VisArtico: a visualization tool for articulatory data

Slim Ouni[1], *Loïc Mangeonjean*[1], *Ingmar Steiner*[2–4]

[1]Université de Lorraine, LORIA, UMR 7503,Vandoeuvre-lès-Nancy, F-54500, France;
[2]Inria, Villers-lès-Nancy, F-54600, France;
[3]School of Computer Science and Informatics, University College Dublin;
[4]Speech Communication Laboratory, Trinity College Dublin
{slim.ouni|ingmar.steiner}@loria.fr

Abstract

In this paper, we present *VisArtico*, a visualization tool for articulatory data acquired using the AG500 3D electromagnetic articulograph (EMA). This software allows displaying the positions of the EMA coils that are simultaneously animated with playback of the acoustic speech signal. It is also possible to display contours for the tongue and lips. The software helps to find the midsagittal plane of the speaker and offers data-based palate shape discovery. In addition, *VisArtico* allows labeling the articulatory data into phonetic segments. Our main goal is to provide an efficient, easy-to-use tool to visualize articulatory data for researchers working in the field of speech production.
Index Terms: speech production, articulatory data, vocal tract, visualization, electromagnetic articulography

1. Introduction

1.1. Background

When we study speech production, we are concerned with both acoustic and articulatory information. Unlike acoustic recordings, articulatory data acquisition is a delicate task which must overcome several practical difficulties.

X-ray imaging was used for several decades [1, 2]. But its use has been greatly reduced or even prohibited in most countries, due to health and safety concerns. Nowadays, non-hazardous modalities such as magnetic resonance imaging (MRI), ultrasound tongue imaging (UTI) or electromagnetic articulography (EMA) are widely used for speech production research. The data obtained by these techniques can be used to track the movement of the tongue, jaw, velum and lips during speech. Some techniques provide good spatial resolution (MRI) and others, a high temporal resolution (EMA).

In this paper, we focus on the visualization of articulatory data obtained by EMA. This acquisition technique tracks the positions of small electromagnetic coils attached to the speech articulators. The positions and orientations of these coils are calculated by measuring the electrical currents produced within multiple low-intensity electromagnetic fields, whose characteristics depend on time [3,4]. This technique is known to present no risk to the health of the speaker [5].

We are particularly interested in the visualization of data acquired by the AG500 3D articulograph (Carstens Medizinelektronik GmbH, Germany), which simultaneously measures the movements of 12 sensor coils at a frequency of 200 Hz. The coils are usually glued to the tongue surface, lips, and on one lower incisor; three reference coils are generally used to subtract the speaker's head movements and transform the articulatory data into a fixed space. Indeed, the speaker is not restrained within in the measurement volume of the articulograph and may move relatively freely.

1.2. Motivation

The *Calpos* software supplied with the articulograph can be used to post-process the raw measurement data, producing coils position data with 5 degrees of freedom (DOF) per coil (three Cartesian and two spherical coordinates per frame). Researchers must then use their own resources (scripts, programs, etc.) to interpret these data according to their objectives. Unfortunately, this severely limits the usefulness of EMA for researchers who do not possess the necessary programming skills. For example, phoneticians are very interested in articulatory data, but in the absence of suitable software for visually inspecting it, EMA data remains all but inaccessible.

A small collection of programs are already available to inspect and visualize EMA data, including EMATOOLS [6], MVIEW [7], Carstens *JustView* [8], and a few others. Unfortunately, some of the existing software tools are no longer maintained or even available for download. Others can be used only on computers running certain versions of Windows, or require a (fairly expensive) license of the commercial MATLAB computing and simulation platform. The latter prerequisite involves significant initial effort with a steep learning curve, and can intimidate or frustrate the non-technically minded user who is interested only in analyzing articulatory speech data.

