Lingua 122 (2012) 145–163

Contents lists available at ScienceDirect

Lingua

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

A phonetically based metric of sound similarity[☆]

Jeff Mielke [*]

Department of Linguistics, University of Ottawa, Arts 401, 70 Laurier Ave East, Ottawa, ON K1N 6N5, Canada

ARTICLE INFO

Article history:
Received 21 January 2011
Received in revised form 20 April 2011
Accepted 21 April 2011
Available online 24 June 2011

Keywords:
Phonetic similarity
Phonological similarity
Ultrasound
Electroglottographym
Airflow
Acoustic

ABSTRACT

This paper examines similarity measures based on acoustic and articulatory data from a set of crosslinguistically frequent consonants and vowels, and compares this phonetic similarity with measures of phonological similarity that are based on the crosslinguistic patterning of phonemes associated with these sounds.

1. Introduction

A basic question in phonological theory is *Why is the class of obstruents active in many different languages?* Many similarly phrased questions can be asked about other groups of sounds. The leading answer to this question may once have been *Because the feature [sonorant] is in Universal Grammar*. In the absence of this UG-based explanation, a comparable answer would be *Because obstruents are phonetically similar to each other*. Supporting this second answer requires a means of quantifying phonetic similarity. Phonetic similarity is frequently invoked for explaining a wide range of phonological observations, but to this day no objective metric of phonetic similarity is widely available.

The purpose of the project described in this paper is to provide a resource for quantifying phonetic similarity, distinguishing articulatory, acoustic, and perceptual similarity from each other and from phonological notions of similarity, such as those based on features (e.g. Frisch, 1996; Frisch et al., 2004; Kondrak, 2003, *et seq.*) or on phonological patterning. This paper examines similarity measures based on acoustic and articulatory data from a set of crosslinguistically frequent consonants and vowels, and compares this phonetic similarity with measures of phonological similarity that are based on the crosslinguistic patterning of phonemes associated with these sounds.

☆ Collection of the phonetic data reported here was made possible by Diana Archangeli and Natasha Warner and supported by James S. McDonnell Foundation grant #220020045 BBMB to Diana Archangeli and funding from the College of Social and Behavioral Sciences at the University of Arizona. Data analysis was supported by SSHRC grant #410-2007-0735, ''Measuring the Phonetic Similarity of Speech Sounds'' and funding from the University of Ottawa Faculty of Arts and the University Research Committee. The collection of P-base was supported by a Presidential Fellowship from the Ohio State University. Research assistants for data analysis at the University of Ottawa were Tanya Awad, Daryl Chow, Chris Delaney, Tabish Ismail, Lyra Magloughlin, Kayleigh Markell, Joseph Roy, Seiko Sagae, Ricardo Tabone, Agnesa Terziyska, and Marie-Claude Tremblay. This project has benefited from comments from Natasha Warner, Osamu Fujimura, Shigeto Kawahara, Søren Wichmann, an anonymous reviewer, and audience members at NELS 40. All mistakes are mine.

* Tel.: +1 6135625800x2701.
E-mail address: jmielke@uottawa.ca.

0024-3841/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.lingua.2011.04.006

The original motivation for creating the similarity metric was to investigate the role of phonetic similarity in determining the sets of segments that are involved in sound patterns. It has long been known that the phonologically active classes appearing in languages' sound systems are not randomly assembled, but reflect general crosslinguistic tendencies. Distinctive feature theory (Chomsky and Halle, 1968; Clements, 1985; Clements and Hume, 1995, *inter alia*) was put forth as a model of the phonetic parameters used to form contrasts and define natural classes. More recently, it has been argued that an all-purpose feature model is in conflict with many attested phonologically active classes, including recurrent ones (Mielke, 2008a), and that the preference for certain classes is driven by physiological and perceptual factors and their role in diachronic change (Ohala, 1981; Blevins, 2004, *inter alia*). The goal of this project is to provide tools for investigating the role of ubiquitous phonetic factors in accounting for which sounds tend to pattern together and which sounds tend not to.

The purpose of this project is not to replace an all-purpose phonological model with an all-purpose phonetic model. However, the study of phonology does deal with a wide range of observations that by their nature are large-scale and language-independent, and make use of broad "all-purpose" statements, such as markedness generalizations, general phonetic explanations for typological patterns, and the types of phenomena that are often accounted for in terms of distinctive features and other primitives of phonological theory. These observations have often received language-independent phonetic and phonological accounts. One of the reasons for focusing on phonetically based explanations for phonological patterns (in spite of the fact sound patterns are also influenced by social, cognitive, lexical, and grammatical factors), is that statements about phonetic tendencies apply generally to all spoken languages, and the other factors vary too much for general statements to be made as easily. This idea of phonetics as a common denominator is the basis for historical phonetic accounts of phonological patterns (Ohala, 1981, 1983, 1993, 1994; Blevins, 2004). Formal primitives such as distinctive features have been useful in investigating crosslinguistic patterns, because they are designed to apply to all languages. This project aims to provide a measure of similarity that does not rely on formal primitives. Nevertheless, there is no substitute for accounting for phonological patterns with phonetic data from the languages where the patterns are observed, and to use particular phonetic data produced by speakers of another language as a stand-in is obviously still a simplification.[1]

This project aims to provide a more direct route from phonetic observations to phonological patterns, unmediated by formal primitives, but it still does not avoid idealization. All of the sounds described here were produced in intervocalic position in a laboratory by linguists whose native language is English. There are many important ways in which a particular consonant or vowel produced under these circumstances does not adequately represent a consonant or vowel in a particular language that could potentially be identified with it. Any segmental properties that are specific to other languages, prosodic positions, vowel contexts, speech styles, or anything like that would not be represented in the data resulting from this project.[2] However, many phonologically relevant properties are expected to hold up. For example, the [t]s produced here are all produced as voiceless coronal stops, as are most of the [t]s of the world's languages. So while not every phoneme described as /t/ is pronounced the same way, it is still useful to have some phonetic data points for segments described as [t], as compared to other segments, as an alternative to using a bundle of distinctive features or a three-part articulatory description as the common denominator. For many of the phonetic properties linguists are interested in, and that are relevant for language-independent statements, phonetic measurements of the articulatory and acoustic properties of speech sounds produced under specific circumstances bear non-trivial relevance.

The presentation of the data in this paper involves a specific set of measurements of phonetic data that could be measured in numerous different ways. As the measurements are discussed in the next section, advantages and limitations and other alternatives will be described where relevant.

In addition to the phonetic similarity data, this paper describes a metric of phonological similarity based on shared phonological patterning in a database of sound patterns. This database is organized in terms of phonemes. Using phonetic data of the kind reported here to address data involving phonemes requires a simplifying assumption, specifically the association of a phoneme with a particular allophone of that phoneme. This is similar to the assumptions that linguists use when making use of resources such as segment inventory databases like UPSID (Maddieson and Precoda, 1990), which uses IPA symbols to describe phoneme inventories in a range of languages that do not all realize these phonemes in the same ways. See Simpson (1999:349) for a critique the use of UPSID in these terms and see also Mielke (2008b) for a discussion of these issues. For the same reasons, the phonetic representations of the type described here clearly do not interface directly with language-specific phonological representations, which involve various phonemic oppositions and allophonic patterns. What allows this approach to make sense is the assumption that a phoneme that is described with a particular IPA symbol will have some connection with a speech sound produced by a linguist reading the same symbol. There is obviously not a direct connection between a phonetic representation resulting from an English-speaking linguist in a laboratory and a phonology-phonetics mapping in a different language, but the potential connections are also obvious, and no more remote than the connection between a set of distinctive features and a phonological representation in a particular language (until shown otherwise).

[1] Note that using phonetic data from the same language also involves a simplification in may cases: if the role of phonetic effects is primarily in diachrony, then the most appropriate phonetics experiment would require access to speakers of an earlier stage of the same language.

[2] See, e.g. Sproat and Fujimura (1993) on the importance of prosodic position for segment realization.

p		t	ʈ	c	k	kʷ		i	ɨ	ɯ	u
b		d	ɖ	ɟ	g		ʔ	iː	ɨː		uː
		ts		tʃ				e			o
		dz		dʒ				eː			oː
ɸ	f	s		ʃ	x		h	ɛ	ə		ɔ
β	v	z		ʒ	ɣ			æ			a
m		n	ɳ	ɲ	ŋ						aː
		l	ɭ	ʎ							
		ɾ	ɽ								
		r									
			ɻ	j		w					

Fig. 1. Consonants and vowels analyzed in this paper.

2. Phonetic similarity

To generate data for phonetic measurements, four phonetically trained linguists each produced the same 58 crosslinguistically frequent consonants and vowels shown in Fig. 1. An additional set of 72 less frequent sounds were divided up among the speakers and produced by one speaker each.[3] The linguists were all native speakers of American English, and included one female and three males. In addition to English, Subject #1 speaks French and Hebrew, Subject #2 also speaks Romanian and Russian, Subject #3 (female) also speaks Dutch, German, Japanese, and Korean, and Subject #4 also speaks German, Spanish, and Turkish.

Each subject produced a total of nine repetitions of each consonant and vowel in three contexts ($3 \times a\_a$, $3 \times i\_i$, $3 \times u\_u$). Thus, all of the consonants were produced in onset position and all of the vowels were produced as a vowel in between two vowels. Vowels were produced in this way so that all of the segments would be produced in the same contexts, in order not to introduce expected differences between classes of sounds. Nevertheless, the fact that vowels and consonants are not ordinarily produced in the same contexts is a reason for them to be have different phonological patterning. Vowels were also recorded in isolation, but those data are not discussed here.

2.1. Data collection

Each subject was recorded in two sessions, because some of the data collection techniques are incompatible with each other. Session 1 involved audio, video, and ultrasound tongue imaging, and session 2 involved audio, electroglottography, and airflow measurement.

In session 1, audio was recorded using an Audio-Technica PRO 49Q condenser microphone and a Symetrix 302 microphone preamplifier. The tongue was imaged using a SonoSite TITAN portable ultrasound machine with a C-11/7-4 11-mm broadband curved array transducer placed under the chin, generating a mid-sagittal section from near the tongue root to near the tongue tip at a rate of 28 scans per second, output as 29.97 fps analog video.[4] The palate is not typically visible during imaging of the tongue, but can be made visible by filling the mouth with water. At the beginning of each session, the palate was imaged by asking the subject to drink water through a straw. Movements of the velum are visible during swallowing, and this motion was used to identify a tissue point where the soft palate meets the hard palate.

The subject's head was recorded in profile with a Sony Mini-DV Digital Handycam, and this video was used to record lip movements and to record head and transducer movements for later correction. The audio channel and the two video channels (ultrasound and camcorder) were combined and digitized using a Videonics MXProDV digital video mixer. The audio was digitized at a sampling rate of 48 Hz with 16 bits per sample. The result of combining the two video channels is shown in Fig. 2. The tongue surface is the bright white contour in the fan-shaped ultrasound view (the sound is [a]), and the four light-colored dots are on blue sticks attached to the glasses and the ultrasound transducer. A blue screen behind the subject allows everything but the dots and the subject's face to be removed from the camcorder video before it is superimposed on the ultrasound video.

