PC score (0.24 for s_2, 0.19 for s_3). The top row corresponds to PC2, the bottom row to PC3. PC2, which explained 24.3% of all variance, captures global amplitude differences predominantly in the HF signal (with almost no change to VP). This is likely caused by speaker-specific variation in energy, e.g. their distance from the microphone during the field recordings and their amplitude level while speaking. Rather than modifying the raw speech signals, we let FPCA catch these differences. We did however try to scale the HF signals by speaker which removed most of this variation (but it did not change the main result reported in this paper), confirming that PC2 indeed captures global, speaker-specific energy levels. PC3, which explained only 14.5% of the variance in the input signals, encodes a compression and expansion of the HF curve with some slight parallel changes to VP. This kind of variation is more difficult to interpret; however, it is very clear that the PC3 curves do not contribute to an explanation of the sound change from pre- to post-aspiration in Andalusian Spanish which the present study has shown to be a phase shift of the closure in /sC/ with respect to the voiceless interval.

We constructed the same LMER models with s_2 and s_3 as response variables as we did for s_1 in Section 3.1. After pruning, all fixed terms with interactions were retained for s_2, whereas all interactions as well as the fixed factor age were dropped for s_3 (however, we added age back into make the analysis homogeneous across the three PC scores). For s_2, the random intercept for word as well as the random slope for speaker were retained. All random slopes were retained for s_3. The post hoc tests were again computed using the R package `emmeans`. This allowed us to construct Fig. 16 in the same way as Fig. 6, but for the second and third PC scores. These boxplots show that s_2 was generally higher for alveolar and velar than for bilabial clusters. Given that PC2 captures amplitude differences in the HF signal, it is perhaps unsurprising that s_2 was lower for bilabial than for the other two cluster types as /p/ typically has a weaker burst than /t, k/. The results of the mixed model confirmed that s_2 was significantly influenced by cluster type ($F[2, 60.1] = 18.6$, $p < 0.001$). The LMER model for s_2 also shows a significant interaction between age and cluster ($F[2, 44.4] = 5.3$, $p < 0.01$) as well as a significant three-way interaction between the fixed factors ($F[2, 44.3] = 3.3$, $p < 0.05$). The former interaction can be observed in Fig. 16 where s_2 takes slightly higher values for older than for younger speakers from West Andalusia producing /sp/ or /sk/. The post hoc tests showed that there was a significant difference between older and younger speakers from West Andalusia producing /sp/ ($t = 2.6$, $p < 0.05$). Furthermore there were significant s_2-differences between /sp/ and /st/ for older East ($t = 4.6$, $p < 0.001$), younger East ($t = 5.1$, $p < 0.001$), older West ($t = 2.9$, $p < 0.05$), and younger West Andalusian speakers ($t = 6.6$, $p < 0.001$) as well as