For these reasons, we have developed *VisArtico*, a lightweight, easy-to-use software tool which allows visualizing EMA data, and which can be run on any computer that supports Java. The software has been designed so that it can directly use the data provided by the articulograph [9] to display the articulatory coil trajectories, synchronized with the corresponding acoustic recordings. Moreover, *VisArtico* not only allows viewing the coils but also enriches the visual information by indicating clearly and graphically the data for the tongue, lips and jaw, and offers some advanced functionality. In the following sections, we describe the software and its main features.

2. User interface of *VisArtico*

VisArtico is a data visualization tool and thus the graphical user interface (GUI) is the most important component. Figure 1 and Figure 3 show an overview of this GUI. In the following subsections, we present the main features of the software. Through this description, the main functionalities will be also explained.

The user interface is composed of three main view panels:

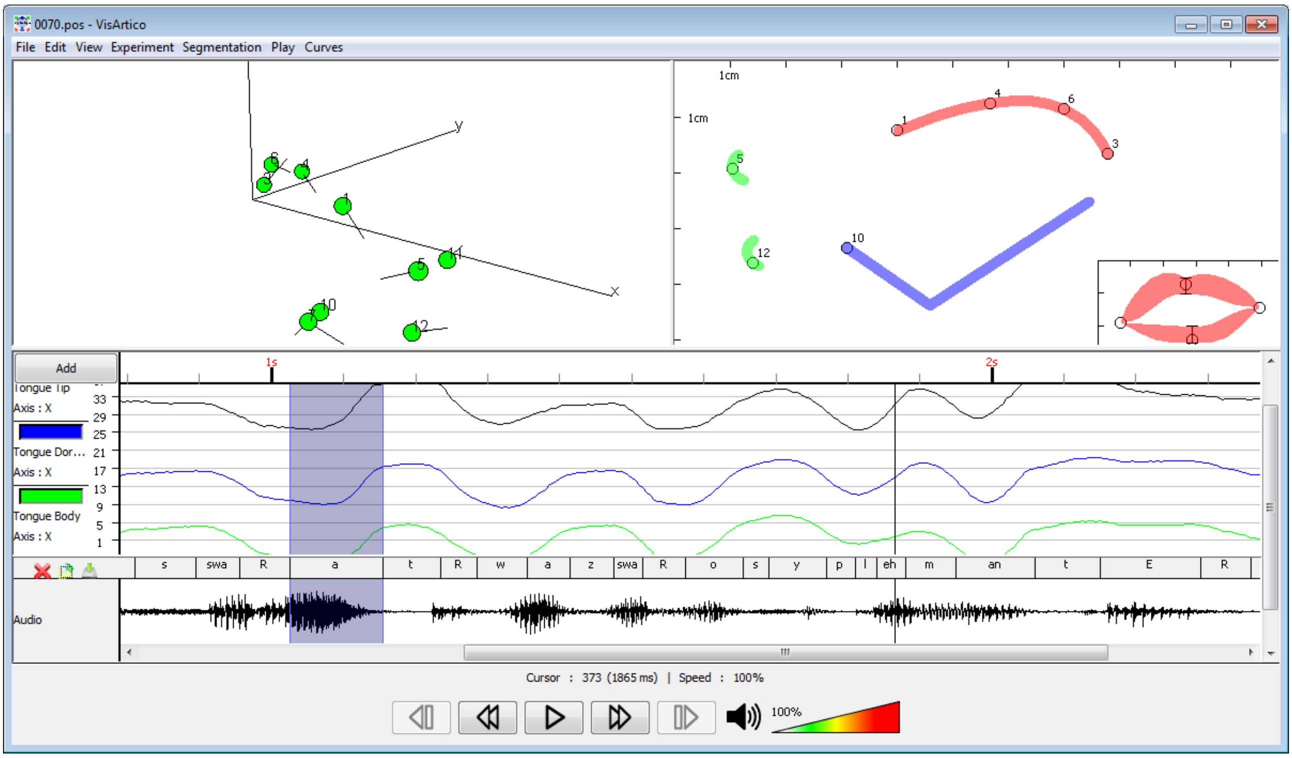

Figure 1: *Main user interface of* VisArtico *(running in Windows). The coils are shown in green in the 3D view (top left panel), with the lip coils in the foreground; the top right panel shows the midsagittal view, with the tongue contour (red), jaw (blue), and lips (green, and in the front-view insert, red). The temporal view is shown in the lower half of the window, displaying several channel trajectories, a phonetic segmentation, and the acoustic waveform.*