[3] The additional sounds were divided up as follows: Subject #1 produced [ʋ ɬ ĩ ũ ẽ õ ɛ̃ ɔ̃ ã y ɪ ʊ ø ʌ œ ɛː ɔː ɑ], Subject #2 produced [ᵐb ⁿd ᵑg pʲ tʲ kʲ bʲ dʲ gʲ fʲ sʲ xʲ hʲ ʋʲ mʲ nʲ lʲ rʲ], Subject #3 produced [ʘ | ‖ ! ǂ pʰ tʰ kʰ ɓ ɗ p' t' k' q' kʷ' tsʰ tʃʰ ts' tʃ'], and Subject #4 produced [k͡p q g͡b θ χ ħ ð ʁ ʕ pʷ qʷ bʷ gʷ xʷ χʷ hʷ zʷ mʷ ŋʷ].

[4] It is often not possible to see the tongue tip on ultrasound images, because it is obscured by a sublingual air pocket, or the mandible, or leaves the transducer's field of view. For the recordings described here, the transducer was positioned so that the tongue tip was visible in sounds produced as far forward as the alveolar ridge, but not in interdental consonants. No interdental consonants are described in this paper. 29.97 fps is the standard rate for NTSC video, and is slightly higher than 28 Hz, which is the maximum rate of the radial scanning of the ultrasound transducer. This means that while it takes about 36 ms for the transducer to complete a sweep from one side of the fan to the other, each video frame represents about 33 ms. The only consequence of this is that each frame contains about 3 ms of scanning that also appeared on the previous frame.

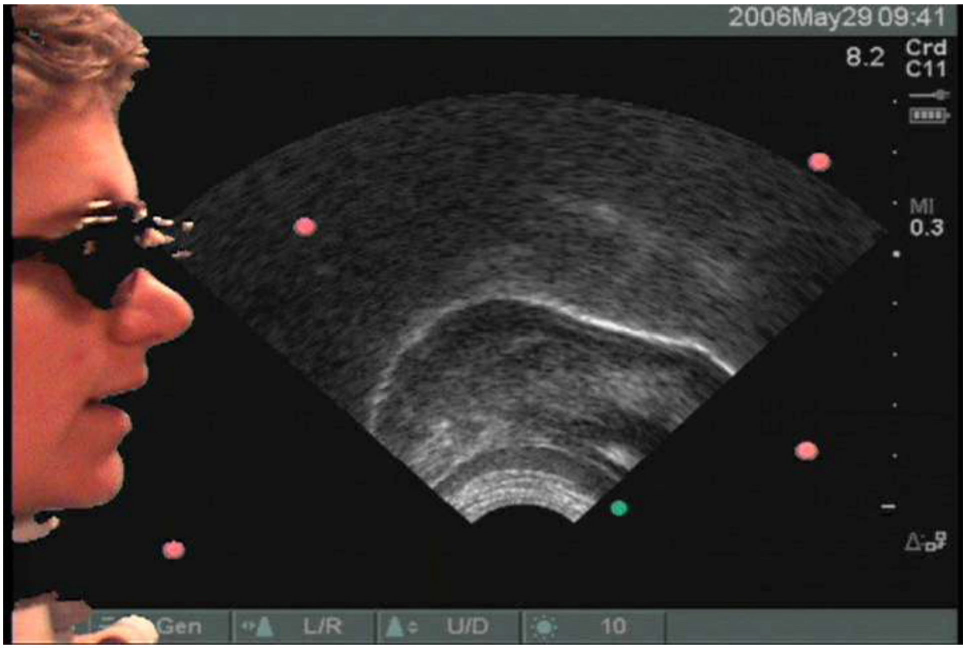

Fig. 2. Video data collection.

In session 2, oral and nasal airflow were measured using a SCICON R&D Macquirer 516 transducer interface system and the associated oral and nasal masks. A Glottal Enterprises EG2-PC electroglottograph (EGG) was used to collect laryngeal data. The primary EGG signal is a voltage representing the amount of contact between the vocal folds (Childers and Krishnamurthy, 1985). During phonation, this signal is periodic. The peaks correspond to maximal contact between the vocal folds and the troughs correspond to minimal contact. The EG2-PC also produces a measure of larynx height. This makes use of the same parameter used for initial placement of the electrodes over the larynx, which is based on a comparison of the EGG signals detected by two pairs of electrodes. Subsequent changes in the apparent position of the electrodes (relative to the larynx), are interpreted as changes in larynx height. The two signals output by the EGG system were recorded together with the airflow data by connecting them to two DC input channels of the Macquirer 516. The four channels of articulatory data were recorded at a sampling rate of 1375 Hz using the SCICON's Macquirer software.[5] Audio was recorded using the microphone built into the oral mask but used only as a reference for segmenting the articulatory data, and not for further analysis of the audio itself.

2.2. Data analysis

All of the recordings from both sessions were segmented using TextGrids in Praat (Boersma and Weenink, 2007). The TextGrids were used for all subsequent measurements and comparisons in all of the channels.

2.2.1. Airflow and larynx activity

The articulatory data from session 2 (airflow and EGG) was measured as follows. A C++ program (Baker, 2006) was used to convert articulatory data from a proprietary format into wav files and tab-delimited text files. Each text file, containing 1375 samples per second for each articulatory data channel, was converted to four wav files using Python's `wave` module and measured using a Praat script. The oral and nasal airflow measures used for comparison of the sounds were the average airflow values during the target segment interval. Since the larynx height measurement is very sensitive to differences of anatomy and electrode placement, these factors were corrected for by taking the larynx height value at the midpoint of the target segment interval and subtracting the average of the larynx height values at the beginning and the end of the target segment interval, giving a relative measure of larynx height. The

[5] This is not an ideal sampling rate for the main EGG signal. The analysis of EGG data benefits from recording at sampling rates appropriate for audio signals. 1375 Hz is adequate for studying the frequency and amplitude of the signal, because it is much higher than twice the frequency of vocal fold vibration. Other measures of larnyx activity that are concerned with the shape of the glottal waveform, such as the Open Quotient (Rothenberg, 1979), benefit from a higher sampling rate.

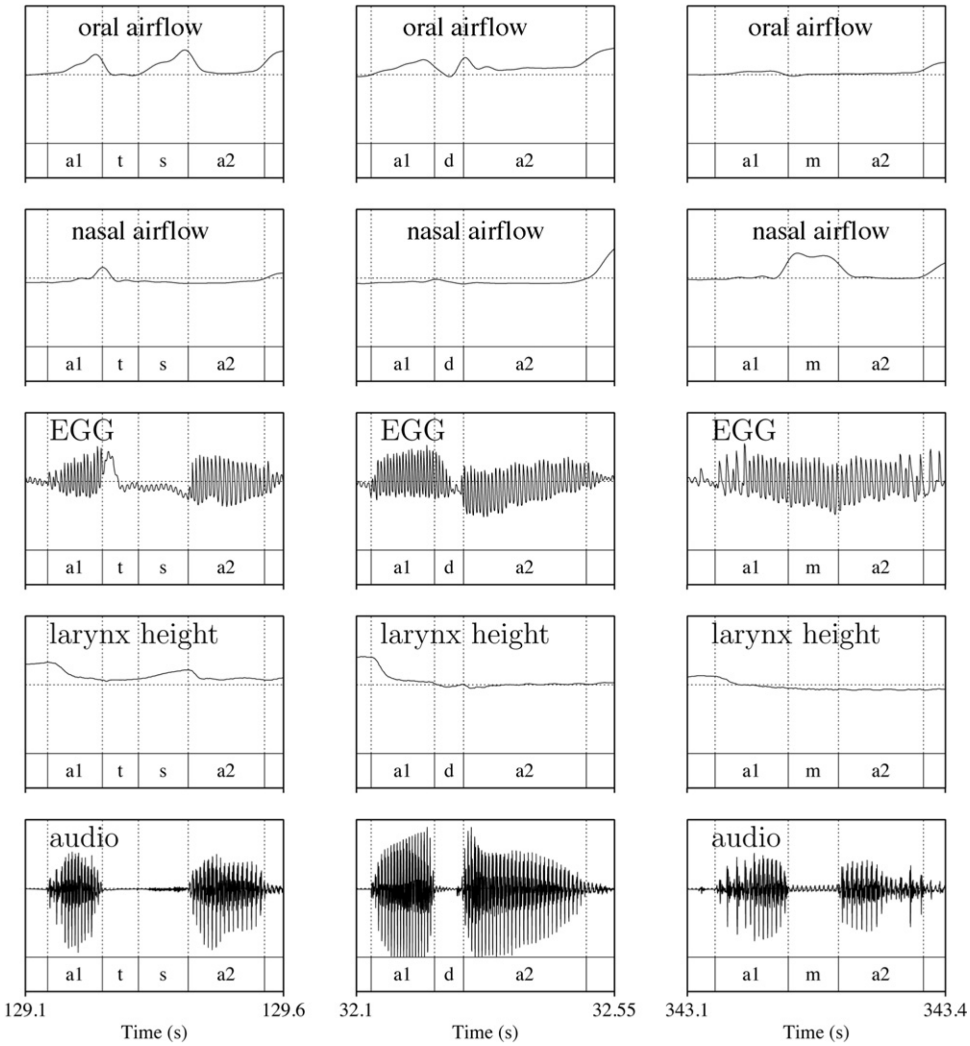

Fig. 3. Sample articulatory data for [atsa], [ada], and [ama], showing segmentation used to make measurements.

average intensity of the main EGG signal (vocal fold contact area) in decibels was taken as a measure of vocal fold activity.[6]

Fig. 3 shows sample airflow and EGG data for three tokens produced by Subject #3. Oral and nasal airflow were measured by calculating the mean value (in milliliters per second) during the target sound interval (in these examples, the target intervals are the ones labeled "t", "s", "d", and "m"). For clarity, oral and nasal airflow have both been smoothed by low-pass filtering at 30 Hz for this figure, in order to remove periodicity caused by voicing, but filtering was not used for the analysis, since mean values were calculated over intervals much longer than one voicing period. The EGG signal was measured by calculating the average intensity during the target interval, and larynx height was measured at the segment boundaries and the midpoint of the target segment (for [ts] and other target segments that were segmented internally, the measurement was taken at the midpoint of the entire target segment). The audio waveform is shown for reference.

To validate the airflow and laryngeal measurements and to see how consistent they are across the four subjects, the data from all four subjects are plotted in Figs. 4 and 5. Not surprisingly, Fig. 4 shows that for all four subjects, nasals are clearly distinguished from all other segments by nasal airflow, and fricatives are distinguished from all other segments by oral airflow. Fig. 5 shows that EGG signal intensity distinguishes segments by voicing, which was its main reason for being measured. Voiced obstruents show some larynx lowering relative to sonorants for all but Subject #1, but for the most part larynx height is not as good as EGG signal intensity for showing expected differences. The female subject (#3) shows the most separation between the three groups of sounds, and Subject #4 shows the least. The differences observed with Subject #3 seem likely to have an anatomical explanation, and the others could be due to individual differences or electrode

[6] Many other measures are possible. At the present time, the main purpose of these data is to represent the amount of vocal fold vibration.

