Super-imposing maxillary and palatal locations for electroarticulometry: A SIMPLE method[a]

Wei-rong Chen,[1,b] Yueh-chin Chang,[1] Catherine T. Best,[2] and Donald Derrick[3]

[1]*Graduate Institute of Linguistics, National Tsing Hua University, Number 101, Section 2 Guangfu Road, Hsinchu 30013, Taiwan*
[2]*MARCS Institute and School of Humanities and Communication Arts, University of Western Sydney, Locked Bag 1797, Penrith, New South Wales 2751, Australia*
[3]*New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, Private Bag 4800, Christchurch, New Zealand*
p@weirongchen.com, ycchang@mx.nthu.edu.tw, c.best@uws.edu.au, donald.derrick@canterbury.ac.nz

Abstract: This study proposes a method of superimposing a physical palatal profile, extracted from a speaker's maxillary impression, onto real-time mid-sagittal articulatory data. A palatal/dental profile is first obtained by three-dimensional–scanning the maxillary impression of the speaker. Then a high resolution mid-sagittal palatal line, extracted from the profile, is sub-divided into articulatory zones and superimposed, by Iterative Closest Point algorithm, onto reconstructed palatal traces in electromagnetic articulometric (EMA) data. Evaluations were carried out by comparing consonant targets elicited by EMA with the proposed method and by static palatography. The proposed method yields accurate results, as supported by palatography.
© 2015 Acoustical Society of America
[AL]

1. Introduction

This paper presents a method of extracting the maxillary profile with details of articulatory landmarks (henceforth palatal profile) from a dental maxillary impression of a speaker and then superimposing the profile onto mid-sagittal articulometric data of the speaker, providing a high resolution, standardized, and anatomically landmarked palatal trace for real-time articulometry.

Constriction location (or place of articulation) in phonetics is commonly defined as the place of main obstruction of the air stream in the vocal tract and is described in terms of a point on the upper/passive articulators (orthodontic and maxillary structures). Electro-articulometry techniques allow for dynamic tracking of articulatory constrictions/movements along with time-aligned acoustic recordings.

But a commonly acknowledged limitation of all the current techniques of real-time articulometry is that none of them provides a high resolution profiling of the upper articulators or palatal trace with clear anatomical landmarks, and there is therefore no ideal subject-normalized method of quantifying constriction location. Fitzpatrick and Chasaide[1] trace electro-palatograph (EPG) images for the hard palate, and x-ray images for the soft-palate, to construct the shape of vocal tract for electromagnetic articulometry (EMA) representations (see Jones and Hardcastle[2] or Fitzpatrick[3] for details). They imported the traces into EMA data by aligning the coordinates of referential teeth. Geng and Mooshammer[4] also used a similar method to manually import the coordinates of an EPG palate into EMA data by visual adjustments. However, the construction of an EPG palate for each participant is expensive and the technique requires many manual measurements and adjustments. Most recently, Yunusova *et al.*[5] described a thin plate spline (TPS) technique for reconstructing the whole palate surface from *in vivo* EMA tracings and found a good correspondence with a plaster palatal impression (3D-scanned as the "gold standard" reference) of the same subject. Neufeld and Anghelescu[6] and Neufeld and van Lieshout[7] also demonstrated the use of a very similar method (*in vivo* 3D EMA tracing of the whole

palate) for redefining articulator positions as palate-relative coordinates (i.e., constriction location, CL) and noted that palate-relative coordinates may be more adequate in capturing underlying phonological structure than the traditional Cartesian coordinates of the lingual articulators. While this technique is cost effective and can successfully provide anatomical information of palate morphology, the anatomical invariances (for cross-subject studies) employed in this technique rely solely on the dentition (i.e., upper incisors, first molars). Limited by the sampling errors of EMA tracing, the fine distinctions within the critical alveolar region, (e.g., dentialveolar vs alveolar vs postalveolar) cannot be clearly demarcated by this technique. Thus, in this study, we extended the same objectives and revisited the benefits of using a dental/palatal impression. We created a Super-Imposing Maxillary and Palatal Locations for Electro-articulometry (SIMPLE) method to retrieve holistic maxillary information from a dental impression and import it into EMA data using an automatic shape matching algorithm, and evaluated it by comparing the results of this method with those of static palatography.

2. Methods

The proposed method consists of four steps, detailed in the following subsections.