(1) the 3D spatial view, (2) the midsagittal view, and (3) the temporal view. *VisArtico* is able to simultaneously animate the three different views of the data in a synchronized way. When the user changes the cursor position in the temporal view, the two other views are adjusted automatically.

2.1. 3D spatial view

The position and orientation of each coil in the articulograph's reference space are displayed in the 3D view panel. The user also has the possibility to graphically connect certain coils together using segments or spline curves, if the arrangement of coils exploits the third dimension. Furthermore, in addition to the current coil positions (as selected in the temporal view), a selection or all of the time frames of the coil trajectories can be displayed as 3D point clouds. This view allows to fully exploit the AG500 to display EMA data in 3D (older articulograph models, the AG100 and AG200, can only measure the coils in a plane).

2.2. Midsagittal view

The Midsagittal view panel presents, in a 2D plane, a midsagittal slice of the vocal tract that can reveal the contour of the tongue, lips, and jaw. The tongue contour is shown as a spline interpolation through the coils on the tongue. The angular shape of the jaw is an approximation and is used only for a better interpretation of the data. In fact, the position of the jaw is calculated based on the position of the coil attached to a lower incisor (if available); in the absence of complementary information, it is very difficult to accurately predict the position of the jaw. The palate contour can also be displayed if palate trace data is available (otherwise, it can be calculated from the data, see Section 3.1). Finally, a front view of the lips is also available to observe the degree of lip opening. The accuracy of the displayed shape of the lips depends on the number of corresponding coils (between two and four).

The midsagittal view provides a reduced and clear representation of the data that interests a lot of researchers in the field of speech production, such as phoneticians. It displays the EMA data in a way that can be compared to classical vocal tract diagrams.

2.3. Temporal view

The temporal view panel displays a time signal representation of individual coil trajectories, synchronized with the acoustic recording and the corresponding phonetic labeling. The user can select and display the curves for any selection of individual EMA channels, viz. the Cartesian coordinates (x, y, z) and spherical coordinates (ϕ, θ), as well as the reported root-mean-square error (RMSE) for any number of EMA coils. The trajectories can be displayed in a single combined panel, or in several separate ones. The temporal representation also allows selecting time intervals for local inspection. This selection can be done at the phone level if a phonetic segmentation is available.

The acoustic recording can be played for individual selections, or for the entire file, while the 3D and midsagittal views display the animated EMA coils. Both the acoustic playback and the coil animation can be slowed down for detailed analysis.

2.4. Phonetic labeling

Adding the phonetic information is very useful for a better interpretation of articulatory gestures; therefore, *VisArtico* allows labeling of articulatory data. The user can create the labeling directly in the GUI's temporal view, or import a labeling file created with external software, e.g., Winsnoori [10], Praat [11], WaveSurfer [12], or a simple custom format. It is possible to add multiple levels of labeling in a hierarchical or autosegmental paradigm; e.g., there can be one segmentation for phones, another for words, phrases, etc.

2.5. EMA configuration

When data from an EMA acquisition session is loaded into *VisArtico* for the first time, the user can supply information about the configuration of EMA coils, i.e., the layout of coils on the tongue, lips, etc. This is made very intuitive using the configuration module, shown in Figure 2, which creates a mapping between coils and articulators.