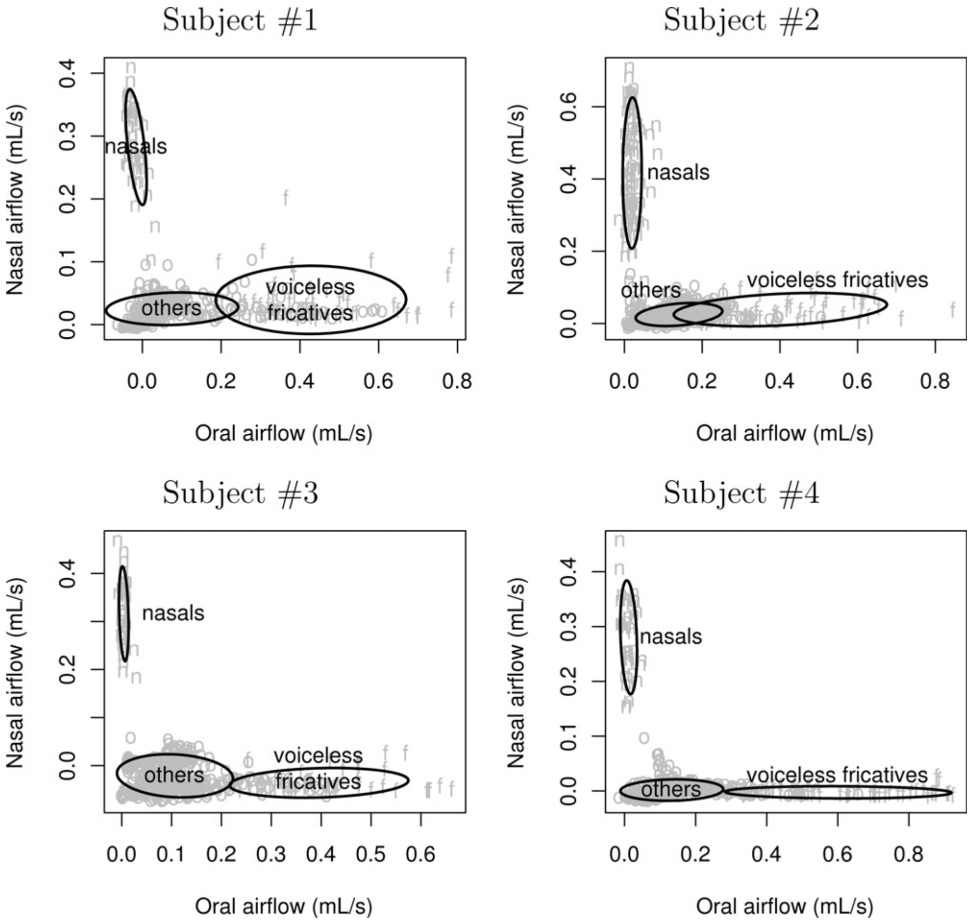

Fig. 4. Airflow measurements by subject. Gray letters represent individual tokens, and ellipses represent one standard deviation from the mean for each group of sounds.

placement. Larynx height appears to be more suspect than the other measurements, although it may be more useful when analyzing the glottalized sounds that were produced by some of the subjects but not analyzed here.

2.2.2. Acoustic data

Acoustic distances between segments were calculated using a dynamic time warping algorithm (Holmes and Holmes, 2001) implemented in Python. This technique allows a comparison of utterances that are not necessarily time-aligned.[7] First, the waveforms for each token (each VXV sequence produced by the speakers) were converted into matrices of 12 Mel-frequency cepstral coefficients (MFCCs) and 12 delta coefficients based on 15ms windows at 5ms intervals, using Praat. MFCCs provide a representation of the spectral properties of the signal at different points in time, and the delta coefficients represent the spectral changes over time.

Fig. 6 illustrates the use of the dynamic time warping technique (DTW) to measure acoustic similarity in three tokens from Subject #1. This figure illustrates pairwise comparisons between tokens of [aba] and [ada] and between tokens of [aba] and [afa]. The *y*-axes of both figures represent the spectral properties of the token of [aba], and the *x*-axes represent [ada] and [afa] in the same way. The darkness of shading in the DTW plot represents the acoustic distance between any pair of points in time in the two sounds being compared.[8] The greatest acoustic difference is between the fricative interval of [afa] and the vowel intervals of [aba].

The purpose of the DTW algorithm is to find the optimal path from the bottom left corner (the beginning of both sounds) to the top right corner (the end of both sounds). If the two sounds were identical, the optimal path would be the diagonal. The path deviates from the diagonal in order to travel through points where the acoustic distance (darkness) is less, with the

[7] Dynamic time warping discards temporal differences between segments, meaning that the acoustic similarity of sounds that differ only in duration or timing is overstated. A more complete measure of acoustic similarity would be based at least in part on comparisons that do not involve time warping.

[8] The illustration includes spectrograms and DTW plots made using Praat, and the path found by Praat's DTW algorithm, but the comparisons reported here were based on matrices of MFCCs created in Praat and fed to a DTW algorithm implemented in Python.

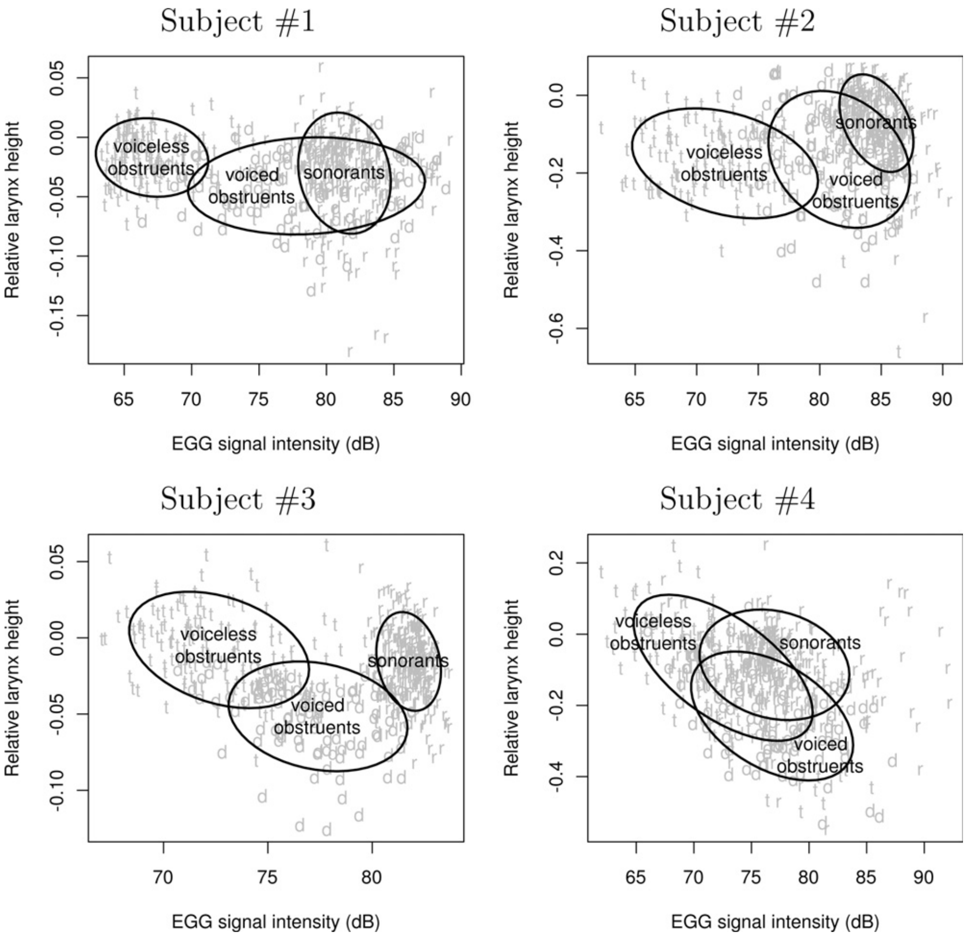

Fig. 5. EGG by subject. Gray letters represent individual tokens, and ellipses represent one standard deviation from the mean for each group of sounds.

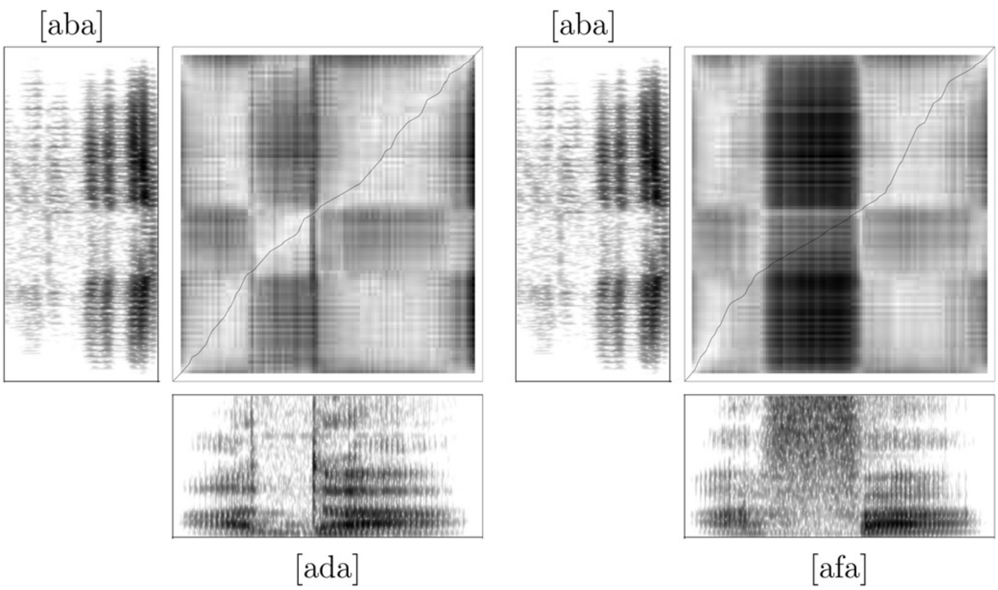

Fig. 6. Dynamic time warping illustration. Left: [aba] vs. [ada], right: [aba] vs. [afa].