2.1 Maxillary impression and dental cast

The first step is to take an alginate impression of the speaker's maxillary contour. A cast is then made by pouring plaster paste into the negative maxillary impression [Fig. 1(a); see Ladefoged[8] or Anderson[9] for details on making such plaster casts].

2.2 Digitizing the maxillary dental cast

In order to extract the mid-sagittal palatal outline from the dental cast (hereafter PLC), the cast needs to be digitized first. We used a *3shape TRIOS* active 3D scanner to retrieve the 3D coordinates of the whole cast. 3D-scanning returns a standard STL file containing a point cloud of [x,y,z] coordinates for the cast. We manually rotated that 3D object to align its occlusal plane to the z-axis, consistent with the EMA coordinate system [Fig. 1(c)]. The PLC was then defined as the data points extracted from the mid-sagittal suture line [Figs. 1(c) and 1(d)].

2.3 Articulatory landmarks

A number of articulatory landmarks were digitally marked on the 3D model, following Lee's approach,[10] which divides the whole upper articulatory space into six zones [Figs. 1(d) and 1(e)]: (1) Dental—from the apex of the front incisors to the tip of the gum between the two front incisors; (2) Dentialveolar—from the end of zone 1 to the root base of the front incisors; (3) Alveolar—from the end of zone 2 to the most convex point in the alveolar ridge; (4) Postalveolar—from the end of zone 3 to the end of the alveolar ridge; (5) Prepalatal—from the end of zone 4 to the point on the hard palate at which the slope changes from upward-backward to backward (or backward-downward); and (6) Palatal—from the end of zone 5 to the boundary between the hard palate and the soft palate (velum).