The standard articulators include the tongue, lips, jaw and velum. The software can determine the shape of the lips from two coils (interpolated from one coil on the upper and lower lip, respectively) to four coils (one or two additional coils at the corners of the mouth). This information is valuable for studies of lip rounding, for instance.

The configuration module is also used to determine the speaker's midsagittal plane, allowing a better interpretation of the midsagittal view, since the orientation of the speaker's head in the cube of the articulograph may not be well-aligned with the reference space of the articulograph. Hence, *VisArtico* uses the coil positions of the tongue to determine the midsagittal plane. It first calculates the directrix of a cloud of points (positions of these coils). Subsequently, the software performs a rotation with respect to this directrix. In Figure 2, the three schematic heads (representing the axial, sagittal, and coronal plane, respectively) show the result of this correction. The red arrow indicates the orientation of the midsagittal plane.

The configuration only needs to be done once per set of EMA data. *VisArtico* stores the configuration for other files from the same acquisition, and loads it automatically when the files are reopened at a later time.

3. Advanced features

3.1. Palate contour detection

The visualization of the palate allows a better interpretation of the tongue movement. In particular, it helps to better display the places of constriction for obstruent phonemes.

During an EMA acquisition session, one of the sweeps is commonly reserved for tracing the contour of the palate. This can be done by attaching one coil to a stick or pen and manually sliding it along the roof of the mouth. *VisArtico* can then exploit this palate trace sweep to extract the contour of the palate and display it along with the data from the rest of the session.

Sometimes however, for a variety of reasons, palate trace data may not be available. *VisArtico* can then calculate an approximation of the missing palate contour, using a simple but effective algorithm, which predicts the contour from the convex hull of the tongue coil positions. For each time frame, we first determine the contour of the tongue from the coils associated with it. We obtain several contours of the tongue throughout the recording session. Thereby the algorithm recovers the maxima of each contour in the midsagittal plane, which should correspond to a contact between the tongue and the palate. The more data is used for this palate discovery, the more realistic the resulting contour becomes. This solution allows to compensate for missing palate trace data.

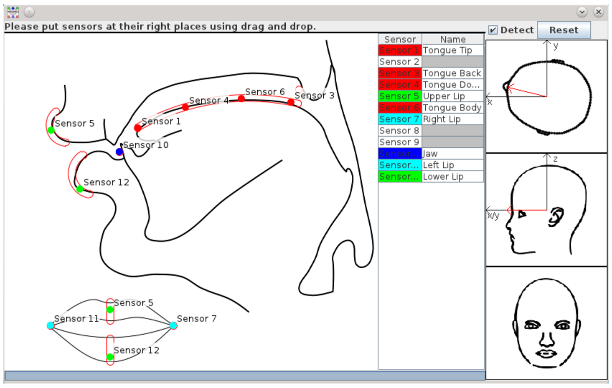

Figure 2: *The configuration module (running in Linux). This tool makes it easy to map the coils accurately to articulators and to determine the midsagittal plane.*

3.2. Filtering

The EMA data obtained by processing the raw articulograph amplitude measurements may contain errors that are usually indicated by a high RMSE, due to machine noise or faulty coils, or errors in processing.

VisArtico offers the possibility of applying filters to one or more coil trajectories to eliminate signal noise or unrealistic coil positions. In a classic approach, we use a cutoff frequency of 20 Hz to smooth the trajectories, but this parameter can be customized by the user in the filtering module. Moreover, the user may choose to ignore signal intervals for which the RMSE exceeds a given threshold and for which smoothing does not correct this error; during such intervals, the data values can be linearly interpolated.

4. Conclusion

We have presented a new software tool for the visualization of articulatory data obtained by EMA. The goal is to expand the user base for electromagnetic articulography to researchers who are not necessarily computer experts.

An overview of *VisArtico*, as well as a detailed user guide, are available at a dedicated website, `http://visartico.loria.fr/`. The software is freely available to researchers and students.