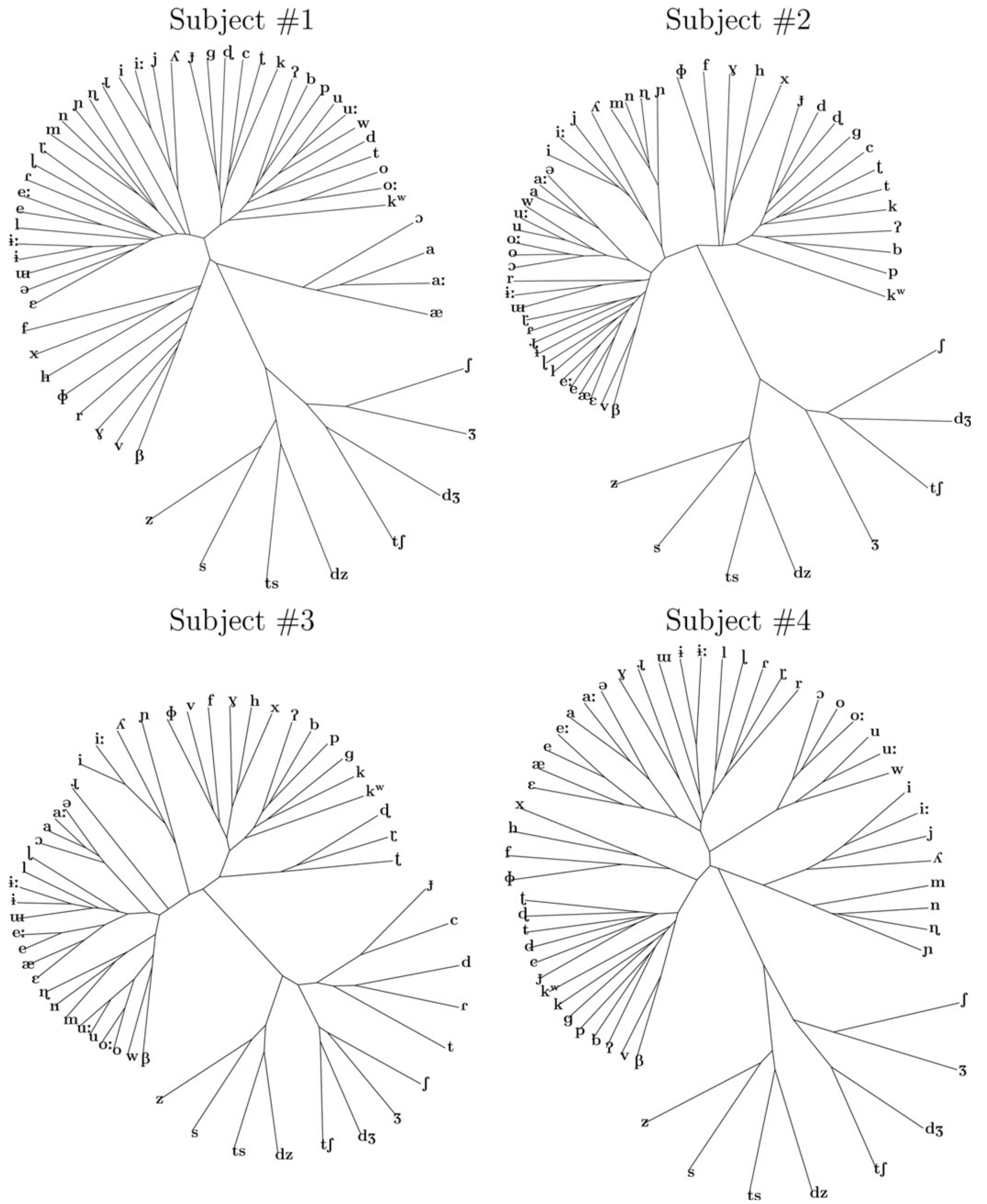

Fig. 7. Neighbor-joining trees for averaged acoustic distances, by subject.

consequence that acoustically similar intervals are compared with each other. For both comparisons, the path avoids comparing vowel and consonant intervals. The acoustic distance for each pair of sounds is the average of the distances at all the points the path passes through. Because the stop consonant intervals are acoustically more similar, the acoustic distance between these tokens of [aba] and [ada] is 2.73, whereas the acoustic distance between these tokens of [aba] and [afa] is 6.87.

To inspect the acoustic data in the four subjects, the distances were averaged across repetitions and contexts, to produce one acoustic distance for each pair of segments for each subject, and neighbor-joining trees (Saitou and Nei, 1987) were created using `neighbor` and `drawtree` from the PHYLIP phylogeny inference package (Felsenstein, 1989), based on each subject's distance matrix. These are shown in Fig. 7. The distance between two segments along the branches of a tree represent the acoustic distance between them.

All four subjects have clusters of alveolar and postalveolar sibilants that are quite distinct from the other segments. For Subject #3, palatal and alveolar stops and the alveolar flap are also clustered with the sibilants. Beyond the sibilant clusters,

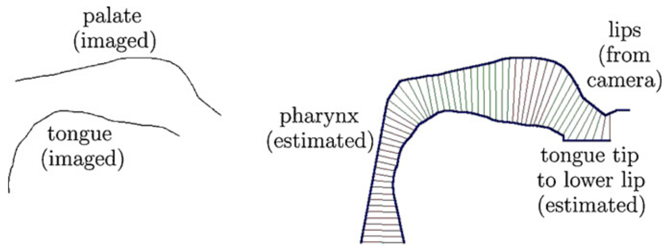

Fig. 8. Estimation and measurement of vocal tract shape from ultrasound images.

all four subjects show similar divisions between obstruents and sonorants. While the large clusters mostly have an interpretation involving manner of articulation, some place-based clusters are visible. Additionally, voiced nonsibilant fricatives such as [ɣ] and [β] group with sonorants in the trees for some subjects. One of the reasons why the clusters are different is that the clustering technique does not allow segments to differ from each other in more than one dimension at a time. In order to interpret the similarity in terms of multiple dimensions, a principal component analysis of the combined acoustic distances will be used in the results section.

2.2.3. Vocal tract shape

To generate measurements of the shape of the vocal tract, video frames were extracted during the target segment intervals and during palate imaging. Because the NTSC video has a rate of 29.97 frames per second, the video frames are about 33 ms apart, so a 150 ms-long speech sound is represented by about five frames. A Praat script was used to read the TextGrid files and generate a shell script that uses `ffmpeg` to extract the video frames as images. The vocal tract shape data from Subject #2 are not included in this analysis, because of a problem aligning the video time code to the TextGrids used to extract the appropriate frames. This problem occurred at the time all of the images were extracted, and it can be fixed in the future. The data for the subject are not lost, but will not be mentioned further in this paper.

Palatoglossatron (Baker, 2005) was used to semi-automatically trace the tongue contour and lips in each tongue image, and to trace the palate contour in each palate image. In all, about 13,000 video frames were traced by a research team. The Palatron algorithm (Mielke et al., 2005) was used to transform all of the traces in order to compensate for head/transducer movement, so that the tongue, palate, and lips could be combined in the same coordinate system. A Python script was used to interpolate between the lips and the anterior extent of the tongue and palate traces, to estimate the location of the pharyngeal wall, and to measure, at 5mm intervals, the distance between the opposing walls of the vocal tract along a line perpendicular to the mid-line of the vocal tract).[9]

The measurement of the image shown above in Fig. 2 is shown in Fig. 8. The curves on the left are the result of tracing the tongue and palate contours in Palatoglossatron. The image on the right shows the same curves after other parts of the vocal tract are added or estimated, and after distances have been measured. The palate location is assumed to be fixed in all of the images. Since the movements of the velum cannot be detected by ultrasound under normal circumstances, it is assumed to be in a neutral position. The location of the pharyngeal wall (relative to the palate) was estimated for each subject, and the position of the upper and lower lips (measured in the image) were placed in the same coordinate system as the tongue and palate by the Palatron algorithm. The lips were connected by straight lines to the most anterior extent of the tongue and palate traces.

After these vocal tract surfaces were assembled and estimated, the cross-vocal-tract distances were measured. First a radial grid (with origin below the lips and to the right of the lower pharynx) was superimposed on the vocal tract, and the midpoints of the segments of these lines that occur between the two vocal tract surfaces were used to find the vocal tract midline.[10] A spline was fit to the set of midpoints, and then the distance between the upper and lower surfaces was measured perpendicular to the midline, at 5mm intervals. This process generates a vector of cross-distances for each image, which represents the shape of the measured/estimated vocal tract.

The collection of vocal tract shape vectors for the subjects were normalized in three ways, to make them comparable. First, the vocal tract was normalized lengthwise by dividing the vocal tract into five zones as follows. (1) The lower pharynx zone is defined as the section from the bottom of the estimated pharynx to the point where the estimated pharyngeal wall and velum meet. (2) The upper pharynx/soft palate zone goes from there to the point where the soft palate hinges against the hard palate. (3) The hard palate zone goes from there to the point where the alveolar ridge is judged to begin. (4) The alveolar ridge zone goes from there to the most anterior point measured in the palate image (usually the base of the upper incisors. (5) The lips zone goes from there to the points marked on the lips in the camera video.

The cross-distances were then resampled so that every image has the same number of measurements per zone (based on the average across subjects), resulting in 63 measurements for the length of each vocal tract. This data reduction has the

[9] The generation of these cross-distances is similar to an automated version of the technique described by Harshman et al. (1977), who made manual measurements based on 13 grid lines superimposed on x-ray tracings.

[10] This technique is similar to Maeda's (1979) technique for finding the midline of the vocal tract using circles.

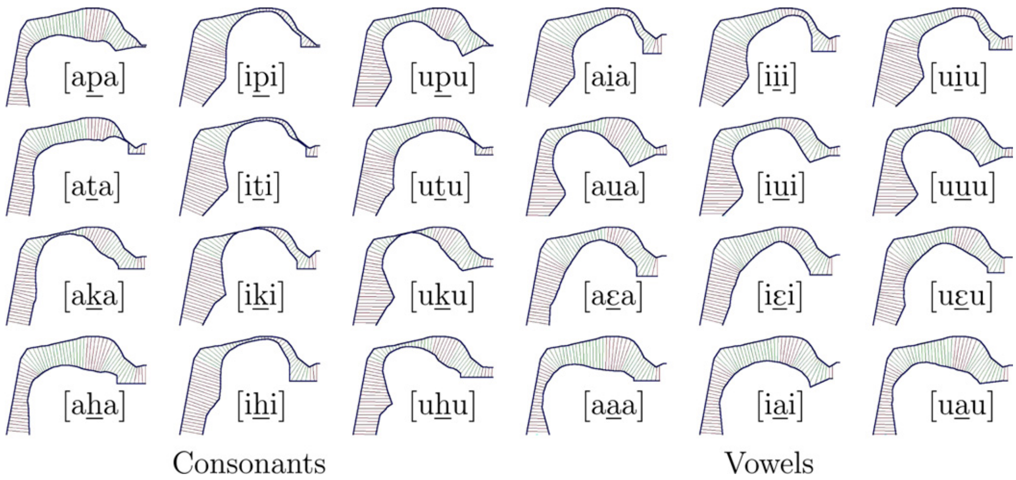

Fig. 9. Center frames from four target consonants and four target vowels produced in three contexts, showing effects of vowel context on tongue shape for target segments.

disadvantage of removing information pertaining to the lengths of particular sections, e.g., about protrusion of the lips, which is phonetically significant. The cross-distances were normalized across speakers by measuring the average distance between the upper and lower surfaces and scaling all the measurements so that the averages are the same for each speaker. Since the target segments differ in duration, the cross-distances were resampled in the time dimension so that each token is represented by five frames. The result of these steps is that each token is represented by a 5×63 matrix, which can be treated as a 305-dimensional vector.

The ultrasound data are consistent with models of speech production in which consonant gestures are superimposed on vowel gestures (Öhman, 1967; Fujimura, 2000), in that the vocal tract shapes for consonant intervals clearly reflect the gestures for the adjacent vowels. Zharkova and Hewlett (2008) recently showed a similar pattern using ultrasound images of the tongue, that the effect of neighboring vowels on the vocal tract shape of consonants is greater than the opposite effect. Fig. 9 illustrates this situation. The vocal tract shapes in the columns, which were produced in the same vowel context, bear more similarity to each other than the vocal tract shapes in the rows, which represent the same target segment. Performing a principal component analysis directly on these tokens causes different consonants produced in similar contexts to cluster together more noticeably than similar consonants produced in different contexts, because the overlapping vowel gestures have a greater effect on the overall tongue shape than most of the consonants do. While intervocalic position is an unnatural context for vowels, the vowels in Fig. 9 show that context vowels had very little effect on tongue shapes during the target vowel intervals.