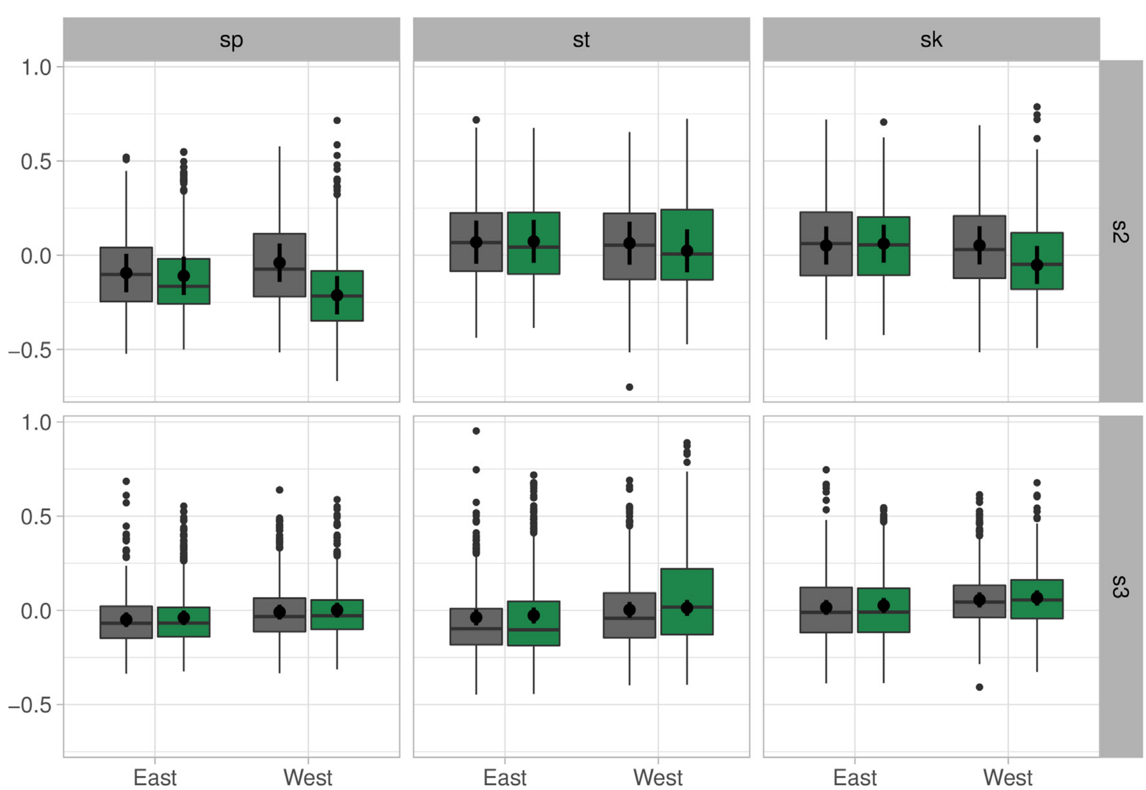

Fig. 16. Boxplots of s_2 (top row) and s_3 (bottom row) as well as their estimated marginal means (black dots within the boxes) with related confidence intervals (black vertical bands) based on the LMER models described in the text. Younger speakers are shown in green, older ones in dark grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

between /sp/ and /sk/ for older East (t = 3.9, p < 0.001), younger East (t = 4.7, p < 0.001), older West (t = 2.5, p < 0.05), and younger West Andalusian speakers (t = 4.4, p < 0.001).

PC score s_3, on the other hand, was significantly influenced by region (F[1, 50.3] = 5.8, p < 0.05) and cluster type (F[2, 65.1] = 6.2, p < 0.01), as shown by the mixed model. Fig. 16 shows that s_3 was slightly higher for speakers from West than from East Andalusia for all three places of articulation, most visibly so for younger speakers producing /st/. For both age groups and all cluster types, the post hoc tests confirmed a significant regional difference in s_3 (t = 2.3, p < 0.05). Furthermore, there was a significant difference between /sp/ and /sk/ (t = 3.2, p < 0.01) as well as between /st/ and /sk/ (t = 2.6, p < 0.05).

Table 1 reports marginal and conditional coefficients of determination or Pseudo-R^2 scores (Johnson, 2014; Nakagawa & Schielzeth, 2013) for the three LMER models. These coefficients were calculated using R package `MuMIn` (Version 1.43.15). The values roughly correspond to the fraction of variance explained by the fixed factors only and by the whole model, respectively. The low marginal and conditional coefficient values for PC3 indicate that this component did not contribute much to explaining the variance in the input data, as was shown previously. The random elements (Conditional - Marginal) seem to be more relevant to the s_2 model than to the s_1 model. This could indicate that the vertical shift of the high-frequency energy signal shown in the PC2 panels of Fig. 15 was conditioned more by the variation introduced by individual speakers or words, while the timing shift of the energy signal modelled by PC1 was a systematic effect governed by the fixed factors. We therefore assume PC2 to be a correcting influence on PC1 rather than to be conceptually relevant to the sound change in progress itself.

Together, the analysis of the second and third PC in this Appendix as well as Appendix C show that neither of them contributes any information essential to the sound change modelled in Fig. 1, and that both need to be factored out from the main analysis because they would otherwise obscure the relation between pre- and post-aspiration in Andalusian Spanish.

Table 1
Rounded marginal and conditional coefficients of determination for the three Linear Mixed Effects models.

PC Score	Marginal	Conditional
s_1	0.11	0.46
s_2	0.13	0.70
s_3	0.03	0.27

Table 2
Specifications for scatter plots in Fig. 19.

Fig.	Method	Signals	Time norm.	Measures
19a	Manual annot.	–	No	Durations
19b	Manual annot.	–	Yes	Durations
19c	$HF(t), VP(t)$	Raw	Yes	Durations
19d	$HF(t), VP(t)$	PC1-based	Yes	Durations
19e	$HF(t), VP(t)$	Raw	Yes	Areas
19f	$HF(t), VP(t)$	PC1-based	Yes	Areas

Appendix E. Including duration in FPCA

The analysis presented in the main text is based on the linearly time-normalised $HF(t)$ and $VP(t)$ signals, as a result of which the total duration d between time points A and B in Fig. 2 has not been taken into account. Here we show that total duration d, though obviously varying among tokens, does not play a significant role in characterising the pre-/post-aspiration trade-off.

In order to preserve the original duration of $HF(t)$ and $VP(t)$ we adopt an extended version of multi-dimensional FPCA that incorporates time warping $r(t)$ as an extra dimension (Gubian, Boves, & Cangemi, 2011). This additional curve encodes the relationship between the original and normalised time axis and decouples the information about curve shape from its duration (i.e. this is a special case of non-linear time warping). In the simple case of linear time normalisation, the time warping curve $r(t)$ is a flat horizontal line taking the value $-\log\frac{d}{\mathrm{mean}(d)}$, i.e. (minus) the log of the normalised token duration (see Gubian, Boves, & Cangemi, 2011 and Appendix A in Asano & Gubian, 2018 for an extended explanation[4]). The analysis was then carried out as a standard FPCA on the three-dimensional signals $(HF(t), VP(t), r(t))$, where the first two dimensions are the same as the ones used in the main text, thus still expressed in normalised time, while $r(t)$ separately encodes total duration in the way described above. For the analysis of the results, $r(t)$ was converted back to ordinary duration values.