2.4 Shape matching

Last, to import those articulatory landmarks into articulatory data (EMA data for demonstration in this study), we performed shape matching between the PLC and the EMA palatal trace by adapting the Iterative Closest Point algorithm (ICP),[11] as follows: (1) for each point in the PLC data, find the closest point in the EMA palatal trace data; (2) estimate a rotation and translation for the PLC data, using a mean squared error function to provide the best alignment for each EMA source point to its PLC match; (3) apply the transform. (4) repeat 1–3 until the mean squared error no longer improves appreciably (converges at the minimum). Mathematically, let A be the matrix of PLC data points [Fig. 1(d)], and B the palatal outline traced by EMA [Fig. 1(f)]. Function $D(X, Y)$ represents the mean squared error between any two sets of data points X and Y. A proper rigid transformation of X can be formulated as $T(X) = R * X + t$, where R represents a rotation (orientation-preserving orthogonal transformation), and t a translation of the origin. The final output of the ICP algorithm (transformed A, denoted as A') is formulated as: $A' = arg\,min_{T(A)}\,D(T(A), B)$ [Fig. 1(g)]. The implementations of the ICP algorithm in this study were carried out by using the *fmincon* function in MATLAB,[12] which enables iterative searching for minimums of an input function under user-defined constraints. Before entering into the algorithm, PLC was first oriented in roughly the same direction as the EMA palatal line in the previous stage [the editing stage in Fig. 1(c)], and then translated to share the same origin as the EMA palatal line. As the two shapes were pre-aligned, the final
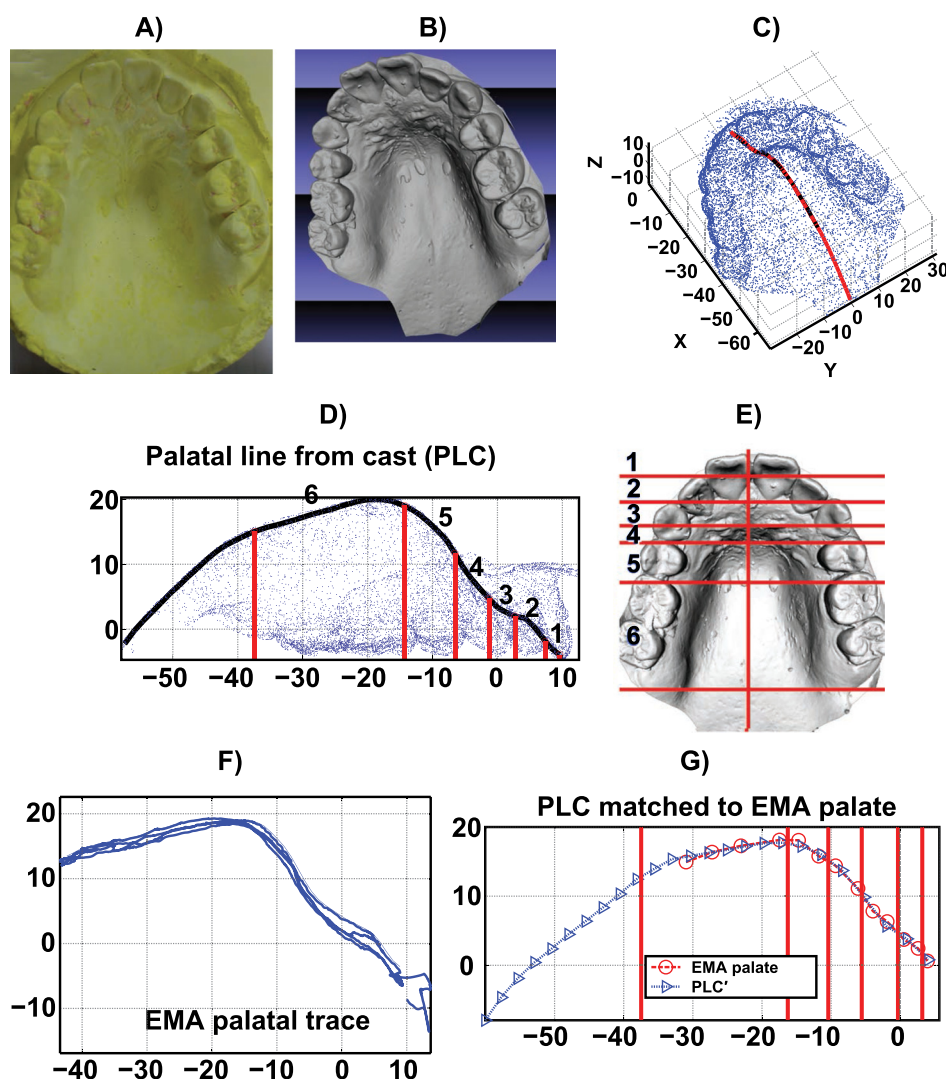
Fig. 1. (Color online) Ordered stages of the SIMPLE method (scales are in millimeters): (A) plaster dental/palatal cast, (B) 3D-scanned model of the cast, (C) pre-aligned model in MATLAB, (D-E) palatal line extracted from the cast (PLC) and divided into six zones, (F) palatal trace obtained by the means of EMA, (G) PLC matched to EMA palatal trace (PLC′) by ICP algorithm. Reference names of articulatory zones in (D) and (E) are (1) dental, (2) dentialveolar, (3) alveolar, (4) postalveolar, (5) prepalatal, and (6) palatal.

ICP transformation was constrained to be within $-90° \sim 90°$ of rotation. This reduces the amount of computation and allows the algorithm to converge at a local minimum, maximizing efficiency and reducing the risk of undesired solutions.

3. Evaluations

To evaluate the reliability of the proposed SIMPLE method, we carried out EMA and static palatographic experiments on the same speakers. Despite possible drawbacks of hyperarticulation or ceiling effects,[13] palatography has been used for over 50 years and is still considered a robust measure of places of articulation. If the consonant constrictions acquired by the SIMPLE method in the EMA data correspond to those of palatography for the same speaker, the reliability of SIMPLE is supported.

3.1 Experiments

Four (2 females) native speakers of Taiwan Mandarin (TM) around 30 years of age served as the subjects in both EMA and palatography experiments. The stimuli were CV syllables in TM consisting of two coronal affricates, /ts/ and /tʂ/, followed by an alveolar approximant (Mandarin's *apical vowel*), which is commonly known to have essentially the same tongue position with the preceding consonant.[14] The stimuli were randomized with fillers and repeated three times. For the EMA experiments, the Carstens AG500 was employed, and the sensors were attached mid-sagittally to (1) tongue dorsum (TD)—the point on the tongue as far back as the subject can tolerate; (2) tongue tip (TT)—about 0.5 cm from the apex of the tongue; (3) tongue body