Ideally, the articulatory information about these sounds would distinguish what is necessary to produce the target sounds from what is necessary for the contexts where they occur. In other words, there is a difference between saying that intervocalic [h] requires three distinct tongue shapes and saying that intervocalic [h] is generally unconcerned with tongue shapes. Algorithmically reweighted principal component analysis (Mielke and Roy, 2009) is a technique, implemented in R, for isolating important parts of the vocal tract for each target sound and ignoring the others. Since the vocal tract shapes will be subject to a principal component analysis, a vocal tract measurement can be made ''invisible'' by setting it to the same value as the mean, across the data set, for that vocal tract position. This is achieved in three steps: (1) A neutral vocal tract shape for each speaker is established by averaging across all tokens of all of the sounds under consideration, at each of the 63 normalized vocal tract positions. (2) For every target segment, the variance across repetitions in all vowel contexts is measured at each of the same 63 locations. (3) Vocal tract positions with variance that exceeds an upper threshold are replaced by the measurement for the neutral vocal tract, and vocal tract positions with variance below a lower threshold are maintained at the measured values. For variances in between the two thresholds, the proportion of original measurements is scaled.[11] Consequently, only relatively invariant vocal tract positions contribute to the differences between different segments. For example, the vocal tract shapes for the [p]s in [apa], [ipi], and [upu] have nothing in common except for the lips. Rather than identifying three distinct tongue shapes as important for articulating [p]s, algorithmic reweighting essentially ignores the non-lip portions of the vocal tract on the basis of their variability across contexts.

Fig. 10 shows the effects of algorithmic reweighting on the segments shown in Fig. 9, as produced by Subject #3. Dotted lines indicate the averages of the original vocal tract shapes in each the three vowel contexts. The dashed line indicates the overall average vocal tract shape. The dark lines are the vocal tract shapes after reweighting. Where there is a lot of variance in the original vocal tract shapes, the original shapes have been replaced with the overall mean. Where there is less variance, the original measurements have been kept. Only the parts that are not set to the mean can distinguish the sounds from other

[11] The thresholds vary along the length of the vocal tract, according to the amount of variance observed at each point.

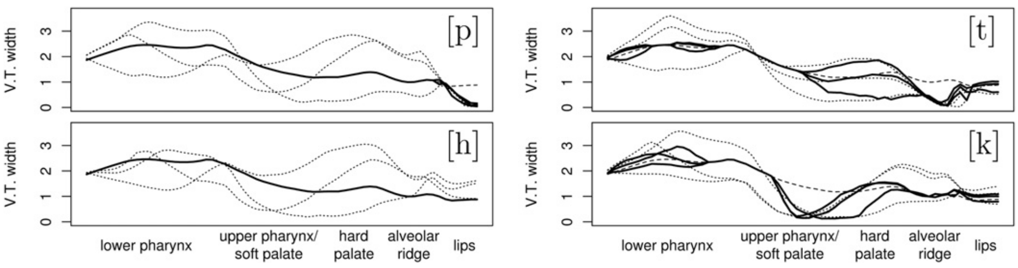

Fig. 10. An illustration of algorithmic reweighting.

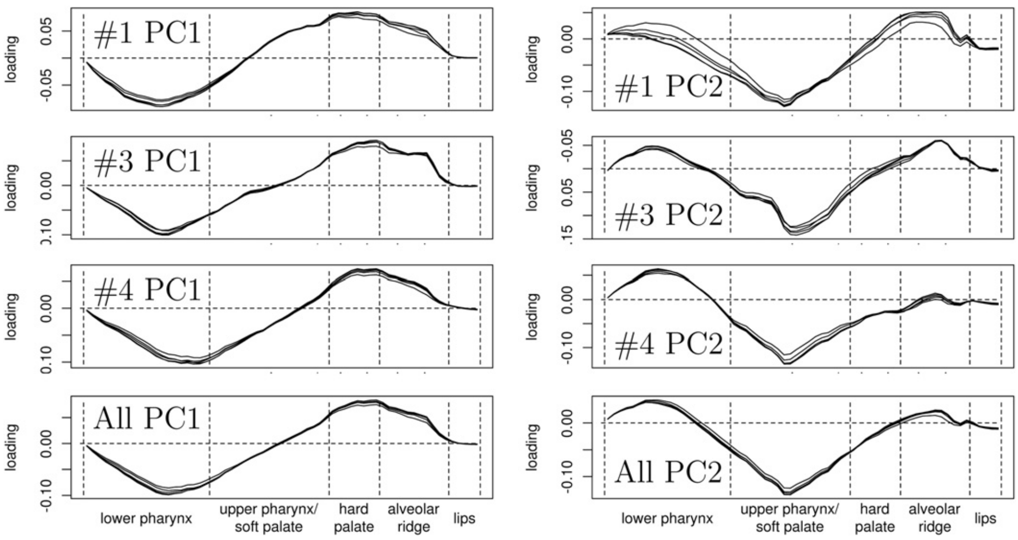

Fig. 11. Vocal tract: principal component loadings by subject.

sounds. In Fig. 10, only the parts of the vocal tract where the transformed vocal tract shape consistently deviate from the dashed line will contribute to its representation in the model. Thus, [p] is represented only by lip position (the only part that is not set to the mean), and [h] is not distinguished from other sounds by anything involving supralaryngeal vocal tract shape. [t] and [k] are distinguished mostly by the measurements in the soft palate and alveolar ridge regions, respectively.

A principal component analysis was performed over all of the reweighted vectors (using `princomp` and `predict` in R), and tokens of each segment were then averaged. Principal component analysis maximizes the variance accounted for by each successive principal component, meaning that the first principal component accounts for as much variance as possible, and the second principal component accounts for as much as possible of the variance that is left, and so on. The analysis was repeated for each of three subjects and for the three subjects combined. Fig. 11 shows the loadings of the first two principal components for each subject. Each plot contains five curves, which correspond to the five time slices of each segment. The value of the curves across the length of the vocal tract indicates its relationship with that principal component. The sign does not matter, and PC2 of Subject #3 has been inverted to match the others. All of the first principal components show an opposition between the pharyngeal region and the palatal region, indicating that for all three subjects, most of the variance between vocal tract shapes (after reweighting) is related to whether the tongue is advanced and raised or retracted and lowered. For all three subjects, the second principal component shows an opposition between constrictions in the velar and alveolar regions.

Fig. 12 shows the segments, averaged across tokens, plotted according to the first two principal components. There is one plot for each of the three subjects whose ultrasound data are available, and one for all three combined. In all four plots, the first principal component (labeled PC 1) can be interpreted as palatal vs. pharyngeal, and the second principal component can be interpreted as coronal vs. dorsal. These correspond directly to the loadings in Fig. 11, i.e., the segments with the lowest values for PC 1 are the ones with the greatest palatal constriction and the least pharyngeal constriction.

The vowels have been connected with lines, to show how the vowel space varies slightly across the four PCAs, but a familiar-looking vowels space is apparent in the PCA based on all three subjects. The sets of alveolar, palatal, and velar consonants are indicated with convex hulls (using `chull` in R). A convex hull is the smallest convex polygon that encloses all of a set of points. They are indicated here to show how compact the place-defined classes are, and that they fall in roughly the same place for all

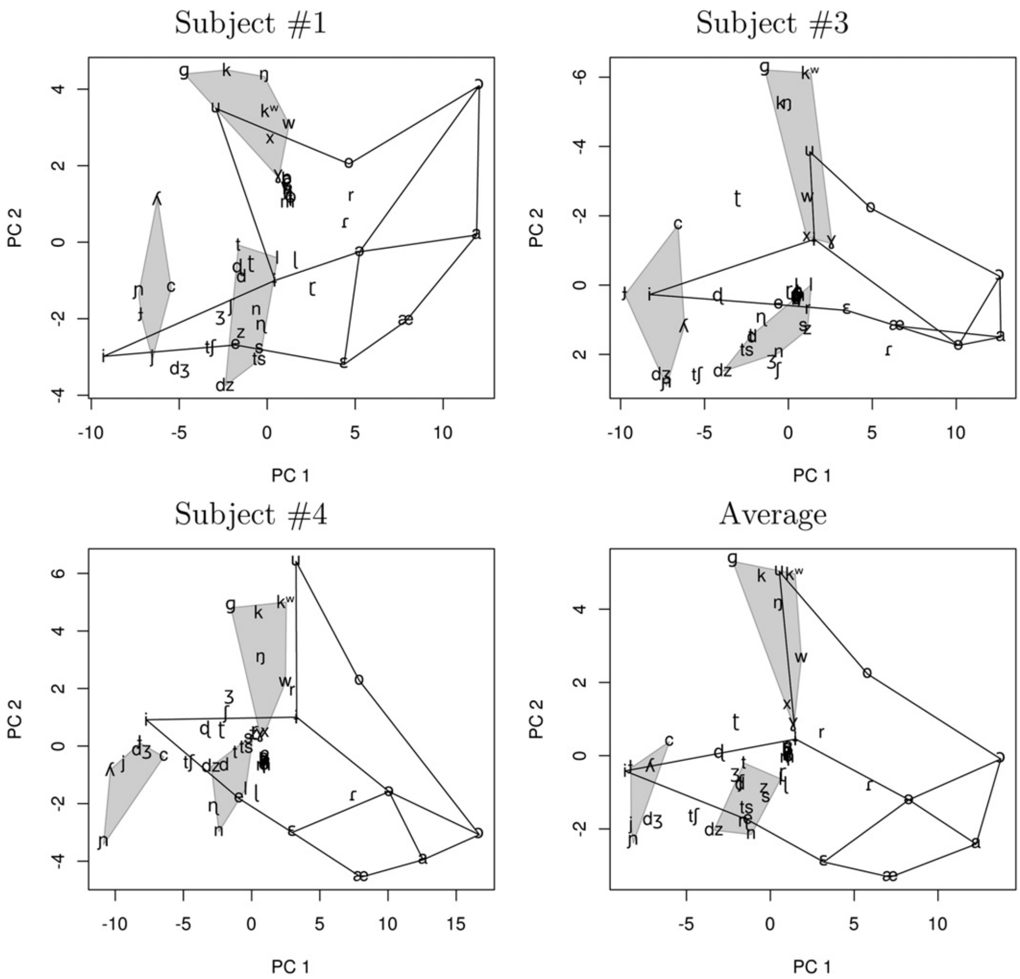

Fig. 12. Vocal tract: first two PCs by subject.

four PCAs.[12] In all four PCAs, the labials and glottals clump together near the origin, because they are not distinguished from any of the other segments by tongue position, and tongue position is all that is important for the first two PCs.

2.3. Data not appearing in this paper

Perceptual data from a speeded AX discrimination task using a subset of the audio recordings have not been analyzed yet, but will ultimately be compared with the acoustic measure of similarity. All of the segments produced by only one speaker (see footnote 3) have been withheld from the analysis for the time being. As mentioned in the last section, all of the ultrasound data from Speaker #2 had to be excluded from the analyses reported here due to a fixable problem involving data extraction.

2.4. Results and evaluation

Having compared the results for each type of data for each subject, the data for each target segment type will now be examined in the context of the phonological behavior of related segments. Summary plots for each kind of phonetic data are displayed in Figs. 13–15. These results will be discussed in the context of a set of phonologically relevant classes, specifically the top 15 most frequent phonologically active classes described by Mielke (2008a:214–217): vowels ($n = 433$), consonants ($n = 180$), nasals ($n = 162$), high vowels ($n = 86$), tense front vowels ($n = 80$), round vowels ($n = 77$), voiceless sounds ($n = 73$), front vowels ($n = 64$), velar obstruents ($n = 62$), nonlow vowels ($n = 57$), back vowels ($n = 57$), obstruents ($n = 53$), glides ($n = 46$), nonhigh vowels ($n = 44$), and voiced obstruents ($n = 43$).[13]

[12] Flaps and the trill are not included in the hulls, because their gestures were too fast to be registered well by ultrasound at 28 Hz.