Fig. 17 shows the variation of $HF(t)$ and $VP(t)$ when approximated by PC1 only. Different curve shapes are associated with different durations, from 210 ms (left panel), to 235 ms (middle panel), to 263 ms (right panel).

Despite the different type of signals, two- vs. three-dimensional, FPCA captured basically the same trends for the shape of $HF(t)$ and $VP(t)$, as can be seen by comparing Fig. 17 with Fig. 4 in the main text, i.e. PC1 still encodes a phase shift of $HF(t)$. The preservation of the trends found in the duration-agnostic analysis is not a general rule, since added information on duration can break and rearrange statistical associations (encoded by PCs). In this case, we note that longer tokens are associated with more post-aspiration (right panel), which is an expected result as post-aspiration is inherently longer than pre-aspiration.

Fig. 18 shows a corrected version of Fig. 5, where $HF(t)$ and $VP(t)$ are approximated using PC1 from the duration-aware FPCA in which the areas were computed on unnormalised time intervals (note the different scale on the y-axis, reflecting a multiplication by duration in ms). The main difference is that A_{tot} incorporates the duration trend associated with PC1 in which a higher s_1 and hence a longer integration interval are derived from longer tokens. Despite that, the trends

[4] The cited sources introduce $r(t)$ as a result of landmark registration, a procedure that was not applied here. The reader consulting those sources should consider linear time normalisation as a special case of landmark registration, where the only landmarks are placed at signal start and end. The resulting time warping function $h(t)$ is a segment whose inclination is higher (resp. lower) than 45° when total duration is higher (resp. lower) than average. The corresponding $r(t)$ is a flat line defined on the normalised time axis taking the value reported in the main text.

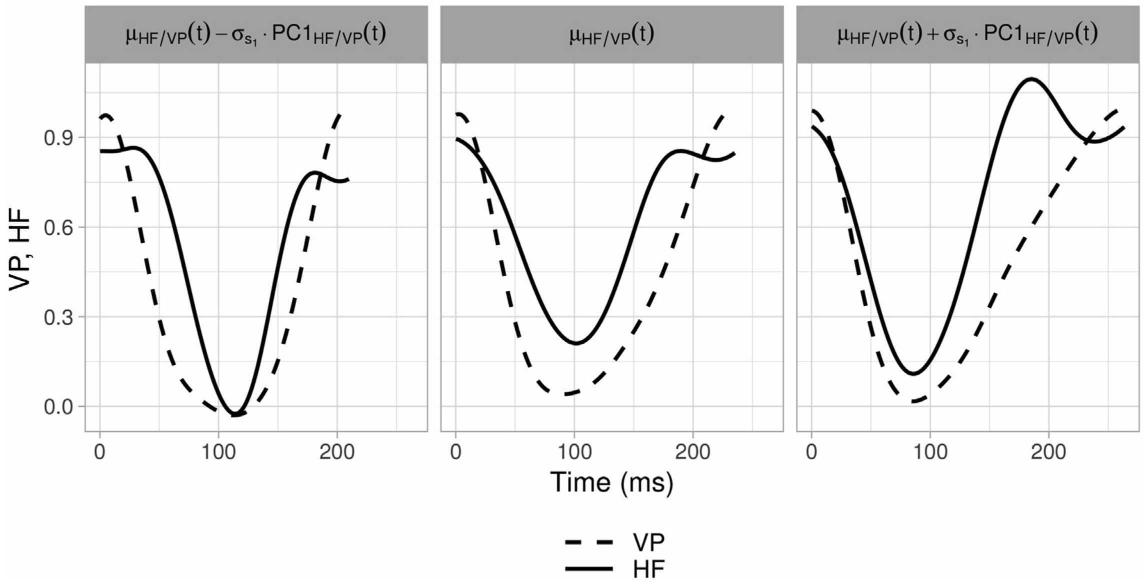

Fig. 17. Variation of $HF(t)$ (solid) and $VP(t)$ (dashed) as modulated by score s_1 (cf. Fig. 4 in the main text). The curves were obtained applying FPCA to three-dimensional signals $(HF(t), VP(t), r(t))$, where $HF(t)$ and $VP(t)$ are in normalised time and $r(t)$ (not shown) encodes duration. The corresponding durations are, from left to right: 210 ms, 235 ms, 263 ms.

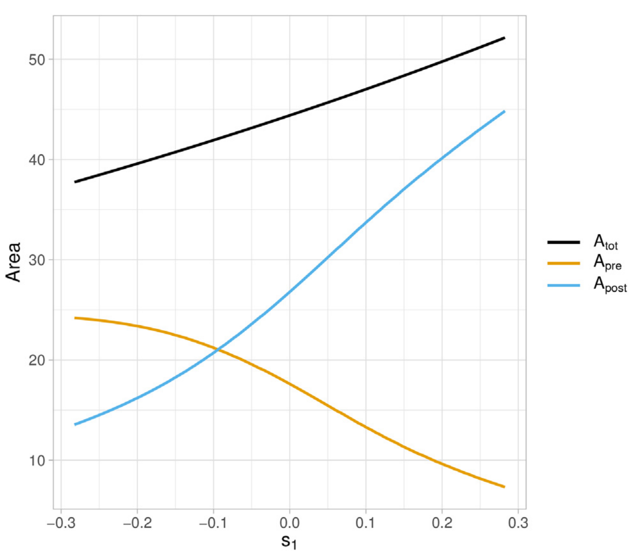

Fig. 18. A_{pre} (yellow), A_{post} (blue), and $A_{tot} = A_{pre} + A_{post}$ (black) computed as in Fig. 5, based on $HF(t)$ and $VP(t)$ corrected for duration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