We plan to make further improvements to *VisArtico*, including the possibility for more more sophisticated filtering and data analysis (such as providing coil velocity and acceleration); displaying a spectrogram with the acoustic speech signal, and providing additional acoustic information (such as formant tracks); extending the configuration module for 3D tongue coil layouts; and adding localization for other languages (currently, the GUI supports English and French).

The design of the software permits easy adaptation for use with other articulograph models beside the AG500, such as older 2D models, the recent AG501, or the Wave system from NDI. Moreover, while the current version can load only the data format produced by the Carstens *CalcPos* software, support is planned for other common EMA data formats, such as that pro-

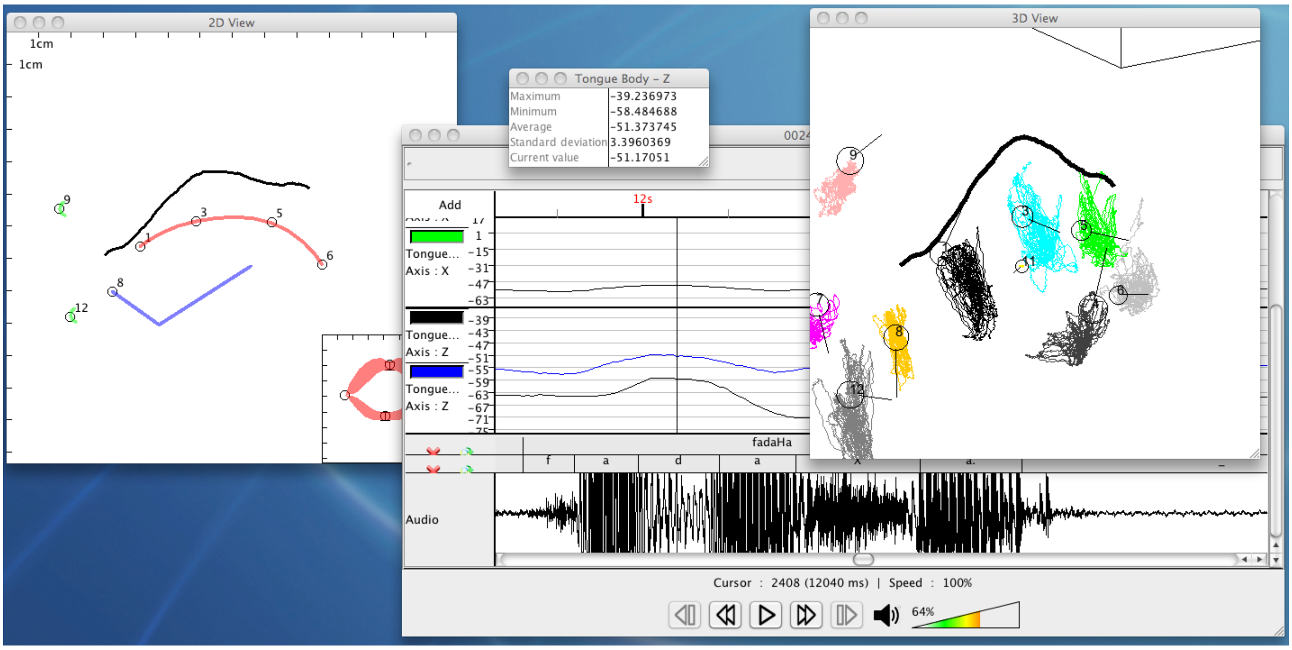

Figure 3: *Several detached view panels of* VisArtico *(running in Mac OS X): (left) midsagittal view with the palate contour drawn in black; (center) temporal view, including two phonetic annotation tiers; (right) 3D view with coil trajectories rendered as colored clouds along with the palate contour; (top) properties panel for one EMA channel, which displays some useful statistics.*

duced by TAPADM [13], or the Edinburgh Speech Tools (EST) track files distributed in some publicly available corpora, such as MOCHA-TIMIT [14] or the more recent *mngu0* database [15].