[13] The frequencies given here are according to the categorization of phonologically active classes using *SPE* features (Chomsky and Halle, 1968), but similar numbers are observed when other feature systems are used for classification.

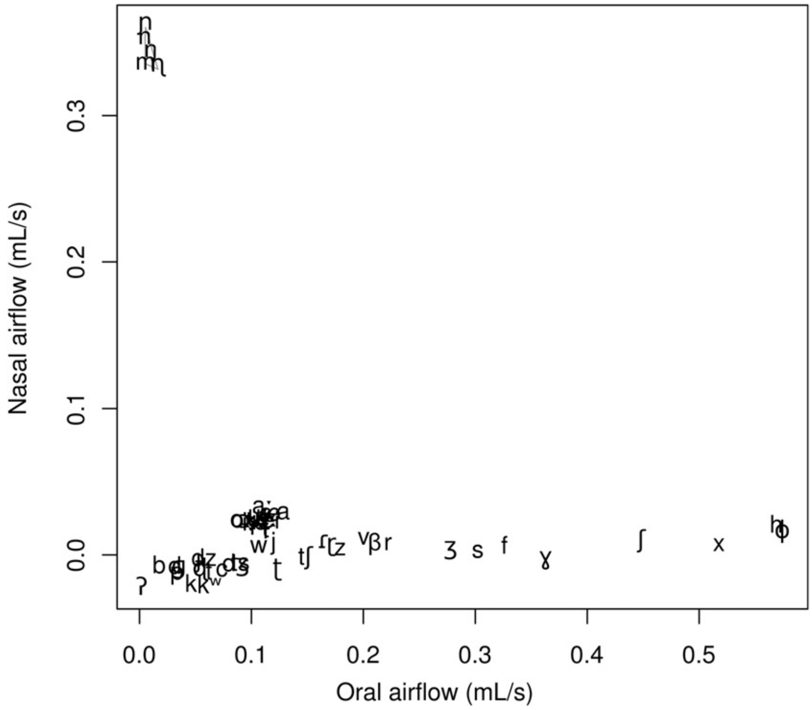

Fig. 13. Airflow measurements.

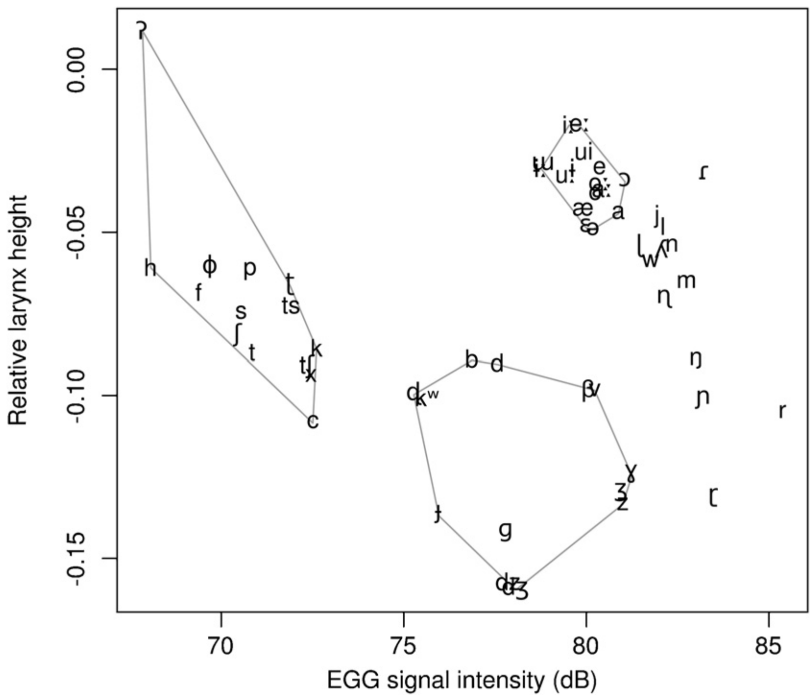

Fig. 14. Electroglottography measurements.

It will be seen that the class of nasals is distinguished by airflow, the classes of obstruents, voiceless segments, and vowels are distinguished by the laryngeal dimensions, the classes of obstruents is mostly distinguished by the second acoustic principal component, and various classes of vowels and classes of consonants defined by place of articulation are distinguished by the first two vocal tract principal components. While this discussion focuses on classes that are definable using the pairs of related phonetic dimensions, other classes can be observed by combining the dimensions in different ways.

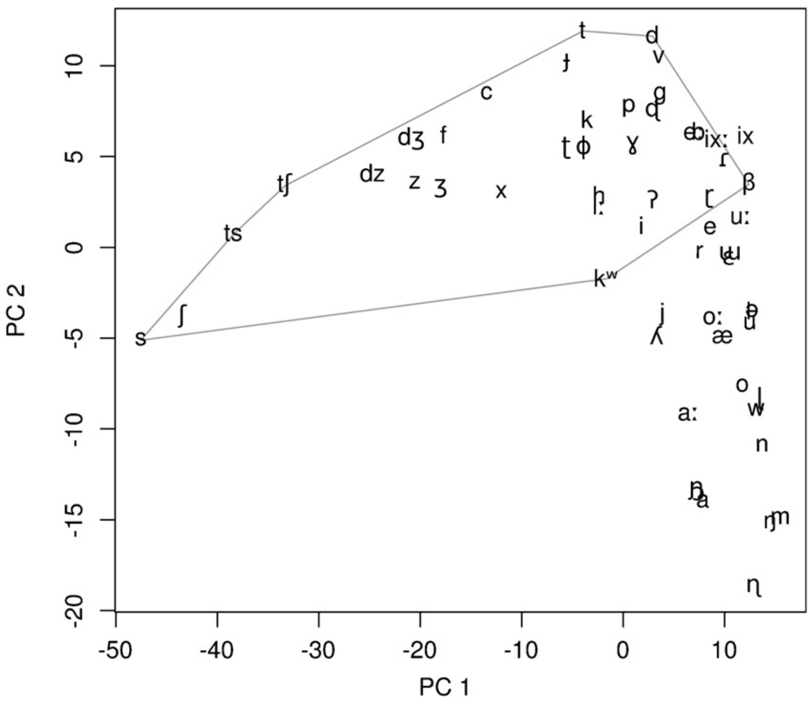

Fig. 15. Acoustic principal components.

Figs. 13–14 show the directly measured articulatory dimensions. In these two figures, the dimensions are determined by the parameters available for measuring the articulatory data. Fig. 13 shows the phonologically relevant classes of nasals (which are distinguished, not surprisingly, by nasal airflow). Fricatives and trills are also distinguished by oral airflow. These are sounds that require a certain level of sustained airflow in order to produce their characteristic frication or trilling. The relative position of the trill and fricatives in this figure is consistent with results reported by Solé (2002), who also discusses the phonological implications of these aerodynamic properties.

Fig. 14 shows that laryngeal dimensions distinguish obstruents from sonorants, and voiceless sounds from voiced obstruents.[14] Notice that the glottal consonants, which have been controversial in their traditional status as [+sonorant] (Chomsky and Halle, 1968) are grouped with the obstruents by the laryngeal measurements. The phonologically important class of vowels is also distinguished. While the classes of sonorants and voiced sounds can also be distinguished, it is worth noting that these classes are much less active phonologically the than the classes of voiceless sounds, obstruents, voiced obstruents, and vowels are. The voiced fricatives, some of which pattern occasionally with sonorants, are all on the sonorant side of the voiced obstruent group.

The next two figures show the segments plotted according to the first two principal components of the acoustic and vocal tract data. In both cases, the PCA was performed across all tokens, and tokens of the same target sound were then averaged within the resulting principal components.

Fig. 15 shows that most of the acoustic variance is related to sibilance: the eight sibilants (which share high-frequency noise) are distinguished from the other segments by the first principal component, which can be interpreted as sibilance (dominance of high-frequency acoustic energy). [f], which has sometimes been treated as a sibilant, is closest to the sibilants. Its voiced counterpart, which sometimes patterns phonologically with sonorants, is nowhere near the sibilants. The obstruents are highlighted in the figure, showing that the second principal component can be interpreted as sonority (i.e., dominance of lower frequency acoustic energy). The class of obstruents overlaps somewhat with the sonorants. Some of the vowels are also near the sonorant-obstruent boundary, probably due to the fact that they were produced intervocalically. While the classes of sibilants and sonorants are easy to visualize, these classes are much less active phonologically than their acoustic similarities might suggest.

The glottals [ʔ h] are right on in the overlap between obstruents and sonorants, which is very intuitive, but difficult to interpret considering that high vowels are also in this overlap. Nevertheless, it is clear that the acoustic properties of glottals put them much closer to the sonorants than the laryngeal properties, seen above in Fig. 14. Differences between vowels and sonorant consonants are more apparent in the higher principal components, which are not shown. Most acoustic correlates of place of articulation are generally too subtle to be apparent in the acoustic dimensions.

[14] [kʷ] has been excluded from all of the classes in Fig. 14. The way it was segmented, it has a voiceless obstruent interval and a sonorant glide-like interval. If EGG intensity had been measured only for the stop interval, [kʷ] likely would be much closer to the other voiceless stops.

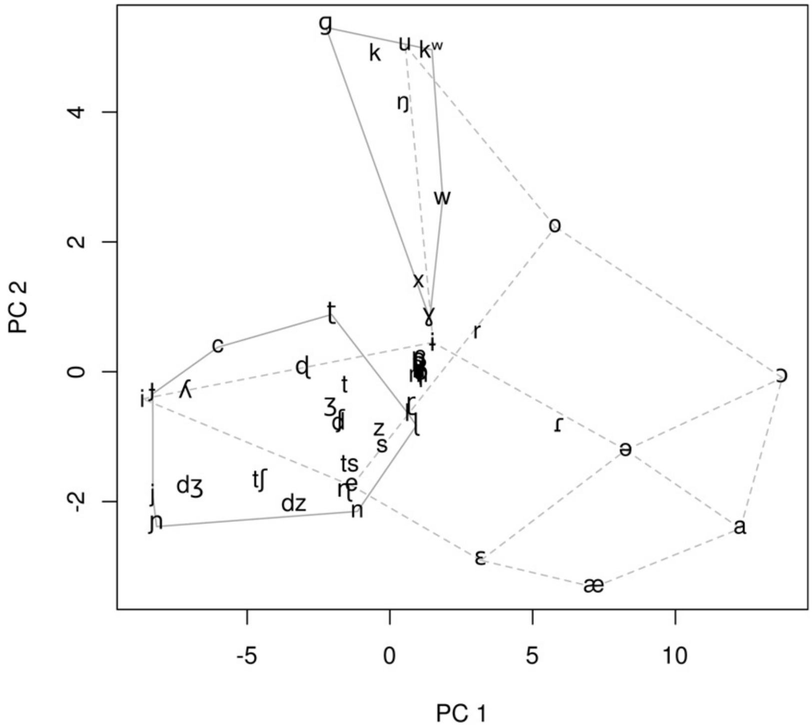

Fig. 16. Vocal tract principal components.