found in the main analysis remain, as A_{pre} and A_{post} are clearly in a trade-off, mildly distorted by the rising trend of A_{tot}. In conclusion, the duration-aware version of FPCA has enriched the analysis of the pre-/post-aspiration trade-off reported in the main text without disrupting it in any significant way. In other words, there is no evidence that the trade-off is confounded by total duration. This is because in both FPCA analyses (with and without total duration) post-aspiration was characterised by an early closure (minimum of $HF(t)$).

Appendix F. Role of areas, time normalisation, and FPCA

Here we illustrate how A_{pre} and A_{post} as defined in Eq. (2) compare to a number of alternative ways to obtain measures of pre- and post-aspiration that reliably and effectively show the underlying trade-off relation between the two. In particular, we explain the role of linear time normalisation, FPCA, as well as the use of the pre- and post-aspiration areas as opposed to using more conventional segmental durations. For this purpose, we present a number of scatter plots showing the distribution of 5845 data points, in which each point corresponds to an /sC/ token in our data set, and the x and y axes are pre- and post-aspiration measures either in the form of areas or durations, according to different definitions and procedures. Table 2 summarises (i) the different combinations of method of computation, either based on manual annotation or based on the $HF(t)$ and $VP(t)$ signals as defined in 2.1.2, (ii) whether the $HF(t)$ and $VP(t)$ signals were the raw versions or the PC1-based reconstructions, (iii) whether or not linear time normalisation was applied, and (iv) whether segmental durations or areas were used as a measure of aspiration.

Fig. 19a and b are based on semi-automatically annotated durations of pre- and post-aspiration, where the annotation was taken from Ruch and Harrington (2014).[5] Fig. 19a is based on unnormalised measures, while 19b shows the effect of dividing the durations in Fig. 19a by the total duration of the /sC/ interval (defined as in Ruch & Harrington, 2014). While pre- and post-aspiration show a mild negative correlation (Spearman's correlation are -0.13 for Fig. 19a and -0.24 for Fig. 19b), it is clear that several factors contribute to the physical duration of the acoustic manifestation of aspiration which blur the underlying trade-off that we hypothesise to be at the base of the planned articulation gesture and that we want to isolate.

Fig. 19c is obtained from the same procedure as Fig. 19b, but with the difference that the durations of pre- and post-aspiration were obtained from the $HF(t)$ and $VP(t)$ signals by computing the length of the integration intervals defined in

[5] In Fig. 19a and b, d_{pre} takes negative values when the voicing of the previous vowel extended into the following closure.