We believe that this software is very useful for the speech science community and makes the use of articulatory data more accessible. Even undergraduate students of linguistics or speech therapy without advanced computer skills should be able to use *VisArtico* with ease to explore articulatory data and improve their understanding of speech production.

5. Acknowledgments

The presentation of this work is supported by the Science Foundation Ireland under Grant No. 07/CE/I1142 as part of the Centre for Next Generation Localisation (`http://cngl.ie`).

6. References

[1] J. S. Perkell, *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*. Cambridge, MA: MIT Press, 1969.

[2] O. Fujimura, S. Kiritani, and H. Ishida, "Computer controlled radiography for observation of movements of articulatory and other human organs," *Computers in Biology and Medicine*, vol. 3, no. 4, pp. 371–378, Dec. 1973, doi:10.1016/0010-4825(73)90003-6.

[3] J. S. Perkell, M. H. Cohen, M. A. Svirsky, M. L. Matthies, I. Garabieta, and M. T. T. Jackson, "Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements," *Journal of the Acoustical Society of America*, vol. 92, no. 6, pp. 3078–3096, Dec. 1992, doi:10.1121/1.404204.

[4] P. Hoole and A. Zierdt, "Five-dimensional articulography," in *Speech Motor Control: New developments in basic and applied research*, B. Maassen and P. van Lieshout, Eds. Oxford University Press, 2010, ch. 20, pp. 331–349.

[5] M. Hasegawa-Johnson, "Electromagnetic exposure safety of the Carstens Articulograph AG100," *Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 2529–2532, Oct. 1998, doi:10.1121/1.423775.

[6] N. Nguyen, "A MATLAB toolbox for the analysis of articulatory data in the production of speech," *Behavior Research Methods, Instruments, & Computers*, vol. 32, no. 3, pp. 464–467, Sep. 2000, doi:10.3758/BF03200817.

[7] M. K. Tiede. MVIEW: Multi-channel visualization application for displaying dynamic sensor movements.

[8] Carstens Medizinelektronik, *JustView: AG500 measuring environment display*, Lenglern, Germany, 2006. [Online]. Available: http://www.ag500.de/manual/ag500/JustView.pdf

[9] ——, *AG500 Data Format and Data Structure*, Lenglern, Germany, 2004. [Online]. Available: http://www.ag500.de/manual/ag500/AG500_ida_format.pdf

[10] Y. Laprie, *WinSnoori version 1.34: A Speech research tool*, LORIA, 2009. [Online]. Available: http://www.winsnoori.fr/

[11] P. Boersma and D. Weenink. (2012) Praat: doing phonetics by computer. [Online]. Available: http://www.praat.org/

[12] K. Sjölander and J. Beskow, "Wavesurfer: An open source speech tool," in *Proc. 6th International Conference on Spoken Language Processing (ICSLP)*, vol. 4, Beijing, China, Oct. 2000, pp. 464–467. [Online]. Available: http://www.isca-speech.org/archive/icslp_2000/i00_4464.html

[13] A. Zierdt. Three-dimensional Artikulographic Position and Align Determination with MATLAB. [Online]. Available: http://wiki.ag500.net/TAPADM

[14] A. A. Wrench, "A multi-channel/multi-speaker articulatory database for continuous speech recognition research," *PHONUS*, vol. 5, pp. 1–14, Dec. 2000. [Online]. Available: http://www.coli.uni-saarland.de/groups/WB/Phonetics/contents/phonus-pdf/phonus5/Wrench_PHONUS5.pdf

[15] K. Richmond, P. Hoole, and S. King, "Announcing the electromagnetic articulography (day 1) subset of the mngu0 articulatory corpus," in *Proc. Interspeech*, Florence, Italy, Aug. 2011, pp. 1505–1508.