Place of articulation is much more apparent in the vocal tract shape data in Fig. 16. Here the first principal component is related to the opposition between palatal and pharyngeal constrictions. Palatal consonants and vowels are at one extreme, opposite the pharyngeal and lax vowels. The second principal component is correlated with velar vs. coronal constrictions, distinguishing the velar stops and [u] from the other segments. The pile of segments close to the point (0,0) in the figure contains the labials and glottals. Because the first two principal components involve constriction locations which the reweighting algorithm does not deem important for any of these sounds, their cross-distances have been set to near the average for all segments, as seen above in Fig. 10. The labials are distinguished by higher principal components (not pictured) which involve the lips, but the glottals are not distinguished from any of the other sounds by supralaryngeal vocal tract shape. Flaps and trills have been omitted from these classes because their articulations are too fast to be reliably imaged at a scan rate of 28 Hz, and in the figure they cluster near the center.

The phonologically important class of velars is apparent, and the class of coronals (which is outside the top 15 most active classes) is also indicated. The phonologically related vowels are connected by dashed lines, making apparent the phonologically important classes of front vowels, back vowels, high vowels, nonlow vowels, and nonhigh vowels. The round vowels and consonants are also close together according to the first two vocal tract shape principal components. This is probably due to the fact that all of the rounded segments also involve a dorsal tongue constriction. While lip position is relevant for higher principal components of the vocal tract data, the loadings for the lip cross-distances are near zero for the first two components, as seen in Fig. 11. There is a lot of overlap between the consonants and the tense vowels, because they involve similar vocal tract shapes. While consonants and vowels are superimposed on this plot, they are distinguished in other phonetic dimensions.

3. Phonological similarity

To compare the phonetic distances more generally with phonological patterning, a measure of phonological similarity was formulated using the sound patterns reported in P-base (Mielke, 2008a). The database contains 6077 phonologically active classes (classes of segments that serve as the trigger or target for an alternation). Many of these classes are involved in multiple sound patterns within the same language. The availability of these data in electronic form provides an easy way to get a measure of similar phonological patterning that does not require selecting particular frequently active classes as in the last section.

3.1. Methods

Phonological similarity was measured by counting the cooccurrences of pairs of sounds in the same phonologically active classes. Specifically, the phonological similarity between two sounds is defined as the number of phonologically active

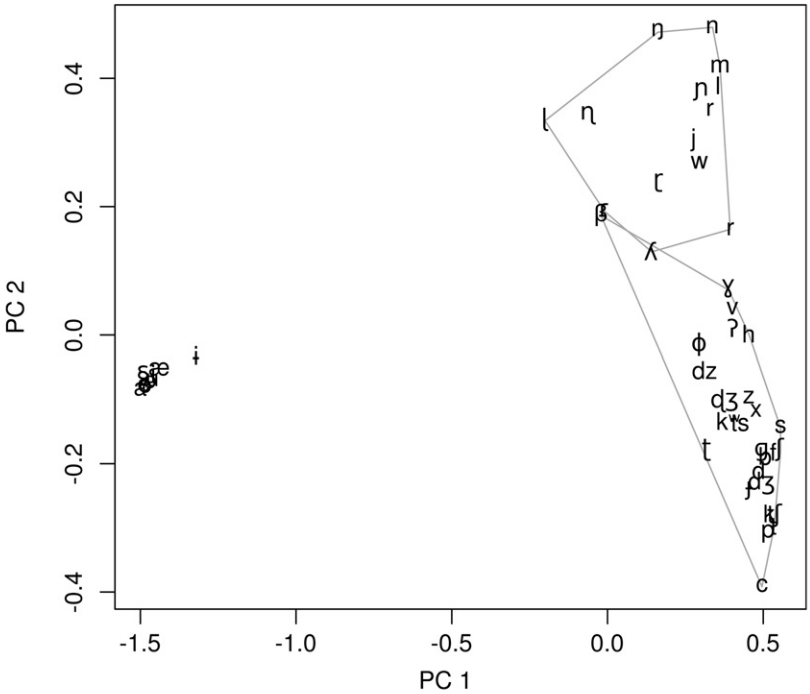

Fig. 17. Phonological similarity (phonologically active classes).

classes involving both segments, divided by the number of phonologically active classes involving either segment (in languages which have both sounds in the inventory). This is the formula used by Frisch (1996) to calculate phonological similarity based on featurally defined classes. Phonological distance is treated as the similarity value subtracted from 1, as in (1).

$$\text{distance}(s_1, s_2) = 1 - \frac{\text{classes with } s_1 \text{ and } s_2}{\text{classes with } s_1 \text{ or } s_2} \quad (1)$$

The distance is 0 for two segments that always pattern together, or 1 for two segments that never pattern together but do occur in the inventories of the same languages. Given the phonologically active classes /p t k s ʃ h/, /t k s ʃ h/, /b d g z/, the phonetic distance between three pairs of segments would be as follows:

$$\text{distance}(/\text{p}/, /\text{t}/) = 1 - \frac{\text{classes with /p/ and /t/}}{\text{classes with /p/ or /t/}} = 1 - \frac{1}{2} = 0.5 \quad (2)$$

$$\text{distance}(/\text{p}/, /\text{b}/) = 1 - \frac{\text{classes with /p/ and /b/}}{\text{classes with /p/ or /b/}} = 1 - \frac{0}{2} = 1 \quad (3)$$

$$\text{distance}(/\text{b}/, /\text{d}/) = 1 - \frac{\text{classes with /b/ and /d/}}{\text{classes with /b/ or /d/}} = 1 - \frac{1}{1} = 0 \quad (4)$$

A principal component analysis was performed for the phonological distances as measured from the classes in P-base.[15]

3.2. Results

Fig. 17 shows the first two principal components of the phonological similarity PCA. The first principal component distinguishes vowels from consonants. The vowels are tightly clustered on the left side of the figure. This is attributable to many sound patterns involving the class of vowels. Since they are clearly separated from the other segments, it would make sense to analyze them separately to look for structure internal to this cluster, but that is not pursued here. The second

[15] Membership in the same phonologically active classes is one way to investigate phonological similarity based on a database of sound patterns. Mielke (2010) and Delaney and Mielke (2011) explore the phonological and phonetic dimensions in which segmental changes occur.

dimension generally distinguishes sonorant consonants from obstruents, as indicated by the convex hulls. As in some of the other figures, the middle ground is held by glottal consonants and voiced fricatives.

4. Discussion of results

Some familiar patterns are seen in the phonetic and phonological similarity reported above, involving the patterning of nonsibilant voiced fricatives, glottals, and the association of particular contrasts with particular types of data. Some apparently strange behavior may also be attributed to methodological issues.

Voiced fricatives often appear with sonorant consonants in the phonetic figures or between obstruents and sonorant consonants. This is especially noticeable in the EGG data (Fig. 14), where they are halfway between sonorant consonants and the rest of the voiced obstruents, and the phonological similarity plot (Fig. 17), where voiced nonsibilant fricatives are near the obstruent-sonorant boundary. The fact that this pattern is seen in both phonetic and phonological data is consistent with the idea that phonetic ambiguity leads to ambivalent phonological patterning across languages (Mielke, 2005a). For instance, Mielke (2005a) reports a few types of segments whose phonological affiliation is variable from language to language, such as lateral liquids and nasals, which are observed in various languages patterning with both continuants and non-continuants. Nasals and lateral liquids are also both potentially ambiguous with respect to continuancy. Further, while fricatives are traditionally classified as obstruents, Mielke (2008a:Ch. 6) reports a number of cases where voiced fricatives pattern with sonorants. This phonological ambivalence is observed in the phonological similarity results in Fig. 17, and in the phonetic similarity results in which they are intermediate between sonorants and other fricatives. It is noteworthy that some of the voiced-voiceless pairs of fricatives pattern more similarly to each other phonetically and phonologically. [f] and [v] are a good example of a pair that differs phonetically and phonologically (compared, for example, to [s] and [z]).

Glottals [ʔ h] are another group of historically ambiguous segments that are observed patterning with sonorants as well as obstruents. See Miller (2011) and Parker (2011) for discussion of the phonological facts. The *SPE* definition that groups them with sonorants defines the opposition in terms of the compatibility of the "vocal tract cavity" with spontaneous voicing (Chomsky and Halle, 1968:302). The only phonetic measure that directly relates to this is the vocal tract shape measure (Fig. 16), where they are located with other sounds that do not require a lingual constriction. Acoustic similarity (Fig. 15) puts them closer to obstruents, laryngeal dimensions (Fig. 14) treat them as extreme obstruents, and phonological patterning (Fig. 17), which presumably is sensitive to a wide range of phonetic dimensions, puts them in the middle.

The association of acoustic similarity with major class and manner distinctions and articulatory similarity with place distinctions is consistent with Lin and Mielke's (2007) proposal that manner distinctions may be learned more easily from acoustic data while place distinctions may be learned more easily from articulatory data (e.g., in tactile feedback during babbling).

It is quite notable that vowels and consonants are massively distinct according to phonological similarity (Fig. 17), and not nearly so different according to any of the phonetic measures reported here. This mismatch must be due at least in part to the fact that consonants and vowels generally occur in different prosodic positions (by definition), but they were all produced intervocalically in the data reported here. This means that the distinctness of consonants and vowels is due mostly to factors other than their intrinsic phonetic properties. A better understanding of the phonetic basis for this phonological difference needs to take into account where these sounds occur in languages. Goldsmith and Xanthos (2009) propose a technique, using Hidden Markov Models, to induce the distinction between consonants and vowels on the basis of the different transitional properties within and between members of the vowel and consonant classes (the fact that consonants tend to be followed by vowels and vowels tend to be followed by consonants, even in languages like English that allow large consonant clusters). Mielke (in preparation) uses this technique in conjunction with the phonetic dimensions described here to model the induction of phonological features from phonetic and phonological data.

The strange patterning of the trill and flaps [r ɾ ɽ] in the vocal tract data are likely due to the fact that they involve closures shorter than the 33ms interval between the video frames. They are difficult to observe without high-speed ultrasound techniques (Miller and Finch, 2011). Another factor may be that these sounds do not correspond to native phonemes for the speakers who produced them. The trill requires a precise aerodynamic configuration that most of the other sounds do not require (Solé, 2002), and it is possible that the speakers (who as linguists have been trained to produce non-native sounds) are able to produce sounds that are acceptable auditorily, but are distinguishable from natively produced sounds by the methods used here.

5. General discussion

At the beginning, the question was posed *Why is the class of obstruents* (or any other recurrent class) *active in many different languages?* This article has illustrated a few different ways in which obstruents and other phonologically important classes of sounds are phonetically distinct from other sounds. It has also been seen that the phonetic and phonological distinctions are gradient in similar ways. Thus there is an apparent phonetic basis for the phonological observations, and the gradience of this relationship suggests that the relationship is not mediated by categorical features.