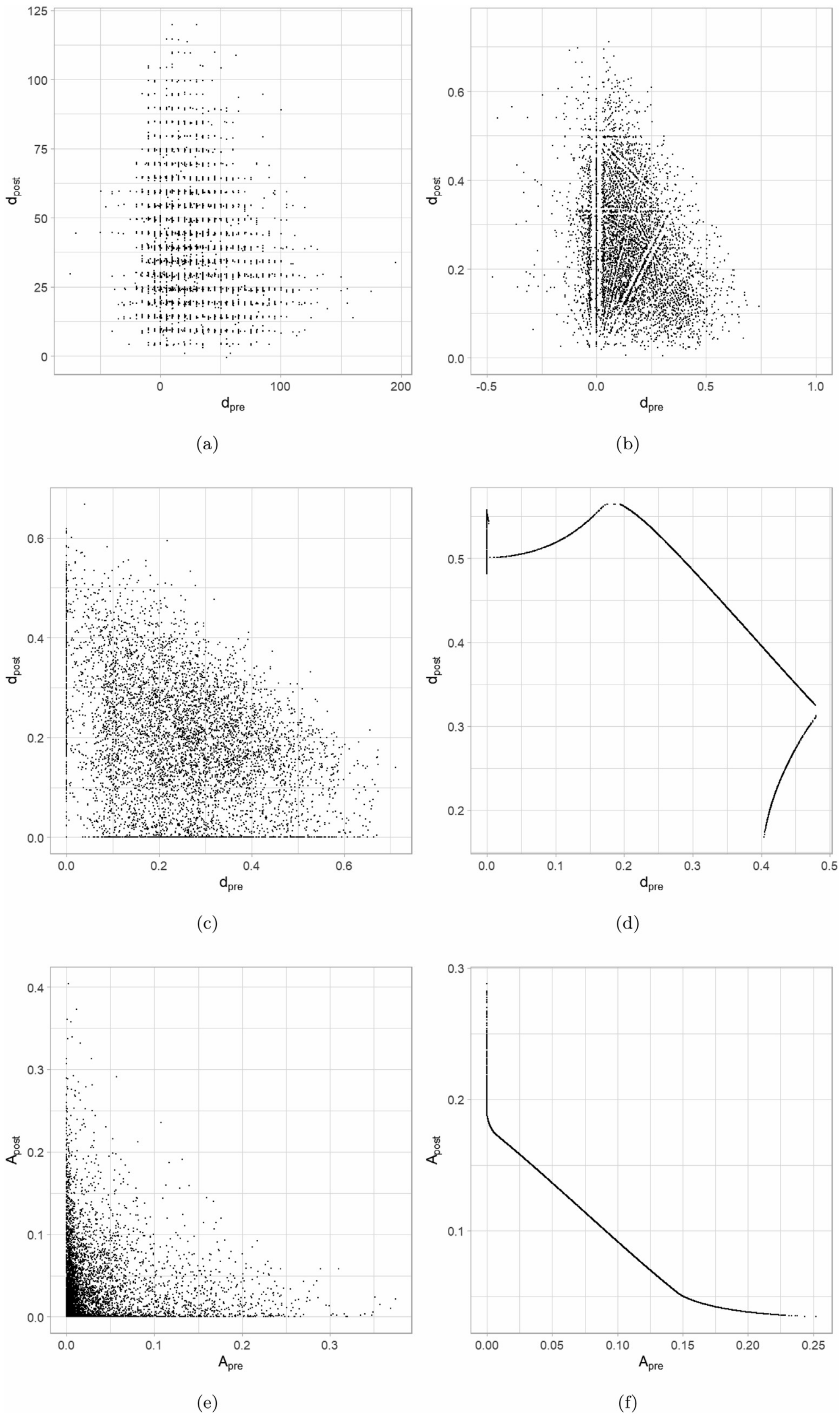

Fig. 19. Scatter plots showing the distribution of pre- (x-axis) and post-aspiration (y-axis) measures obtained using different methods. See Table 2 and text for details.

Eq. (2). In other words, d_{pre} (resp. d_{post}) is the total duration of the interval before (resp. after) the minimum value of $HF(t)$, conditioned by $HF(t) > VP(t)$. There is an improvement in the visible correlation (Spearman's correlation is -0.31), but still the trade-off is not clearly delineated from other factors that contribute to duration. The situation is even worse when computing the areas A_{pre} and A_{post} instead of durations, as shown in Fig. 19e (same as Fig. 9). By contrast, a few lines rather than a cloud of points are obtained when the same parameters were derived from the PC1-based reconstructed signals (Fig. 19d and f). Those lines are still scatter plots, i.e. formed by individual points, but this time the location of the points is constrained by a single degree of freedom, i.e. PC score s_1. In Fig. 19d a large portion of the scatter plot exhibits an obvious trade-off (the segment with roughly -45° inclination), while other parts are affected by what we argue are artefacts. These are the consequence of the fact that $d_{pre} + d_{post} \leqslant 1$ (because the total duration of the signals is 1 in normalised time), but at the same time d_{pre} (resp. d_{post}) cannot be larger than t_M (resp. $1 - t_M$), where t_M is the location of the minimum of $HF(t)$, which is usually around the temporal midpoint $t = 0.5$ and rarely occurs near $t = 0$ or $t = 1$. As a consequence, when either d_{pre} or d_{post} decreases below approx. 0.3, the other stops increasing, i.e. the trade-off between d_{pre} and d_{post} is interrupted. This explains e.g. the roughly vertical line at the bottom right corner in Fig. 19d, where d_{post} keeps decreasing from 0.3 to 0.2 while d_{pre} stops increasing (and even decreases slightly), as according to its definition its value cannot exceed t_M, which is not likely to be far from $t = 0.5$. These artefacts are not present when areas were used instead of durations, as Fig. 19f illustrates, where a clear trade-off relation is preserved even when either A_{pre} or A_{post} are small.

To summarise, with Fig. 19 we have shown that for the purpose of isolating the underlying pre-/post-aspiration trade-off (i) linear time normalisation alone does not bring any particular benefit, (ii) a clear trade-off emerges only by applying the FPCA-based signal decomposition on $HF(t)$ and $VP(t)$, and when doing so, (iii) computing areas instead of segmental durations preserves a clear trade-off trend also at the extremes of pre-/post-aspiration ranges.

References

Aleza Izquierdo, M., & Enguita Utrilla, J. M. (2002). *El español de América: Aproximación sincrónica*. Valencia: Tirant lo Blanch.

Alfonso, P. J., & Baer, T. (1982). Dynamics of Vowel Articulation. *Language and Speech, 25*(2), 151–173.

Alvar, M. (1955). Las hablas meridionales de España y su interés para la lingüística comparada. *Revista de Filología Española, 39*(1/4), 284–313. https://doi.org/10.3989/rfe.1955.v39.i1/4.1136.

Asano, Y., & Gubian, M. (2018). "Excuse meeee!!": (Mis)coordination of lexical and paralinguistic prosody in L2 hyperarticulation. *Speech Communication, 99*, 183–200. https://doi.org/10.1016/j.specom.2017.12.011.

Beddor, P. S. (2009). A Coarticulatory Path to Sound Change. *Language, 85*(4), 785–821. https://doi.org/10.1353/lan.0.0165.