(TB)—roughly at the midpoint between TD and TT; (4) upper lip (UL); and (5) lower lip (LL). The temporal landmark for extracting the articulator positions for consonant targets in the EMA data was defined as the time the primary articulator achieved maximal constriction (per Gafos[15] and Shaw *et al.*[16]). The coordinate spaces for EMA data were subject-normalized by their occlusal bite planes (*x*-dimension: front-back; *y*-dimension: high-low). Tongue tip constriction location (TTCL) for consonants was defined as the closest point on the palate to the TT sensor, and was expressed as the percentage in the total length of the palatal curve from zone 1 to zone 6, defined as

$$\text{TTCL}\% = (L_C/L_T) \times 100\%, \tag{1}$$

where L_C is the curve length from the beginning of zone 1 to TTCL, and L_T the total curve length of the six zones, as well as the percentage within the articulatory zone where TTCL is located, given by

$$\text{TTCL@} = N + \frac{L_I}{L_N}, \tag{2}$$

where N denotes the number of the zone where TTCL is located, L_I the curve length from the beginning of zone N to TTCL, and L_N the curve length of zone N. The procedures of palatography follow Dart.[17] A dental mirror (Polaroid MACRO 5) and CANON EOS M with EF-M 22 mm STM lens were used for taking palatograms.

3.2 Results

Comparison of constriction locations for /ts/ (upper panel) and /tʂ/ (lower panel) acquired by the SIMPLE method (lower row in each panel) and by palatography (upper row in each panel) for the four subjects are shown in Fig. 2 (results in Table 1). For example, for subject F01 in the upper panel of Fig. 2, the most anterior point of constriction for /ts/ in palatography extends to the front end of zone 2 (dentialveolar), whereas in the EMA data, the averaged position of TT sensor is also close to zone 2. Note that there can be a 0.5 cm error between the most anterior point in palatography and TT position, as the TT sensor was attached approximately 0.5 cm posterior to the apex of the tongue. The palatographic data presented in Fig. 2 were the best of three repetitions for each token, and EMA data are all the three repetitions for each token. Values of *TTCL*% and *TTCL@* were averaged across repetitions within each subject.

　　　Within-subject comparisons show that the constriction zones for Taiwan Mandarin /ts/ and /tʂ/ determined by SIMPLE are consistent with those of palatography, except for M03's /tʂ/, which had an alveolar constriction location in palatography, and in one EMA trial only—the TT constriction was dentialveolar for the other EMA trial.
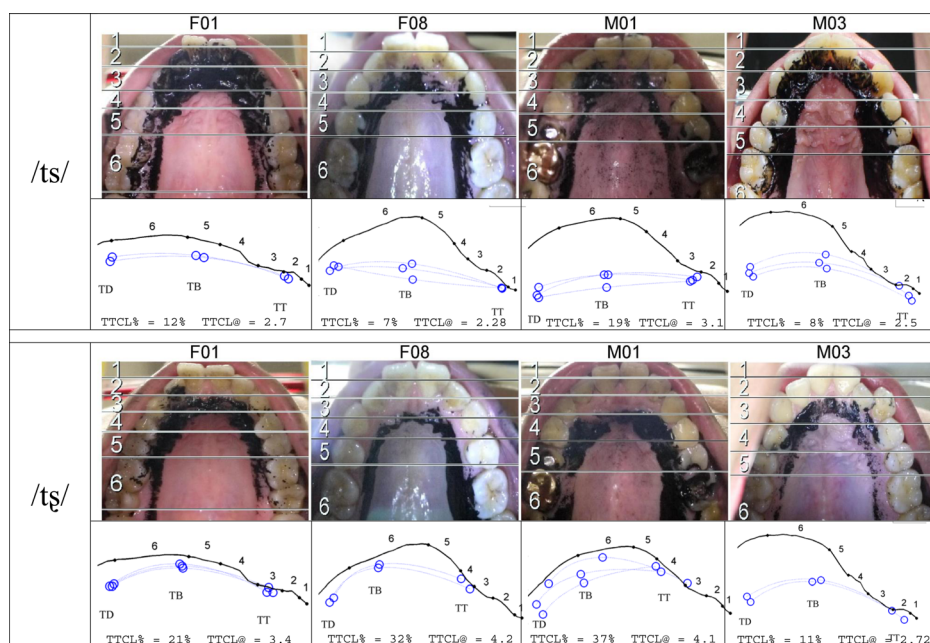

Fig. 2. (Color online) A comparison of consonant constrictions for /ts/ (upper panel) and /tʂ/ (lower panel) elicited by palatography and by SIMPLE method in EMA data for the four subjects.