The phonological similarity measure provides a very rough picture of the phonological relationships between sounds. The data that Fig. 17 is based on include a wide range of sound patterns, many of them involving classes that were looked at in more detail in connection with particular phonetic dimensions. While there is a clear connection between phonetic similarity and phonological activity, it is worthwhile to reflect further on the nature of the relationship. A more sophisticated answer to the question posed above would provide a specific mechanism for how a phonetic fact becomes incorporated into a phonological grammar. Data about the phonological patterns that involve the phonetic distinction provide some insight: [−sonorant] is needed about three times as often [+sonorant] in P-base sound patterns (Mielke, in press), and of 81 instances of the class [−sonorant] in the database, about half involve voicing or devoicing patterns. An important mechanism for the involvement of the sonorant-obstruent opposition in phonology seems to be the tendency of adjacent obstruents to affect each other's voicing specification (Stevens et al., 1992; Myers, 2010), apparently due to the difficulty of producing adjacent voiceless and non-spontaneously voiced intervals. In this case, a very specific notion of phonetic similarity is related to the mechanism of change (a particular phonetic effect involving the vocal folds), and broad measures like acoustic similarity are too blunt or too indirectly related to capture it.

In light of observations like this, a more direct way of accounting for recurrent phonologically active classes would involve comparing the record of sound changes with the record of synchronic sound patterns, something which will be possible upon completion of the Handbook of Phonological Change (Blevins and Mielke, in preparation), a database of regular sound changes.[16] For phonetically natural classes of sounds that are active because of shared participation in a phonetically motivated sound change, the phonetic similarity of the sounds should be tied directly to the phonetic effect that gave rise to the pattern. Some of the phonologically important groups investigated in the last section appear to show this relationship. Obstruents are distinguished from other sounds by the laryngeal dimensions, and a large number of the phonological patterns applying to obstruents involve voicing alternations, which is directly related to the larynx. Obstruents are also mostly distinct from sonorants in acoustic properties as well, but this seems less likely to be the cause of voicing alternations involving obstruents, although the acoustic similarity may be more closely related to pattern that are sensitive to sonority. Many of the phonological patterns involving vowel tenseness and backness involve assimilation patterns, where the tongue position for one vowel affects the tongue position of another. This is consistent with the observation that these classes of sounds are grouped by the dimensions based on vocal tract shape.

Another role for phonetic similarity may be generalization in the acquisition or spread of sound patterns (Mielke, 2005b, 2009; Dinkin, 2006, *inter alia*). Mielke (2005b) argued that a tendency to associate phonological behavior with phonetically similar segments could result in a bias toward phonetically natural classes. In a simulation, phonetically unnatural classes like /p m/ in a language with a /b/ were occasionally mislearned as /p b m/, while the opposite change was not observed. Something like this could account for sound patterns involving phonetically natural classes that do not appear to be related to the original phonetic basis for the pattern. The extent to which generalization is needed to account for natural classes depends on the magnitude of the residue after sound patterns with direct phonetic explanations have been accounted for.

If the role of the more general notion of phonetic similarity is limited to generalization, then it is reasonable to suppose that it would involve phonetic dimensions that are salient to the language user. The salience of articulatory similarity, especially the vocal tract measure in this project, is suspect. Future studies may show that other phonetic dimensions are dominant in generalization, or that the role of articulatory similarity is better described in terms of tactile feedback or muscular activity.

Whether the phonological activity of a group of phonetically similar sounds is due to a sound change or to phonetically based generalization, the connections between phonetic and phonological similarity support the idea that any phonetic property is potentially the basis for a phonological pattern and that some classes are very active phonologically because the phonetic property they share is related to a particular phonetic effect that is the basis for a frequent sound change (Blevins, 2004; Mielke, 2008a).

6. Conclusions

This article has described a phonetic similarity metric that is based on phonetic data, and compared its results with measures of phonological similarity. Some of the connections between phonetic and phonological similarity are transparently related to particular phonetic effects, and others seem to be based on a more vague notion of similarity. The correspondences, both categorical and gradient, between phonetic and typological patterns suggest a relationship whose explanation does not require mediation by abstract primitives such as distinctive features. Whether distinctive features play an active role in the mental representation of phonological patterns is a different question that has not been addressed here.

References

Baker, A., 2005. Palatoglossatron 1.0. University of Arizona, Tucson, Arizona. http://dingo.sbs.arizona.edu/~apilab/pdfs/pgman.pdf.
Baker, A., 2006. Crackquirer (software). University of Arizona, Tucson, Arizona. http://dingo.sbs.arizona.edu/~apilab/pmwiki/pmwiki.php?n=Procedures.ConvertingFlowAndPressureDataToPraatFormat.
Blevins, J., 2004. Evolutionary Phonology. Cambridge University Press, Cambridge.
Blevins, J., Mielke, J. The handbook of phonological change, in preparation.
Boersma, P., Weenink, D., 2007. Praat: Doing Phonetics by Computer [Computer Program]. http://www.praat.org.
Brown, C.H., Holman, E.W., Wichmann, S. Sound correspondences in the world's languages, uCLA/NIU/MPI-EVA ms, submitted for publication.
Childers, D.G., Krishnamurthy, A.K., 1985. A critical review of electroglottography. Critical Reviews in Biomedical Engineering 12 (2), 131–161.

[16] Closely related to this, Brown et al. (submitted for publication) examine sound correspondences in a large number of languages, and compare the correspondences with some of the phonetic similarity data reported in this paper.

Chomsky, N., Halle, M., 1968. The Sound Pattern of English. Harper & Row, New York.
Clements, G., 1985. The geometry of phonological features. Phonology Yearbook 2, 225–252.
Clements, G., Hume, E.V., 1995. The internal organization of speech sounds. In: Goldsmith, J. (Ed.), The Handbook of Phonological Theory. Blackwell, Cambridge, MA, pp. 245–306.
Delaney, C., Mielke, J., 2011. Acoustic and articulatory dimensions in sound patterns. Poster presented at the Undergraduate Research Opportunity Program Symposium, University of Ottawa.
Dinkin, A., 2006. Unnatural classes and phonological generalization in dialect formation. Talk presented at NWAV 35, Columbus, OH.
Felsenstein, J., 1989. PHYLIP – phylogeny inference package (version 3.2). Cladistics 5, 164–166.
Frisch, S., 1996. Similarity and frequency in phonology. Ph.D. Thesis, Northwestern University.
Frisch, S., Pierrehumbert, J., Broe, M., 2004. Similarity avoidance and the OCP. Natural Language and Linguistic Theory 22, 179–228.
Fujimura, O., 2000. The C/D model and prosodic control of articulatory behavior. Phonetica 57, 128–138.
Goldsmith, J., Xanthos, A., 2009. Learning phonological categories. Language 85 (1), 4–38.
Harshman, R., Ladefoged, P., Goldstein, L., 1977. Factor analysis of tongue shapes. Journal of the Acoustical Society of America 62 (3), 693–707.
Holmes, J., Holmes, W., 2001. Speech Synthesis and Recognition, 2nd edition. Taylor & Francis, New York.
Kondrak, G., 2003. Phonetic alignment and similarity. Computers and the Humanities 37 (3), 273–291.
Lin, Y., Mielke, J., 2007. Discovering place and manner features: what can be learned from acoustic and articulatory data? In: Proceedings of 31st Penn Linguistics Colloquium.
Maddieson, I., Precoda, K., 1990. Updating UPSID. In: UCLA Working Papers in Phonetics, vol. 74. pp. 104–111.
Maeda, S., 1979. Une modèle articulatoire de la langue avec des composantes linéares. In: 10émes JEP, GALF. pp. 152–164.
Mielke, J., 2005a. Ambivalence and ambiguity in laterals and nasals. Phonology 22 (2), 169–203.
Mielke, J., 2005b. Modeling distinctive feature emergence. In: WCCFL 24.
Mielke, J., 2008a. The Emergence of Distinctive Features. Oxford University Press, Oxford.
Mielke, J., 2008b. Segment inventories. Blackwell Language and Linguistics Compass 32 (3), 700–718.
Mielke, J., 2009. Accepting unlawful variation and unnatural classes: a model of phonological generalization. In: Féry, C., van de Vijver, R., Kügler, F. (Eds.), Variation and Gradience in Phonetics and Phonology. Mouton de Gruyter, Berlin.
Mielke, J., 2010. The frequency of segmental alternations: implications for sociolinguistic variation. Poster presented at NWAV 39, San Antonio.
Mielke, J. A feature-learning algorithm. Cahiers Linguistiques d'Ottawa, in preparation.
Mielke, J. Phonologization and the typology of feature behavior. In: Yu, A. (Ed.), Origins of Sound Change: Approaches to Phonologization. Oxford University Press, OxfordMielke, in press.
Mielke, J., Baker, A., Archangeli, D., Racy, S., 2005. Palatron: a technique for aligning ultrasound images of the tongue and palate. In: Jackson, S., Siddiqi, D. (Eds.), Coyote Papers, vol. 14.
Mielke, J., Roy, J., 2009. Measuring articulatory similarity with algorithmically reweighted principal component analysis (abstract). Journal of the Acoustical Society of America 125, 2499.
Miller, A., Finch, K., 2011. Corrected high frame rate anchored ultrasound with software alignment. Journal of Speech, Language, and Hearing Research 54, 471–486.
Miller, B., 2011. Feature patterns: their sources and status in grammar and reconstruction. Ph.D. Thesis, Trinity College, Dublin.
Myers, S., 2010. Regressive voicing assimilation: production and perception studies. Journal of the International Phonetic Association 40 (2), 163–179.
Ohala, J.J., 1981. The listener as a source of sound change. In: Masek, C., Hendrik, R., Miller, M. (Eds.), CLS 17: Papers from the Parasession on Language and Behavior. CLS, Chicago, pp. 178–203.
Ohala, J.J., 1983. The origin of sound patterns in vocal tract constraints. In: MacNeilage (Eds.), The Production of Speech. Springer, New York, pp. 189–216.
Ohala, J.J., 1993. Sound change as nature's speech perception experiment. Speech Communication 13, 155–161.
Ohala, J.J., 1994. Towards a universal, phonetically-based, theory of vowel harmony. In: Bamman, D., Magnitskaia, T., Zaller, C. (Eds.), 3rd International Congress on Spoken Language Processing, Yokohama, pp. 491–494.
Öhman, S.E.G., 1967. Numerical model of coarticulation. Journal of the Acoustical Society of America 41 (2), 310–320.
Parker, S., 2011. Sonority. In: van Oostendorp, M., Ewen, C., Hume, E., Rice, K. (Eds.), Companion to Phonology. Wiley–Blackwell.
Rothenberg, M., 1979. Some relations between glottal air flow and vocal fold contact area. In: Conference on Assessment of Vocal Pathology (ASHA Report, vol. 11), American Speech and Hearing Association, Rockville, MD, pp. 88–96.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4 (4), 406–425.
Simpson, A.P., 1999. Fundamental problems in comparative phonetics and phonology: does UPSID help to solve them? In: Proceedings of ICPhS XIV. pp. 349–352.
Solé, M.-J., 2002. Aerodynamic characteristics of trills and phonological patterning. Journal of Phonetics 30, 655–688.
Sproat, R., Fujimura, O., 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. Journal of Phonetics 21, 291–311.
Stevens, K.N., Blumstein, S.E., Glicksman, L., Burton, M., Kurowski, K., 1992. Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. Journal of the Acoustical Society of America 91 (5), 2979–3000.
Zharkova, N., Hewlett, N., 2008. Measuring lingual coarticulation from midsagittal tongue contours: description and example calculations using English /t/ and /a/. Journal of Phonetics 37 (2), 248–256.