Beddor, P. S. (2012). Perception Grammars and Sound Change. In M.-J. Solé & D. Recasens (Eds.), *The Initiation of Sound Change: Perception, Production, and Social Factors* (pp. 37–56). Amsterdam: John Benjamins.

Beddor, P. S., Brasher, A., & Narayan, C. (2007). Applying Perceptual Methods to the Study of Phonetic Variation and Sound Change. In M.-J. Solé, P. S. Beddor, & M. Ohala (Eds.), *Experimental Approaches to Phonology* (pp. 127–143). Oxford: Oxford University Press.

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language, 94*(4), 931–968. https://doi.org/10.1353/lan.2018.0051.

Bermúdez-Otero, R. (2015). Amphichronic Explanation and the Life Cycle of Phonological Processes. In P. Honeybone & J. C. Salmons (Eds.), *The Oxford Handbook of Historical Phonology* (pp. 374–399). Oxford: Oxford University Press.

Bermúdez-Otero, R., & Trousdale, G. (2012). Cycles and continua: On unidirectionality and gradualness in language change. In T. Nevalainen & E. Closs Traugott (Eds.), *The Oxford Handbook of the History of English* (pp. 691–720). Oxford: Oxford University Press.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics, 29*(3), 191–211. https://doi.org/10.3758/BF03207286.

Bladon, A. (1986). Phonetics for hearers. In G. McGregor (Ed.), *Language for hearers* (pp. 1–24). Oxford: Pergamon.

Boersma, P., Escudero, P., & Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1013–1016). Barcelona.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219–252. https://doi.org/10.1017/S0952675700000658.

Browman, C. P., & Goldstein, L. (1992). Articulatory Phonology: An Overview. *Phonetica, 49*(3–4), 155–180. https://doi.org/10.1159/000261913.

Butterworth, S. (1930). On the Theory of Filter Amplifiers. *Wireless Engineer, 7*(6), 536–541.

Canfield, D. L. (1981). *Spanish pronunciation in the Americas*. Chicago: University of Chicago Press.

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *Journal of the Acoustical Society of America, 144*(3), EL172–EL177. https://doi.org/10.1121/1.5052025.

Davidson, L. (2006). Schwa Elision in Fast Speech: Segmental Deletion or Gestural Overlap? *Phonetica, 63*(2–3), 79–112. https://doi.org/10.1159/000095304.

Ettlinger, M. (2007). An exemplar-based model of chain shifts. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 685–688). Saarbrücken.

Fant, G. (1973). *Speech Sounds and Features*. Cambridge: MIT Press.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics, 36*(4), 359–368.

Fowler, C. A. (2003). Speech production and perception. In I. B. Weiner, A. F. Healy, & R. W. Proctor (Eds.), *Handbook of Psychology. Experimental Psychology* (4, pp. 237–266). New York: Wiley.

Fowler, C. A. (2005). Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics, 33*(2), 199–213. https://doi.org/10.1016/j.wocn.2004.10.003.

Fowler, C. A., & Saltzman, E. (1993). Coordination and Coarticulation in Speech Production. *Language and Speech, 36*(2–3), 171–195. https://doi.org/10.1177/002383099303600304.

Fowler, C. A., & Smith, M. R. (1986). Speech Perception as Vector Analysis: An Approach to the Problem of Invariance and Segmentation. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and Variability in Speech Processes* (pp. 123–136). Hillsdale: Lawrence Erlbaum Associates.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics, 62*(8), 1668–1680. https://doi.org/10.3758/BF03212164.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 349–366. https://doi.org/10.1037/0096-1523.28.2.349.

Gerfen, C. (2002). Andalusian codas. *Probus, 14*(2), 247–277. https://doi.org/10.1515/prbs.2002.010.

Gonzalez, S., & Brookes, M. (2014). PEFAC – A pitch estimation algorithm robust to high levels of noise. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22*(2), 518–530. https://doi.org/10.1109/TASLP.2013.2295918.

Gubian, M., Boves, L., & Cangemi, F. (2011). Joint analysis of f0 and speech rate with Functional Data Analysis https://doi.org/10.1109/ICASSP.2011.5947472.

Gubian, M., Harrington, J., Stevens, M., Schiel, F., & Warren, P. (2019). Tracking the New Zealand English NEAR/SQUARE merger using functional principal components analysis https://doi.org/10.21437/Interspeech.2019-2115.

Gubian, M., Torreira, F., & Boves, L. (2015). Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics, 49*, 16–40. https://doi.org/10.1016/j.wocn.2014.10.001.

Hagège, C., & Haudricourt, A.-G. (1978). *La phonologie panchronique: Comment les sons changent dans les langues*. Paris: Presses universitaires de France.

Haggard, M., Summerfield, Q., & Roberts, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: evidence from trading F0 cues in the voiced-voiceless distinction. *Journal of Phonetics, 9*(1), 49–62. https://doi.org/10.1016/S0095-4470(19)30926-X.

Harmon, Z., Idemaru, K., & Kapatsinski, V. (2019). Learning mechanisms in cue reweighting. *Cognition, 189*, 76–88. https://doi.org/10.1016/j.cognition.2019.03.011.