Table 1. Summary for the comparisons of consonant constrictions for /ts/ and /tʂ/ in Fig. 2.

	Subject: F01	Subject: F08	Subject: M01	Subject: M03
/ts/				
SIMPLE	Dentialveolar	Dentialveolar ∼ alveolar	Dentialveolar ∼ alveolar	Dentialveolar
Palatography	Dentialveolar	Dentialveolar	Dentialveolar ∼ alveolar	Dentialveolar
/tʂ/				
SIMPLE	Alveolar	Alveolar ∼ Postalveolar	Postalveolar	Dentialveolar
Palatography	Alveolar	Alveolar ∼ Postalveolar	Postalveolar	Alveolar

     Table 2 compares the Cartesian coordinates of TT positions (TTx and TTy) with palate-related constriction locations ($TTCL\%$ and $TTCL@$) for /ts/ and /tʂ/, averaged across subjects. The values in the parentheses are the standard deviations (SD). The results in Table 2 reveal a larger overlapping between /ts/ and /tʂ/ in terms of the variations (represented by ± 1 SD) of TT position (both TTx and TTy) than that of constriction location (both $TTCL\%$ and $TTCL@$), indicating that transforming TT positions to TTCL parameters may reduce cross-subject variability and better capture the contrast between /ts/ and /tʂ/.

4. Discussion and conclusion

4.1 Usage of the proposed method

The evaluation results show that the consonant targets determined by the SIMPLE method are largely consistent with those in palatographic data, suggesting the method is reliable for providing a fine-grained palatal outline with subject-specific anatomical landmarks in EMA data. This also suggests that SIMPLE can be used not only to determine constriction locations for consonants and vowels, but also for speaker-normalized quantification of constriction locations, given that the few required points of anatomical invariance can be assumed and found on the PLC palatal outline.

     Using the SIMPLE method with EMA data has advantages over static palatography or EPG in that SIMPLE allows for (1) dynamic tracking of well-defined articulator positions, even when there is no tongue-palate contact, and (2) a larger number of repetitions of trials than is usually possible for a static palatographic study. The SIMPLE method has a lot in common with the TPS method[5] (reconstructing 3D palatal surface from *in vivo* EMA tracings). The advantages of the TPS method include lower cost and greater precision of alignment with EMA data, as compared to the SIMPLE method. However, the SIMPLE method provides more detailed articulatory landmarks and can be applied to other electro-articulometric measures, such as real-time MRI and ultrasound. It is then possible to see the combination of these two methods in future studies.

4.2 Alternative to 3D-scanning for extracting the dental cast palatal line

An alternative way to use 3D-scanning for extracting the mid-sagittal line from the dental cast is to cut the cast in halves and hand-trace the mid-sagittal line on a piece of paper, which is a common practice in traditional palatographic studies,[8–10,13] and then digitize it. We also have performed this practice with our data and compared it with the results of 3D-scanning. Figure 3 shows a comparison of the transformed outputs by the ICP algorithm with PLC acquired by 3D-scanning (circle marker, ○) and by hand-tracing (triangular marker, ▷), for subject M01. The mean error in Fig. 3 is 0.52 mm/pt, which represents the point-wise mean distance between the PLC′-3D-scan and PLC′-hand-tracing. The averaged mean error across four subjects is 0.65 mm/pt (SD = 0.17), which suggests that even if 3D-scanning is not available at the lab, a traditional means of hand-tracing can still yield a reliable PLC as an input to the SIMPLE method.

Table 2. Summary of the Cartesian coordinates of TT positions (TTx, TTy) and palate related constriction locations [$TTCL\%$ and $TTCL@$, defined in Eq. (1) and Eq. (2)], averaged across subjects. Values in the parentheses are the standard deviations.