Harrington, J., & Cassidy, S. (1999). *Techniques in Speech Acoustics*. Dordrecht: Springer Netherlands.

Harrington, J., Kleber, F., Reubold, U., Schiel, F., & Stevens, M. (2018). Linking Cognitive and Social Aspects of Sound Change Using Agent-Based Modeling. *Topics in Cognitive Science*, 1–22. https://doi.org/10.1111/tops.12329.

Herrero de Haro, A. (2017). The phonetics and phonology of Eastern Andalusian Spanish: a review of literature from 1881 to 2016. *Revista de Lenguaje y Cultura, 22*(2), 313–357. https://doi.org/10.17533/udea.ikala.v22n02a09.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America, 119*(5), 3059–3071. https://doi.org/10.1121/1.2188377.

Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic Explanations for the Development of Tones. *Language, 55*(1), 37–58. https://doi.org/10.2307/412518.

Hualde, J. I., & Sanders, B. P. (1995). A new hypothesis on the origin of the Eastern Andalusian vowel system. In: Proceedings of the 21st Annual Meeting of the Berkeley Linguistics Society, Berkeley. pp. 426–437.

Hualde, J. I., & Chitoran, I. (2016). Surface sound and underlying structure: The phonetics-phonology interface. In S. Fischer & C. Gabriel (Eds.), *Manual of Grammatical Interfaces in Romance* (pp. 23–40). Berlin: De Gruyter.

Hyman, L. M. (2013). Enlarging the scope of phonologization. In A. C. L. Yu (Ed.), *Origins of Sound Change* (pp. 3–28). Oxford: Oxford University Press.

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance, 37*(6), 1939–1956. https://doi.org/10.1037/a0025641.

Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *Journal of the Acoustical Society of America, 132*(6), 3950–3964. https://doi.org/10.1121/1.4765076.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145–165). San Diego: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: the emergence of social identity and phonology. *Journal of Phonetics, 34*, 485–499. https://doi.org/10.1016/j.wocn.2005.08.004.

Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods in Ecology and Evolution, 5*(9), 944–946. https://doi.org/10.1111/2041-210X.12225.

Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics, 36*(1), 28–54. https://doi.org/10.1016/j.wocn.2007.02.001.

Kiparsky, P. (2015). Phonologization. In P. Honeybone & J. C. Salmons (Eds.), *The Oxford Handbook of Historical Phonology* (pp. 563–582). Oxford: Oxford University Press.

Kirby, J. P. (2014). Incipient tonogenesis in Phnom Penh Khmer: Computational studies. *Laboratory Phonology, 5*(1), 195–230. https://doi.org/10.1515/lp-2014-0008.

Kirchner, R., Moore, R. K., & Chen, T.-Y. (2010). Computing phonological generalization over real speech exemplars. *Journal of Phonetics, 38*(4), 540–547. https://doi.org/10.1016/j.wocn.2010.07.005.

Labov, W. (1994). *Principles of Linguistic Change. Volume 1: Internal Factors*. Oxford: Blackwell.

Lehet, M., & Holt, L. L. (2017). Dimension-Based Statistical Learning Affects Both Speech Perception and Production. *Cognitive Science, 41*, 885–912. https://doi.org/10.1111/cogs.12413.

Lindblom, B., Guion, S. G., Hura, S. L., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica, 7*(1), 5–37.

Martin, J. G., & Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance, 8*(3), 473–488. https://doi.org/10.1037/0096-1523.8.3.473.

McQueen, J. M., Norris, D., & Cutler, A. (2006). The Dynamic Nature of Speech Perception. *Language and Speech, 49*(1), 101–112. https://doi.org/10.1177/00238309060490010601.

Momcilovic, N. B. (2009). A sociolinguistic analysis of /s/-aspiration in Madrid Spanish. LINCOM studies in Romance linguistics. München: LINCOM.

Mondéjar Cumpián, J. (2001). In P. Carrasco & M. Galeote (Eds.), *Dialectología Andaluza: Estudios* (1). Málaga: Universidad de Málaga.

Moya Corral, J. A. (2007). Noticia de un sonido emergente: La africada dental procedente del grupo -st- en Andalucía. *Revista de Filología de la Universidad de La Laguna, 25*, 457–466.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4* (2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

Navarro Tomás, T. (1938). Dédoublement de phonemes dans le dialecte andalou. *Travaux du Cercle Linguistique de Prague, 8*, 184–186.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9.

Ohala, J. J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 258–282). Cambridge: Cambridge University Press.

Ohala, J. J. (1993). Sound change as nature's speech perception experiment. *Speech Communication, 13*(1–2), 155–161. https://doi.org/10.1016/0167-6393(93)90067-U.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical Linguistics: Problems and Perspectives* (pp. 237–278). London: Longman.

O'Neill, P. (2010). Variación y cambio en las consonantes oclusivas del español de Andalucía. *Estudios de fonética experimental, 19*, 11–41.

Parrell, B. (2012). The role of gestural phasing in Western Andalusian Spanish aspiration. *Journal of Phonetics, 40*(1), 37–45. https://doi.org/10.1016/j.wocn.2011.08.004.