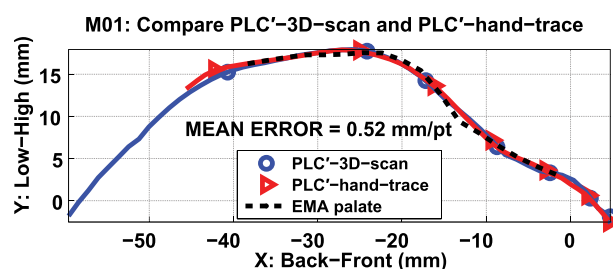|	TTx (mm)	TTy (mm)	$TTCL\%$	$TTCL@$
/ts/	2.8 (5.9)	−1.1 (4)	11.7% (4.7%)	2.7 (0.3)
/tʂ/	−3.9 (9.7)	2.4 (2.8)	25.2% (9.9%)	3.6 (0.6)

Fig. 3. (Color online) Comparison of 3D-scanning and hand-tracing for subject M01.

4.3 Limitations of the proposed method

The proposed method has two limitations. The first caveat of the SIMPLE method is that it relies on shape matching between PLC and the palatal outline traced by EMA; the method will not yield a reliable output if the EMA palatal tracing is not properly performed. The second caveat is that the shape matching by the ICP algorithm relies on shared features on the two shapes. Therefore, if the palatal midline of a subject is comparatively flat, the ICP algorithm may fail to converge. Two possible remedies can be (1) attaching three extra EMA sensors to fixed points on the upper teeth (middle and left and right side) while performing the EMA recording; This can provide three reference points to do rigid-body transformation for the whole dental cast profile transformed into EMA coordinates; and (2) combining 3D *in vivo* EMA tracings[5,7] instead of only the midsagittal line on the palate with SIMPLE method, as mentioned in Sec. 4.1; this would yield more shared features in three dimensions for ICP shape matching. We leave these to future studies.

References and links

[1]L. Fitzpatrick and A. N. Chasaide, "Estimating lingual constriction location in high vowels: A comparison of EMA- and EPG-based measures," J. Phonetics **30**(3), 397–415 (2002).

[2]W. Jones and W. J. Hardcastle, "New developments in EPG3 software," Int. J. Lang. Commun. Disorders **30**(2), 183–192 (1995).

[3]L. Fitzpatrick, "Analysis of lingual articulation using electropalatography and electromagnetic articulography," (Trinity College, Dublin, Ireland, 2001).

[4]C. Geng and C. Mooshammer, "How to stretch and shrink vowel systems: Results from a vowel normalization procedure," J. Acoust. Soc. Am. **125**(5), 3278–3288 (2009).

[5]Y. Yunusova, M. Baljko, G. Pintilie, K. Rudy, P. Faloutsos, and J. Daskalogiannakis, "Acquisition of the 3D surface of the palate by in-vivo digitization with Wave," Speech Comm. **54**(8), 923–931 (2012).

[6]C. Neufeld and A. Anghelescu, "The acoustics and articulation of mandarin sibilants: Improving our data by modeling the palate with EMA," Can. Acoust. **39**(3), 176–177 (2011).

[7]C. Neufeld and P. van Lieshout, "Tongue kinematics in palate relative coordinate spaces for electromagnetic articulography," J. Acoust. Soc. Am. **135**(1), 352–361 (2014).

[8]P. Ladefoged, "Use of palatography," J. Speech Hear. Disorders **22**(5), 764–774 (1957).

[9]V. B. Anderson, "Static palatography for language fieldwork," Lang. Documentation Conserv. **2**(1), 1–27 (2008).

[10]W.-S. Lee, *A Phonetic Study of the Speech of the Cantonese-Speaking Children in Hong Kong* (City University of Hong Kong, Hong Kong, 1999).

[11]P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992).

[12]MathWorks, MATLAB and Optimization Toolbox Release 2014a (MathWorks Inc., Natick, MA, 2014).

[13]V. Anderson, I. Ko, W. O'Grady, and M. Choo, "A palatographic investigation of place of articulation in Korean coronal obstruents," Korean Ling. **12**(1), 1–24 (2004).

[14]P. Ladefoged and I. Maddieson, "The sounds of the world's languages," Language **74**(2), 374–376 (1998).

[15]A. Gafos, "A grammar of gestural coordination," Nat. Language Linguistic Theory **20**(2), 269–337 (2002).

[16]J. A. Shaw, A. Gafos, P. Hoole, and C. Zeroual, "Syllabification in Moroccan Arabic: Evidence from patterns of temporal stability in articulation," Phonology **26**(01), 187–215 (2009).

[17]S. N. Dart, "Comparing French and English coronal consonant articulation," J. Phonetics **26**(1), 71–94 (1998).