Penzl, H. (1949). Umlaut and Secondary Umlaut in Old High German. *Language, 25*(3), 223–240. https://doi.org/10.2307/410084.

Pierrehumbert, J. B. (2003). Phonetic Diversity, Statistical Learning, and Acquisition of Phonology. *Language and Speech, 46*(2–3), 115–154.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics, 34*(4), 516–530. https://doi.org/10.1016/j.wocn.2006.06.003.

Ramsammy, M. (2015). The Life Cycle of Phonological Processes: Accounting for Dialectal Microtypologies. *Language and Linguistics Compass, 9*(1), 33–54. https://doi.org/10.1111/lnc3.12102.

Ramsay, J., & Silverman, B. W. (2010). *Functional Data Analysis*. New York: Springer New York.

Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance, 40*(2), 539–555. https://doi.org/10.1037/a0034409.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*(1), 81–110.

Romero, J. (1994). An Articulatory View of Historical S-Aspiration in Spanish. *Haskins Laboratories Status Report on Speech Research*, 255–266.

Ruch, H. (2013). *Lautvariation und Lautwandel im andalusischen Spanisch: Prä- und Postaspiration bei*. München: /s/ vor stimmlosen Plosiven (Doctoral dissertation, LMU.

Ruch, H. (2018). Perception of speaker age and speaker origin in a sound change in progress: The case of /s/-aspiration in Andalusian Spanish. *Journal of Linguistic Geography, 6*(1), 40–55. https://doi.org/10.1017/jlg.2018.4.

Ruch, H., & Harrington, J. (2014). Synchronic and diachronic factors in the change from pre-aspiration to post-aspiration in Andalusian Spanish. *Journal of Phonetics, 45*, 12–25. https://doi.org/10.1016/j.wocn.2014.02.009.

Ruch, H., & Peters, S. (2016). On the Origin of Post-Aspirated Stops: Production and Perception of /s/ + Voiceless Stop Sequences in Andalusian Spanish. *Laboratory Phonology, 7*(1), 1–36. https://doi.org/10.5334/labphon.2.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics, 71*(6), 1207–1218. https://doi.org/10.3758/APP.71.6.1207.

Savoia, L., & Maiden, M. (1997). Metaphony. In M. Maiden (Ed.), *The Dialects of Italy* (pp. 15–25). London: Routledge.

Solé, M.-J. (2010). Effects of syllable position on sound change: an aerodynamic study of final fricative weakening. *Journal of Phonetics, 38*(2), 289–305. https://doi.org/10.1016/j.wocn.2010.02.001.

Stevens, M., Harrington, J., & Schiel, F. (2019). Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English. *Glossa, 4*(1), 1–30. https://doi.org/10.5334/gjgl.620.

Terrell, T. D. (1980). Diachronic reconstruction by dialect comparison of variable constraints: -s aspiration and deletion in Spanish. In D. Sankoff & H. Cedergren (Eds.), *Variation omnibus* (pp. 115–124). Edmonton: Linguistic Research.

Todd, S., Pierrehumbert, J. B., & Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: an exemplar-based model. *Cognition, 185*, 1–20. https://doi.org/10.1016/j.cognition.2019.01.004.

Torreira, F. (2006). Coarticulation between aspirated-s and voiceless stops in Spanish: an interdialectal comparison. In: Selected Proceedings of the 9th Hispanic Symposium, Somerville. pp. 113–120.

Torreira, F. (2007). Pre- and postaspirated stops in Andalusian Spanish. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and Prosodic Issues in Romance* (pp. 67–82). Amsterdam: John Benjamins.

Torreira, F. (2012). Investigating the nature of aspirated stops in Western Andalusian Spanish. *Journal of the International Phonetic Association, 42*(1), 49–63. https://doi.org/10.1017/S0025100311000491.

Torres-Tamarit, F., Linke, K., Oostendorp, M. v. (Eds.). (2016). Approaches to metaphony in the languages of Italy. Berlin: De Gruyter Mouton.

Twaddell, W. F. (1938). A note on Old High German Umlaut. *Monatshefte für Deutschen Unterricht, 30*(3/4), 177–181.

van der Kooij, E., & van der Hulst, H. (2005). On the internal and external organization of sign language segments: Some modality-specific properties. In M. van Oostendorp & J. van de Weijer (Eds.), *The Internal Organization of Phonological Segments* (pp. 153–180). Berlin: Mouton de Gruyter.

Villena-Ponsoda, J. A. (2008). Sociolinguistic patterns of Andalusian Spanish. *International Journal of the Sociology of Language, 2008*(193/194), 139–160. https://doi.org/10.1515/IJSL.2008.052.

Wedel, A. (2006). Exemplar models, evolution and language change. *The Linguistic Review, 23*(3), 247–274. https://doi.org/10.1515/TLR.2006.010.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1990). Gradient Effects of Fundamental Frequency on Stop Consonant Voicing Judgments. *Phonetica, 47*, 36–49.

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics, 61*, 13–29. https://doi.org/10.1016/j.wocn.2016.12.002.